



**ALÉM DAS NUVENS:
EXPANDINDO AS FRONTEIRAS DA CIÊNCIA DA
INFORMAÇÃO**

Isa M. Freire, Lilian M. A. R. Alvares,
Renata M. A. Baracho, Mauricio B. Almeida,
Beatriz V. Cendon, Benildes C. M. S. Maculan
(Org.)



**ALÉM DAS NUVENS:
EXPANDINDO AS FRONTEIRAS DA CIÊNCIA DA
INFORMAÇÃO**

ISSN 2177-3688

BELO HORIZONTE
ECI/UFMG
2014

DIREITO AUTORAL E DE REPRODUÇÃO

Direitos de autor ©2014 para os artigos individuais dos autores. São permitidas cópias para fins privados e acadêmicos, desde que citada a fonte e autoria. A republicação desse material requer permissão dos detentores dos direitos autorais. Os editores deste volume são responsáveis pela publicação e detentores dos direitos autorais.

E56a
2014

Encontro Nacional de Pesquisa em Ciência da Informação : além das nuvens, expandindo as fronteiras da Ciência da Informação (15. : 2014 : Belo Horizonte, MG).

Anais [recurso eletrônico] / XV Encontro Nacional de Pesquisa em Ciência da Informação : além das nuvens, expandindo as fronteiras da Ciência da Informação, 27-31 de outubro em Belo Horizonte, MG. / Organizadores: Isa M. Freire, Lilian M. A. R. Álvares, Renata M. A. Baracho, Maurício B. Almeida, Beatriz V. Cendon, Benildes C. M. S. Maculan. – Belo Horizonte, ECI, UFMG, 2014.

ISSN 2177-3688

Evento realizado pela Associação Nacional de Pesquisa em Ciência da Informação (ANCIB) e organizado pelo Programa de Pós-Graduação em Ciência da Informação da Universidade Federal de Minas Gerais (PPGCI-ECI/UFMG).

1. Evento – Ciência da Informação. 2. Evento – Pesquisa em Ciência da Informação. I. Título.

CDU: 02(063)(81)

COMISSÃO CIENTÍFICA

Profa. Dra. Renata Maria Abrantes Baracho – UFMG: Presidente
Profa. Dra. Lillian Alvares – UnB
Profa. Dra. Icléia Thiesen – Unirio
Profa. Dra. Brígida Maria Nogueira Cervantes – UEL
Profa. Dra. Giulia Crippa - USP
Profa. Dra. Emeide Nóbrega Duarte – UFPB
Prof. Dr. Clóvis Montenegro de Lima – IBICT
Profa. Dra. Aida Varela - UFBA
Profa Dra. Leilah Santiago Bufrem – UFPE
Profa. Dra. Plácida Amorim da Costa Santos – Unesp/Marília
Profa. Dra. Luisa M. G. de Mattos Rocha – IPJB/RJ
Prof. Dr. Carlos Xavier de Azevedo Netto – UFPB
Profa. Dra. Maria Cristina Soares Guimarães - IBICT/Fiocruz

PARECERISTAS DA COMISSÃO CIENTÍFICA DO GT 8

INFORMAÇÃO E TECNOLOGIA

Ana Maria Pereira
Angela Maria Grossi de Carvalho
Carlos Henrique Marcondes
Cenidalva Teixeira
Cesar Cusin
Claudio Jose Silva Ribeiro
Divino Ignacio Ribeiro Jr
Edberto Ferneda
Elvis Fusco
Fabiano Ferreira de Castro
Fernando Modesto
Gercina Lima
Guilherme Ataíde Dias
Izabel França de Lima
José Eduardo Santarem Segundo
Luis Fernando Sayão
Marckson Sousa
Marcos L Mucheroni
Maria Elisabete Catarino
Maria Inês Tomaél
Maria José Vicentini Jorente
Maurício Barcellos Almeida
Murilo Bastos Cunha
Nanci Oddone
Patrícia Zeni Marchiori
Plácida L. V. Amorim Costa Santos
Rachel Cristina Vesú Alves
Renata Maria Abrantes Baracho
Renato Fernandes Corrêa
Renato Rocha Souza
Ricardo César G. Santana
Rogério Ap. Sá Ramalho
Silvana Ap. Borsetti Gregorio Vidotti
Silvana Drumond
Vinícius M. Kern
Wagner Junqueira de Araujo
Zaira Regina Zafalon

Realização



Agências de Fomento



Grupos de pesquisa



Grupo de Pesquisa e Estudos em Metodologia
Ação e Lettrice na Tecnologia, Educação e Cérebro



Apoio



UNIVERSIDADE
FUMEC
DE MINAS GERAIS PARA O MUNDO

De olho na CI



livrosdebiblio



GT 8

INFORMAÇÃO E TECNOLOGIA

SUMÁRIO

PREFÁCIO	6
GT 8 – INFORMAÇÃO E TECNOLOGIA.....	3710
Modalidade da apresentação: Comunicação oral	3710
TECNOLOGIA E SISTEMAS DE INFORMAÇÃO NA CIÊNCIA DA INFORMAÇÃO: PERCURSO DA TEMÁTICA TECNOLÓGICA NO ARIST.....	3710
<i>Ronaldo Ferreira Araujo</i>	
<i>Marlene Oliveira</i>	
IMAGEM & TECNOLOGIA EM CATÁLOGOS DE MUSEUS DE ARTE: INTERSEMOSE EM QUESTÃO.....	3726
<i>Fábio Rogério Batista Lima</i>	
<i>Plácida Leopoldina Ventura Amorim da Costa Santos</i>	
UTILIZANDO O GOOGLE ANALYTICS PARA MONITORAR A AUDIÊNCIA DE BLOGS	3745
<i>Célio Andrade Santana Júnior</i>	
<i>Camila Oliveira Lima</i>	
<i>Diego Andres Salcedo</i>	
<i>Andréia Candida de Souza</i>	
BASES METODOLÓGICAS PARA ARQUITETURA DA INFORMAÇÃO PERVASIVA	3765
<i>Henry Poncio Cruz de Oliveira</i>	
<i>Silvana Aparecida Borsetti Gregorio Vidotti</i>	
<i>Virginia Bentes Pinto</i>	
ENCONTRABILIDADE DA INFORMAÇÃO: ATRIBUTOS E RECOMENDAÇÕES PARA AMBIENTES INFORMACIONAIS DIGITAIS.....	3780
<i>Fernando Luiz Vechiato</i>	
<i>Silvana Aparecida Borsetti Gregorio Vidotti</i>	
CONVERSÃO DE REGISTROS EM XML PARA MARC 21: UM MODELO BASEADO EM XSLT.....	3802
<i>Fabrício Silva Assumpção</i>	
<i>Plácida Leopoldina Ventura Amorim da Costa Santos</i>	
APLICAÇÃO DE DADOS INTERLIGADOS ABERTOS APOIADA POR ONTOLOGIA	3822
<i>Linair Maria Campos</i>	
<i>Maria Luiza de Almeida Campos</i>	
USUÁRIOS EM PRÁTICAS PROFISSIONAIS DE BIBLIOTECÁRIOS E ANALISTAS DE TI	3842
<i>Eliane Cristina de Freitas Rocha</i>	
<i>Adriana Bogliolo Sirihal Duarte</i>	
WEB SEMÂNTICA: INTRODUÇÃO A RECUPERAÇÃO DE DADOS USANDO SPARQL	3863
<i>José Eduardo Santarém Segundo</i>	

<i>DESIGN SCIENCE: FILOSOFIA DA PESQUISA EM CIÊNCIA DA INFORMAÇÃO E TECNOLOGIA.....</i>	3883
<i>Marcello Peixoto Bax</i>	
<i>LIVROS DIDÁTICOS DIGITAIS EM AMBIENTES COMPUTACIONAIS NAS NUvens</i>	3904
<i>Charles Rodrigues</i>	
<i>Angel Freddy Godoy Viera</i>	
<i>ESTUDOS BRASILEIROS SOBRE INDEXAÇÃO AUTOMÁTICA NO AMBITO DA CIÊNCIA DA INFORMAÇÃO (1973-2012).....</i>	3924
<i>Remi Correia Lapa</i>	
<i>Renato Fernandes Correa</i>	
<i>DESENVOLVENDO UMA PUBLICAÇÃO CIENTÍFICA AMPLIADA NA PLATAFORMA e-QUITL.....</i>	3942
<i>Adriana Carla Silva de Oliveira</i>	
<i>Guilherme Ataíde Dias</i>	
<i>Mariana Cantisani</i>	
<i>ACESSIBILIDADE A INFORMAÇÃO: ANÁLISE DO WEBSITE DA FUNDAÇÃO APOIO À PESSOA COM DEFICIÊNCIA</i>	3963
<i>Célia Medeiros Dantas</i>	
<i>Hellosman de Oliveira Silva</i>	
<i>Marckson Roberto Ferreira de Sousa</i>	
<i>UMA FERRAMENTA PARA RECUPERAÇÃO DE TAGS DE BLOGS BASEADA EM MICROFORMATOS.....</i>	3980
<i>Célio Andrade Santana Júnior</i>	
<i>Nilton Heck Santos</i>	
<i>Steffane Ramires de Lima</i>	
<i>Amanda Maria de Almeida Nunes</i>	
<i>POLÍTICAS DE FORMATOS DE ARQUIVOS PARA OBJETOS DE APRENDIZAGEM: PRESERVAÇÃO DIGITAL NO SABER TECNOLOGIAS EDUCACIONAIS E SOCIAIS</i>	3998
<i>Vildeane da Rocha Borba</i>	
<i>Sandra de Albuquerque Siebra</i>	
<i>Marcos Galindo</i>	
<i>Josiane Lemos Machiavelli</i>	
<i>Cristine Martins Gomes de Gusmão</i>	
<i>A INTERAÇÃO DO USUÁRIO COM CATÁLOGOS BIBLIOGRÁFICOS <i>ON-LINE</i> DE BIBLIOTECAS UNIVERSITÁRIAS</i>	4016
<i>Flavia Maria Bastos</i>	
<i>Silvana Aparecida Borsetti Gregorio Vidotti</i>	
<i>REDES SOCIAIS EM BIBLIOTECAS: TRABALHOS APRESENTADOS NO XVII SNBU.....</i>	4034
<i>Maira Nani França</i>	
<i>Angela Maria Grossi de Carvalho</i>	
<i>METADADOS ARQUIVÍSTICOS: CONSIDERAÇÕES SOBRE CONCEITOS, TIPOS E INSTRUMENTOS</i>	4050
<i>Rachel Cristina Vesu Alves</i>	

Plácida Leopoldina Ventura Amorim da Costa Santos

UM ROTEIRO PARA MODELAGEM CONCEITUAL DE SISTEMAS DE INFORMAÇÃO BASEADA EM PRINCÍPIOS ONTOLÓGICOS	4072
<i>Stefane de Melo Silva</i>	
<i>Fernanda Farinelli</i>	
<i>Mauricio Barcellos Almeida</i>	
SISTEMA PARA GERENCIAMENTO E RECUPERAÇÃO DE INFORMAÇÃO	4096
<i>Renata Maria Abrantes Baracho</i>	
<i>Cátia Rodrigues Barbosa</i>	
<i>Christiano Benvindo dos Santos</i>	
BIG DATA: UMA INVESTIGAÇÃO COM USO DE DADOS ABERTOS SOBRE ACIDENTES DE TRABALHO	4116
<i>Cláudio José Silva Ribeiro</i>	
ONTOLOGIA BASEADA NOS FRBR: PROPOSTA DE APLICAÇÃO EM CATÁLOGOS ONLINE	4132
<i>Rogério Aparecido Sá Ramalho</i>	
<i>Zaira Regina Zafalon</i>	
<i>Marcos Teruo Ouchi</i>	
AVALIAÇÃO DO ACESSO E VISUALIZAÇÃO DA INFORMAÇÃO EM REPOSITÓRIOS INSTITUCIONAIS	4151
<i>Sandra de Albuquerque Siebra</i>	
<i>Júccia Nathielle do Nascimento Oliveira</i>	
<i>Críssia de Santana Marcelinoa</i>	
Modalidade da apresentação: Pôster.....	4171
INFLUÊNCIA DA INOVATIVIDADE E FINANCIAMENTO NO RESULTADO EM START-UPS	4171
<i>Anselmo Battisti</i>	
<i>Carlos O. Quandt</i>	
TECNOLOGIAS PARA APLICAÇÃO DA WEB SEMÂNTICA NAS BIBLIOTECAS E ARQUIVOS	4178
<i>Marcelo Tomita</i>	
<i>Maria Elisabete Catarino</i>	
FERRAMENTA DE SOFTWARE LIVRE PARA CONSTRUÇÃO DE BASES DE DADOS REFERENCIAIS DE ARTIGOS CIENTÍFICOS DIGITAIS	4185
<i>Flávio Ribeiro Córdula</i>	
<i>Wagner Junqueira de Araújo</i>	
MODELO PARA O DESCARTE SEGURO DA INFORMAÇÃO EM SUPORTE DIGITAL	4193
<i>Silvio Lucas da Silva</i>	
<i>Wagner Junqueira Araújo</i>	
ACESSIBILIDADE WEB NA PERSPECTIVA DOS SISTEMAS DE RECUPERAÇÃO DA INFORMAÇÃO.....	4201
<i>Lucinéia Souza Maia</i>	

ANÁLISE DA AÇÃO DE INFORMAÇÃO RELACIONAL NO LAViD DA UFPB.	4212
<i>Emy Pôrto Bezerra</i>	
<i>Ítalo José Bastos Guimarães</i>	
<i>Zayr Cláudio Gomes da Silva</i>	
BIBLIOTECAS E WEB SEMÂNTICA: ANÁLISE SOBRE O ESTADO DESTE RELACIONAMENTO	4219
<i>Edgar Bisset Alvarez</i>	
ANÁLISE DA ARQUITETURA DA INFORMAÇÃO DO WEBSITE DO SIBI UNIRIO	4228
<i>Alanda do Valle Vitorino</i>	
<i>Claudio Jose Silva Ribeiro</i>	
AVALIANDO A EDITORAÇÃO DE E-BOOKS EM AMBIENTES DE EDITORAS UNIVERSITÁRIAS: UMA APLICAÇÃO DO OPEN MONOGRAPH PRESS	4235
<i>Adriana Carla Silva de Oliveira</i>	
<i>Guilherme Ataíde Dias</i>	
BIG DATA APLICADO A SISTEMAS CIBER-FÍSICOS DA LOGÍSTICA: PROPOSTA DE MODELO CONCEITUAL	4242
<i>Moisés Lima Dutra</i>	
<i>William Barbosa Vianna</i>	
<i>Enzo Morosini Frazzon</i>	
VISUALIZAÇÃO DE CORRESPONDÊNCIAS SEMÂNTICAS NO UNIVERSO BIG DATA	4250
<i>Moises Lima Dutra</i>	
<i>Marcio Matias</i>	
SÍNDROME DE GABRIELA” EM UMA COMUNIDADE RURAL DE JOÃO PESSOA: RESISTÊNCIA E ACEITAÇÃO DAS TECNOLOGIAS DE INFORMAÇÃO E COMUNICAÇÃO NO PROCESSO DE ALFABETIZAÇÃO E INCLUSÃO DIGITAL	4258
<i>Raíssa Carneiro de Brito</i>	
A WEB 2.0 COMO CANAL DE COMUNICAÇÃO ENTRE A BIBLIOTECA E OS USUÁRIOS	4265
<i>Nivea Camara Rocha de Souza</i>	
<i>Ana Paula de Oliveira Villalobos</i>	

PREFÁCIO

A Ciência da Informação é um campo científico de natureza interdisciplinar devotedo à busca por soluções para a efetiva comunicação da informação, bem como de seus registros, no contexto social, institucional ou individual de uso e a partir de necessidades específicas. A evolução da Ciência da Informação está inexoravelmente ligada à tecnologia da informação, uma vez que o imperativo tecnológico tem gerado transformações que culminaram em uma sociedade pós-industrial, a sociedade da informação. Nesse contexto, a Ciência da Informação desempenha importante papel na evolução da sociedade da informação por suas fortes dimensões social e humana, as quais vão além das fronteiras da tecnologia.

O tema do ENANCIB 2014 – *Além das nuvens: expandindo as fronteiras da Ciência da Informação* – remete ao cenário atual caracterizado pelo contínuo desenvolvimento das tecnologias da informação e comunicação, assim como pela evolução constante do ambiente Web, os quais têm proporcionado novas formas de acessar, recuperar, armazenar e gerir a informação. Telefonia móvel, nuvens, big data, linked data, dentre outras formas de interagir com a informação têm exigido novas abordagens para os estudos em Ciência da Informação. O ENANCIB 2014 oferece a oportunidade para refletir sobre essas mudanças, as quais impactam na interação humana com a informação, bem como sobre suas implicações para o futuro da Ciência da Informação.

Promovido pela Associação Nacional de Pesquisa em Ciência da Informação (ANCIB), o ENANCIB, em sua décima quinta edição, foi organizado pelo Programa de Pós-Graduação em Ciência da Informação da Universidade Federal de Minas Gerais (PPGCI-ECI/UFMG) e realizado na Escola de Ciência da Informação da Universidade Federal de Minas Gerais (ECI/UFMG), em Belo Horizonte, Minas Gerais, no período de 27 a 31 de outubro de 2014. O evento foi financiado pela Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), pela Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG), pela UFMG e outras organizações apoiadoras.

Pesquisadores em Ciência da Informação foram convidados a submeter pesquisas teóricas e empíricas, de acordo com a orientação temática dos onze Grupos de Pesquisa (GTs) da ANCIB. A chamada de trabalhos foi aberta para duas categorias de submissões. A primeira categoria é a comunicação oral (máximo de 20 páginas), que consiste de artigo escrito em português, descrevendo trabalho original com demonstração efetiva de resultados. As comunicações orais aprovadas foram convidadas para apresentação no evento. A segunda

categoria é o pôster (máximo de 7 páginas), que consiste de artigos curtos escritos em português, descrevendo pesquisa em desenvolvimento. Os pôsteres aceitos foram convidados para exposição nas dependências em que ocorreu o evento.

O ENANCIB 2014 recebeu mais de 600 trabalhos, dos quais mais de 300 foram aceitos para publicação nos Anais, sendo cerca de 240 para apresentação oral e 80 para exibição em pôsters. Este volume é então constituído por 74% de comunicações orais e 26% de pôsteres, selecionados pelo comitê de programa dos GTs, os quais são compostos por pareceristas especializados, definidos no âmbito de cada GT.

Agradecemos à Comissão Organizadora e à ANCIB pelo seu comprometimento com o sucesso do evento, aos autores por suas submissões e à Comissão Científica pelo intenso trabalho. Agradecemos ainda aos alunos, funcionários e colaboradores que contribuíram para a efetivação do evento.

Belo Horizonte, outubro de 2014

Isa M. Freire

Lilian M. A. R. Alvares

Renata M. A. Baracho

Mauricio B. Almeida

Beatriz V. Cendon

Benildes C. M. S. Maculan

GT 8 – INFORMAÇÃO E TECNOLOGIA

Modalidade da apresentação: Comunicação oral

TECNOLOGIA E SISTEMAS DE INFORMAÇÃO NA CIÊNCIA DA INFORMAÇÃO: PERCURSO DA TEMÁTICA TECNOLÓGICA NO ARIST

*TECHNOLOGY AND INFORMATION SYSTEMS IN THE INFORMATION SCIENCE: THE
TECHNOLOGICAL THEME ON ARIST*

Ronaldo Ferreira Araujo
Marlene Oliveira

Resumo: A revolução científica e tecnológica promoveu o surgimento de disciplinas científicas emergentes como a ciência da informação, e o avanço tecnológico desempenha um papel relevante no desenvolvimento dessa área. Ao longo das décadas percebemos várias passagens que marcam esse avanço e demonstram sua contribuição para as problemáticas da informação. Para contribuir com essa reflexão, a temática tecnológica presente nas publicações do Annual Review of Information Science and Technology (ARIST) é analisada. A partir dos descritores ‘informatics’, ‘infomation technology’ e ‘information systems’ foram identificados 87 artigos de revisão distribuídos entre os anos de 1966 e 2008. Por meio da análise dos títulos e palavras-chave apresentamos os descritores mais recorrentes por período e uma clusterização das temáticas. O número de trabalho decresce ao longo do período analisado ao passo que o quantitativo de descritores se amplia, indicando uma evolução terminológica. As tecnologias estão presentes em questões interdisciplinares, aspectos educacionais e profissionais, e tem como núcleo os sistemas de recuperação de informação.

Palavras-chave: Ciência da Informação; Tecnologia da Informação; Sistemas de Informação; ARIST.

Abstract: The scientific and technological revolution promotes the arising of emerging sciences and in-formation science and technological advancement plays a role in the development of this area. Over the decades we noticed several passages that mark this advance and demonstrate their contribution to the information problems. To contribute to the discussion we analyze the issue in this technology publications on Annual Review of Information Science and Technology (ARIST). From the descriptors' infomation technology 'and' information systems'foram identi-fied 87 review articles. Through the analysis of titles, keywords and summary present the most recurrent themes. The technologies are present in interdisciplinary issues, educational and professional aspects, the information retrievable systems are central.

Keywords: Information Science; Information Technology; Information Systems; ARIST.

1 INTRODUÇÃO

A revolução científica e tecnológica promoveu o surgimento de disciplinas científicas emergentes como a Ciência da Informação (CI), e o avanço tecnológico desempenha um papel relevante no desenvolvimento dessa área. Ao longo das décadas percebemos várias passagens que marcam esse avanço e demonstram sua contribuição para as problemáticas da informação.

Para Santos et. al. (2013) as tecnologias de informação e comunicação estão cada vez mais presentes no desenvolvimento da CI e na produção científica resultante de seus estudos e pesquisas. Segundo os autores, ao avanço tecnológico que caracteriza este momento considerado histórico se deve à essência da atuação do campo no tratamento da informação, na gestão de recursos informacionais e na mediação da informação para a apropriação do conhecimento.

A pesquisa na CI envolvendo a temática ‘tecnologia’ conduziu a criação do “GT8 Informação e Tecnologia”, um dos grupos da Associação Nacional de Pesquisa e Pós-Graduação em Ciência da Informação (ANCIB). O GT em questão iniciou-se em 2008 no Encontro Nacional de Pesquisa e Pós-Graduação em Ciência da Informação (ENANCIB) e dedica-se aos estudos e pesquisas teórico-práticos sobre e para o “desenvolvimento de tecnologias de informação e comunicação que envolvam os processos de geração, representação, armazenamento, recuperação, disseminação, uso, gestão, segurança e preservação da informação em ambientes digitais” (GRUPOS DE TRABALHO, GT8, 2014).

De acordo com Vidotti, Oliveira e Lima (2013, p. 3) “o GT 8 da ANCIB tem contribuído com a socialização de pesquisas de cunho tecnológico focadas no objeto de investigação da Ciência da Informação”.

É sob tal afirmação que Santos e outros (2013) desenvolveram uma pesquisa que procurou mapear a temática tecnológica e suas variações na comunidade científica da CI em periódicos da área, indexados no *Web of Science* (WOS), classificados no WebQUALIS 2013 e disponíveis no Portal CAPES de Periódicos.

Nesse sentido propõe-se outra forma de reflexão sobre o componente tecnológico na produção científica da CI, mas elege como fonte de pesquisa a base de conhecimento acumulado no *Annual Review of Information Science and Technology* (ARIST) ao longo dos seus 45 anos de publicação (1966 a 2011).

2 TECNOLOGIA, CIÊNCIA DA INFORMAÇÃO E O ARIST

Os Estados Unidos da América (EUA) pode ser considerado, segundo alguns pesquisadores, como o berço da Ciência da Informação (PINHEIRO, 2002), tendo sido a sociedade americana marcada por várias transformações (científicas, tecnológicas e sociais) que propiciaram o nascimento dessa nova ciência. Alguns acontecimentos marcantes antecederam seu nascitudo, entre os quais a Segunda Grande Guerra; o desenvolvimento da Ciência e Tecnologia (C&T) e da Pesquisa e Desenvolvimento (P&D); muito em função da própria guerra e, entre as tecnologias, o *Memex*, ideia de Vannevar Bush de 1945.

A expressão “Ciência da Informação” foi se consolidando aos poucos ao longo da década de 1960. Sua denominação aconteceu em duas conferências realizadas no *Georgia Institute of Technology*, em 1961 e 1962; depois, com a criação, em 1966, do *Annual Review of Information Science and Technology* (ARIST); Em 1968, com a mudança do nome do *American Documentation Institute* (ADI) para *American Society for Information Science* (ASIS), como nova alteração em 2000 para *American Society for Information Science and Technology* (ASIS&T). A nova denominação indica sua adesão à tecnologia da informação; e recentemente, em 2013 muda-se para *Association for Information Science and Technology*, mantendo a sigla, mas refletindo seu crescimento internacional. Em sua página da internet, a ASIS&T disponibiliza acesso ao projeto *Pioneers of Information Science in North America Web* <<http://faculty.libsci.sc.edu/bob/ISP/contents.htm>>, que tem se constituído importante instrumento para a análise histórica da ciência da informação com destaque para algumas personalidades (profissionais e pesquisadores) da área que contribuíram com a formação, constituição e desenvolvimento do campo.

O projeto teve como fontes de informação documentos pessoais e institucionais, além de entrevistas e algumas informações bibliográficas. Ao acessá-lo, tem-se inicialmente um índice, formado com nomes de 100 pioneiros apresentados em ordem alfabética de sobrenome. Ao selecionar um nome são fornecidas informações sobre: datas de nascimento e morte (se for o caso), local de trabalho, as contribuições na área e a localização de arquivos nos quais se encontram os documentos de onde as informações foram extraídas.

Moraes (2002) se dedicou ao estudo desse índice para elucidar indícios históricos da Ciência da Informação nos EUA e observou dentre outras variáveis a categorização temática dos 166 trabalhos listados como contribuição dos pioneiros. Ficou evidente no trabalho que as temáticas com abordagens técnico-científicas e tecnológicas representam 50,60% dos trabalhos, com destaque para: ‘sistemas de recuperação da informação’, ‘tecnologia da informação’, ‘administração e processos técnicos de sistemas de informação’ e ‘redes e sistemas de informação’.

Também com abordagem histórica sobre o desenvolvimento da ciência da informação Buckland e Liu (1995) fazem um artigo revisão no qual ao discorrerem sobre os inúmeros trabalhos analisados sistematizam entre os tópicos a importância e o impacto das técnicas e tecnologias para a CI na representação do documento, nas multimídias e hipermídias, nas técnicas de busca e seleção de informação, e nas aplicações tecnológicas (processamento e comunicação de dados).

A partir do traçado do domínio epistemológico da Ciência da Informação no exterior, e com abordagem historiográfica, Pinheiro (1997, 2005) apresenta o núcleo de disciplinas da área bem como suas tendências por meio de um estudo da frequência dos artigos de revisão do ARIST (1966-1995). A autora identificou 17 temas, dos quais 07 (sete) indicavam relação dialógica com a tecnologia, conforme QUADRO 1.

Disciplinas	Frequência
1. Sistemas de informação	49
2. Tecnologia da informação	36
3. Sistemas de recuperação da informação	35
4. Políticas de informação	28
5. Necessidades e usos de informação	25
6. Representação da informação	25
7. Teoria da Ciência da Informação	16
8. Formação e aspectos profissionais	16
9. Gestão da informação *	14
10. Bases de dados	14
11. Processamento automático da linguagem	11
12. Economia da informação	10
13. Bibliometria	6
14. Inteligência competitiva e Gestão do conhecimento	5
15. Mineração de dados	5
16. Comunicação científica eletrônica	3
17. Bibliotecas digitais/virtuais	2

Fonte: PINHEIRO (2005, p.7)

É possível observar no estudo de Pinheiro (2005) que dentre as disciplinas com maior ocorrência, lideram, figurando nos três primeiros lugares, com frequências 49, 36 e 35 respectivamente, temáticas tecnológicas, como ‘sistemas de informação’, ‘tecnologia da informação’ e ‘sistemas de recuperação da informação’.

Essa presença marcante das temáticas tecnológicas na ciência da informação reforça definições sobre o campo, como a de Foskett (1980) que evidencia sua configuração interdisciplinar como área que “... surge de uma fertilização cruzada de ideias que incluem a velha arte da biblioteconomia, [e] a nova área da computação”, e cuja sua forma moderna relaciona-se diretamente com “todos os problemas da comunicação – a transferência da informação”.

O mesmo pode ser observado em Le Coadic (2004) que considera que o próprio desenvolvimento da ciência da informação é seguido, ou mesmo precedido, pelo desenvolvimento excepcional das técnicas e tecnologias, ou ainda nas ideias de Saracevic (1999), que ao discorrer sobre a interdisciplinaridade da Ciência da Informação com outros

campos do conhecimento, destaca sua proximidade com a área de Ciência da Computação sugerindo que há uma relação especial, entre ambas as áreas, a qual se apresenta de forma significativa e desenvolvida.

O prestigiado ARIST encerrou seus 45 anos de publicações em 2011. Para Bawden (2010) ele representa o principal fórum para artigos de revisão em ciência da informação. E de acordo com Hjørland (2000) ele pode ser considerado a fonte de informação mais importante sobre o estado da arte da CI sendo referência para sua comunidade científica com discussões que apresentam de forma densa a configuração do campo proporcionando uma visão analítica de seu desenvolvimento bem como suas tendências.

E sua pesquisa Ladeira (2010, p.22) considerou o conhecimento acumulado no ARIST sobre ‘processamento de linguagem natural’ como referência para selecionar e analisar as publicações nacionais. Como fonte, o ARIST serviu para a autora de respaldo e garantia literária, e foi escolhido “dada a sua importância no panorama da ciência da informação no Brasil e no mundo”. Ainda segundo a autora o ARIST procura apresentar ao leitor uma revisão geral, analítica, acessível e com autoridade das tendências e desenvolvimentos significativos nas áreas de interesse da ciência da informação. Os tópicos abordados variam de ano para ano, refletindo o dinamismo da disciplina e a diversidade das perspectivas teóricas da ciência e da tecnologia de informação. Apesar de alguns tópicos clássicos continuarem em evidência (bibliometria, recuperação de informação), o ARIST tem ampliado a sua abrangência com o intuito de conectar a ciência da informação a outras comunidades acadêmicas e profissionais (LADEIRA, 2010, p.22).

Rabello (2013, p.154) considerou o ARIST como fonte de pesquisa para analisar aspectos da literatura internacional da área de CI relacionados aos temas ‘necessidade, busca, comportamento e uso da informação’, tendo como objetivo explorar a trajetória histórica e as interpretações sobre os conceitos usuário e uso de informação. A escolha do autor por esta fonte se deu por considerá-la “uma importante referência internacional para a área de informação, com revisões de literatura que trazem um panorama dos assuntos debatidos em CI e em tecnologia”.

Tendo em vista a relevância das revisões do ARIST para a CI, bem como o papel da tecnologia no percurso histórico e desenvolvimento do campo, a presente pesquisa pretende analisar o percurso da temática tecnológica nessa publicação, procurando traçar um panorama dos estudos sobre o componente tecnológico na área.

3 PROCEDIMENTOS METODOLÓGICOS

O estudo faz parte da pesquisa de doutorado em andamento, que tem como objetivo geral investigar ‘As bases tecnológicas da Biblioteconomia e Ciência da Informação no Brasil’, analisando a tecnologia como assunto de pesquisa e conteúdo formativo. A presente comunicação está direcionada para a representatividade da temática no ARIST (como referência para selecionar e analisar a produção científica da comunidade acadêmica nacional da área) por meio de uma análise documental com abordagem quantitativa e descritiva.

A coleta de dados se deu por meio de pesquisa junto ao diretório de busca do website do ARIST <<http://www.asis.org/Publications/ARIST/search.php>> em consulta manual por meio de descritores para o componente tecnológico. Segundo Costa e Moura (2013) termos, conceitos ou expressões utilizados como palavras-chave ou descritores são signos que estão em relação com o real e visam representar, ainda que parcialmente, o objeto que se estuda.

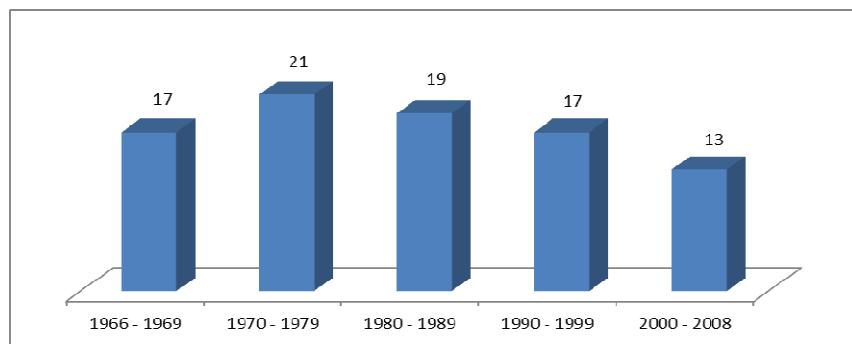
Considerando essa representação elegeu-se os descritores ‘informatics’ (11), ‘infomation technology’ (45) e ‘information systems’(44) em busca realizada em novembro de 2013. Obteve-se um total de 100 trabalhos de revisão, os quais, após leitura inicial e verificação de duplicidades foram reduzidos a 87, que por sua vez, forneceram um total de 245 descritores, que seguindo o mesmo princípio de eliminação por duplicidades, foram reduzidos para 135.

A análise documental com abordagem no conteúdo segundo Fox (2005) pode ser considerada como um processo comunicativo no qual um conjunto de operações cognitivas de natureza analítico-sintética transforma, por meio do reconhecimento e da representação do conteúdo, fornecendo especial apoio à pesquisa científica. Com base nessa premissa procedeu-se a leitura dos artigos considerando para análise os títulos e as palavras-chave das publicações. Os artigos de revisão são apresentados na sua distribuição por década de publicação, frequência de termos e clusterização temática.

4 RESULTADOS

Os 87 trabalhos de revisão recuperados foram publicados no ARIST entre os anos de 1966 e 2008. O GRÁFICO 1 mostra a quantidade de trabalhos por períodos, e com base nele é possível perceber que a década de 1970 apresentou o maior número de trabalhos (21), seguida da década de 1980 (19), tendo as décadas de 1960 e 1990 com o mesmo quantitativo (17) e os anos 2000 com menor quantidade (13).

GRÁFICO 1. Distribuição de publicação sobre tecnologia (1966-2008)



Fonte: elaboração dos autores

Ao analisar o gráfico percebemos um aumento de trabalhos da década de 1960 para a década de 1970, contudo percebe-se que nas décadas seguintes não há um crescimento da temática tecnológica e sim uma redução.

O papel da tecnologia na formação e desenvolvimento do campo da CI é fundamental à recuperação automática da informação científica e tecnológica, denominada inicialmente, como o primeiro núcleo de pesquisa em Ciência da Informação (ALMEIDA, BASTOS e BITTENCOURT, 2007; SARACEVIC, 1996). Considerou-se que a redução gradativa no número de trabalhos pode ser decorrente da evolução terminológica, de mudanças de enfoque, de especialização ou mesmo da fragmentação desse núcleo.

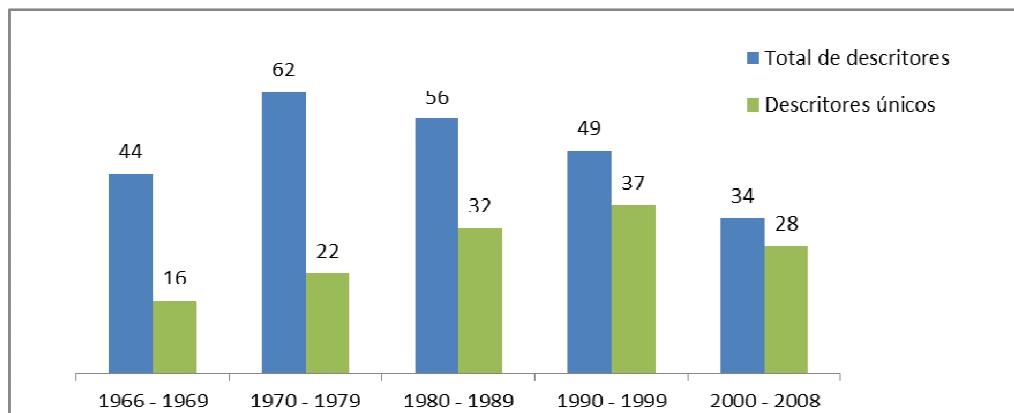
Uma constatação que corrobora com essa vertente é que nos anos de 2009 e 2010, não obteve-se nenhum trabalho recuperado por meio dos descritores pesquisados, no entanto, foram identificados trabalhos com temática tecnológica. Em 2009, por exemplo, temos um artigo de revisão sobre os princípios da ‘Arquitetura de informação’ (JACOB e LOEHRLEIN, 2009) evidenciando a repercussão do tema em outras áreas do conhecimento; e outro que aborda a ‘Classificação semântica colaborativa e os sistemas de notação’, adentrando sobre os recursos da web 2.0 e a etiquetagem social (*social tagging*) como prática da folksonomia (HUNTER, 2009).

Na edição do ARIST de 2010, tem-se um artigo de revisão com ampla discussão em torno dos 50 anos de pesquisa e desenvolvimento sobre Inteligência Artificial (EKBIA, 2010); e ainda um trabalho sobre as aplicações de tecnologias de informação e comunicação no governo eletrônico (ROBERTSON e VATRAPU, 2010).

A análise quantitativa dos descritores também pode contribuir na constatação anterior quanto à fragmentação desse núcleo. Os 87 trabalhos reuniram um total de 245

descritores, mas devido a recorrência de termos, chegou-se a um total de 135 descritores únicos. A distribuição do quantitativo dos descritores por período pode ser observado no gráfico 2, que apresenta o quantitativo geral, com o total de descritores que os trabalhos reuniram por período, e os descritores únicos, que geralmente possuem menor valor por conter termos recorrentes.

GRÁFICO 2. Distribuição dos descritores (1966-2008)



Fonte: elaboração dos autores

A diferença entre a quantidade do total de descritores (TD) e os descritores únicos (DU), equivale à dispersão (D), sendo $TD - DU = D$. Assim, quanto maior o valor de D menor será a dispersão no período, uma vez que temos mais descritores com alta recorrência. Sendo assim – conforme já observado – embora, tenha-se uma redução da década de 1970 para a década de 1990 no quantitativo de artigos publicados, quando observa-se a configuração dos descritores percebe-se o comportamento oposto. Isto é, de crescimento, de evolução temática com surgimento de muitos novos termos. Com isso, a década de 1970 apresenta a menor dispersão ($D = 40$), seguida das décadas de 1980 ($D = 24$) e 1990 ($D = 12$); e dos anos 2000 ($D = 6$) que apresentou a maior dispersão na temática no período analisado.

Ao verificar os descritores recorrentes é possível perceber por período quais termos constituem o núcleo dos trabalhos e ao mesmo tempo comprovar em alguns casos a dispersão temática na qual os trabalhos se encontram. As tabelas 1 e 2 apresentam a distribuição dos termos recorrentes nas décadas de 1960 e 1970, respectivamente.

A década de 1960 apresentou o menor índice de descritores com apenas uma ocorrência, foram 07 (sete). Os artigos de revisão listados nessa década indicam as primeiras questões abordadas pelos teóricos da CI com foco em temáticas tecnológicas, sendo por exemplo, a década que reúne maior reflexão sobre o componente tecnológico e os aspectos profissionais (*Information professional*) da ciência da informação e tecnologia

(TAYLOR, 1966; HARVEY, 1967; ATHERTON e GREER, 1968; SHERA e MCFARLAND, 1969).

TABELA 1 - Descritores da década de 1960

#	Descriptor	Frequência
1	Information technology	12
2	Information professionals	4
3	Information Storage and Retrieval	4
4	Storage and Retrieval systems	4
5	Design	3
6	Evaluation	3
7	Information science	3
8	Biomedical information	2
9	Information systems	2
10	Descritores com 1 ocorrência	7
	Total	44

Fonte: elaboração dos autores

TABELA 2 – Descritores da década de 1970

#	Descriptor	Frequência
1	Information technology	11
2	Information storage and retrieval	10
3	Storage and retrieval systems	9
4	Information systems	5
5	Design	4
6	Evaluation	4
7	Communication	2
8	Computer technology	2
9	information services	2
10	Descritores com 1 ocorrência	13
	Total	62

Fonte: elaboração dos autores

Com destaque para o descritor ‘tecnologia da informação’ (*Information Technology* - 12), a década também apresenta os primeiros trabalhos sobre as atividades de armazenamento e recuperação da informação (*Information Storage and Retrieval* - 4), bem como as aplicações tecnológicas dos sistemas nessas atividades.

A preocupação com o crescimento exponencial da informação e com seus estoques coloca a tecnologia como fundamento da relação entre CI e Ciência da Computação. Nessa relação reside a aplicação dos computadores e da computação no armazenamento e recuperação da informação, assim como nos produtos, serviços e redes associados (SARACEVIC, 1996).

A recuperação da informação tornou-se a grande questão de pesquisa da CI. Para qual se tem apresentado soluções bem sucedidas e para isso está em processo de desenvolvimento até os dias atuais. De acordo com Saracevic (1996, p.45) o trabalho determinado pela necessidade de recuperar informações suscitou questões e promoveu pesquisas exploratórias de fenômenos, processos e variáveis, bem como das causas, efeitos, comportamentos e manifestações relacionados.

As décadas de 1960 e 1970 podem ser consideradas o período de concepção e desenvolvimento dos sistemas de informação, os descritores ‘design’ e ‘evaluation’ confirmam isso. Há um crescimento de trabalhos sobre o armazenamento e recuperação da informação, bem como os sistemas de armazenamento e recuperação da década de 1960

para a década de 1970, firmando essa última com o maior número de artigos de revisão sobre o assunto.

Para estudiosos da recuperação como Lesk (1996) a década de 1960 é vista como o período da adolescência para o desenvolvimento da temática. Segundo o autor é nela que se tem a criação e a experimentação dos primeiros sistemas de pesquisa e sistemas comerciais, bem como o desenvolvimento de tecnologias para avaliação de sistemas de recuperação. O período é considerado “o *boom* da Recuperação da Informação” (ORTEGA, 2002, p.81).

Segundo Lesk (1996) é na década de 1970 que temos a fase adulta nos estudos de recuperação da informação e os sistemas passam a ser utilizados em diversos campos científicos, além do desenvolvimento e distribuição dos sistemas comerciais em larga escala. Aprimoram-se as formas de pesquisa, a recuperação torna-se mais prática e os sistemas desenvolvidos possibilitam ofertas de serviços de informação. Embora “a maioria desses sistemas operasse sem resumos e índices”, vale lembrar que essa década viu “o início dos sistemas de recuperação em texto integral” (ORTEGA, 2002, p.84).

A grande diferença entre os descritores da década de 60 e 70 é que muitos da primeira década desaparecem na segunda. E os novos descritores presentes na década de 1970 demonstram diversificação nas atividades de CI que se ligam à tecnologia. As tabelas 3 e 4 dão continuidade à discussão e apresentam a distribuição dos descritores recorrentes das décadas de 1980 e 1990, respectivamente.

TABELA 3 - Descritores da década de 1980

# Descritor	Frequência
1 Information technology	9
2 Information storage and retrieval	6
3 Storage and retrieval systems	6
4 Information services	5
5 Information systems	2
6 Legal information	2
7 Descritores com 1 ocorrência	26
Total	56

Fonte: elaboração dos autores

TABELA 4 - Descritores da década de 1990

# Descritor	Frequência
1 Information technology	9
2 Information systems	4
3 Information services	2
4 Descritores com 1 ocorrência	34
Total	49

Fonte: elaboração dos autores

Na década de 1980 há uma pequena redução nas pesquisas sobre armazenamento e recuperação da informação e nos sistemas de recuperação, mas há aumento dos trabalhos sobre ‘serviços de informação’. A dispersão temática dobra de 13 descritores com apenas uma ocorrência na década anterior para 26 descritores. Para Lesk (1996) essa década marca a fase de maturidade nos estudos de recuperação da informação, com o crescimento dos

processadores de texto e barateamento do espaço em disco significava ter mais informação disponível na forma legível por computador.

O aumento de recorrência aos serviços de informação pode estar pautado no crescimento do interesse de novas métodos de recuperação da informação (como pesquisa sobre desambiguidade), na difusão do CD-ROM, que atacava o paradigma de publicação tradicional, e no fato que o uso da recuperação da informação *online* expandiu-se, tanto na disponibilização do texto integral – e não somente resumos e índices – quanto, na propagação desse serviço por não especialistas com a chegada dos Catálogos Online de Acesso Público – OPACs (ORTEGA, 2002).

O armazenamento e recuperação da informação e os sistemas de recuperação praticamente desapareceram na década de 1990, sendo ainda a década com maior dispersão da temática com 34 descritores com apenas uma ocorrência. Para Ortega (2002, p.84) “com crescimento e massificação dos sistemas comerciais de bases de dados, houve um recuo na investigação no campo da Recuperação da Informação” e uma grande dispersão na performance dos sistemas de recuperação. Não é por menos que Lesk (1996) considera os anos 90 como a “crise de meia idade” nos estudos de recuperação da informação.

A Tabela 5 apresenta a distribuição dos termos recorrentes na década de 2000. Observa-se, por um lado, uma redução significativa dos descritores únicos, e por outro, um aumento considerável de descritores únicos, revelando assim uma grande dispersão.

TABELA 5 - Descritores da década de 2000

#	Descriptor	Frequência
1	Information technology	4
2	Medical informatics	3
3	Information services	2
4	Descritores com 1 ocorrência	25
	Total	34

Fonte: elaboração dos autores

De acordo com Ortega (2002, p. 87) os anos 2000 marcam “a utilização universal da internet nos meios acadêmicos e empresariais nos Estados Unidos”. A quantidade disponível, a facilidade de acesso e uso da informação *online* é berço de mais um movimento de evolução terminológica (25 descritores com 1 ocorrência) no campo da CI e reforça ainda relações com outras disciplinas, como é o caso da Informática Médica.

Ao analisar os pontos de convergência e divergência entre os campos da ciência da informação e sistemas de informação a partir de descritores de artigos de periódicos de

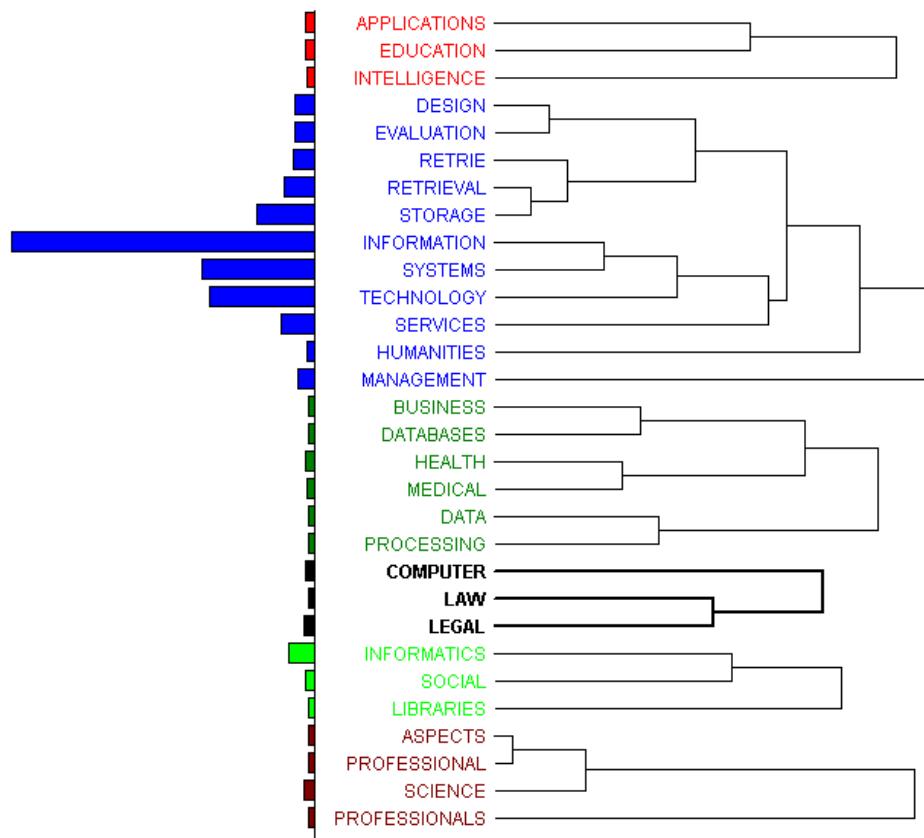
ambas as disciplinas Monarch (2000) apontou a informática médica como disciplina híbrida de convergência entre esses campos.

Para a clusterização analisou-se os títulos e as palavras-chave e utilizamos o *WordStat*¹ para calcular os termos mais incidentes com frequência da ocorrência e quais termos estão mais relacionados. Foram considerados para representação os termos com incidência maior ou igual a 2.

A importância de se conhecer os termos de maior ocorrência, que representam o campo nocional em torno das disciplinas analisadas, está no fato de ser possível por meio delas ter condições de saber sobre quais os elementos mais discutidos.

A partir da FIGURA 1, com apresentação do dendograma, podemos perceber a relação entre esses termos. Essa representação permite avaliar qual o número de *clusters* (temáticos) e sua aproximação e distância por meio de agrupamentos.

FIGURA 1. Dendograma “clusters” dos termos mais recorrentes



O maior agrupamento (na cor azul) se dá em torno dos termos ‘informação’, ‘sistemas’ e ‘tecnologia’, que por sua vez, aparecem com menor grau de distância o que

1 Software de análise de conteúdo e mineração de textos. Versão gratuita de teste disponível em: <<http://provalisresearch.com/products/content-analysis-software/>>.

implica na indicação ‘sistemas de informação’ e ‘tecnologia da informação’ como temáticas centrais. O resultado é similar ao encontrado por Moraes (2002) quando analisou 166 trabalhos listados como contribuição dos pioneiros da CI nos Estados Unidos, com destaque também para os ‘sistemas de informação’.

Estão presentes nesse mesmo agrupamento o ‘design’ e ‘evolução’ dos sistemas de recuperação da informação, os ‘serviços de informação’ e a ‘gestão da informação’, tendo ainda trabalhos sobre aplicações dos sistemas de informação na e para as ‘Ciências Humanas’.

Os demais agrupamentos vão indicar alguns focos das pesquisas, que vão dos aspectos profissionais (na cor marrom) aos educacionais (na cor vermelha), passando por questões interdisciplinares com outros campos, como o do Direito, com trabalhos sobre aspectos legais da informática (na cor preta), ou da Saúde (verde escuro) reunindo trabalhos sobre processamento de dados e bases de dados para negócio. E trabalhos com abordagens do componente tecnológico no contexto da Informática Social (na cor verde claro).

5 CONCLUSÕES

A reflexão sobre a evolução dos aparatos tecnológicos e suas implicações para a Ciência da Informação no campo científico nos artigos de revisão do ARIST leva à percepção de um vasto conteúdo que perpassa a evolução dos suportes e dos dispositivos de armazenamento; os processos que envolvem a automação, informatização e digitalização; as bases e bancos de dados e sistemas de informações; os processos de tratamento e recuperação da informação; aos produtos e serviços de informação no contexto digital (bibliotecas eletrônicas e digitais, catálogos automatizados e online; repositórios digitais; bibliotecas 2.0); dentre outros.

Essa proliferação que a temática tecnológica tem no campo da CI é expressa em dois movimentos percebidos na análise dos artigos de revisão: (a) há uma redução no quantitativo de trabalhos ao longo das décadas; e por outro lado, (b) há um aumento considerável no quantitativo de descritores únicos.

Inclusive os descritores escolhidos “Informatics”, “Information Technology” e “Information Systems” já não foram suficientes para representar os trabalhos que versam sobre tecnologia e foram publicados nas edições de 2009 e 2010 do ARIST. O que nos deixa alguns questionamentos sobre qual seria o limite para uma fragmentação temática ou mesmo até que ponto uma contínua dispersão pode indicar baixa consolidação temática e terminológica, sobretudo no contexto dinâmico da tecnologia.

Ao alertar sobre os perigos do determinismo tecnológico na ciência Morin (2002) chamou de “tecnologização da epistemologia” a crítica feita ao predomínio técnico na formação do cidadão.

Ideias do autor sustentam concepções sobre “mecanização do conhecimento” ou mesmo “tecnologização do currículo escolar” (LIMA JUNIOR, 2003) que situa o processo histórico do uso da tecnologia na educação à logica desenvolvimentista, do modelo econômico em curso no qual a tecnologia educacional, por exemplo, é encarada como parte do modelo tecnicista da educação (GIANOLLA, 2006) cujas questões centrais não estão os rumos a serem seguidos, “mas tão somente, os métodos e técnicas a serem utilizados” (MORAES, 1996, p. 47).

Os trabalhos encontrados voltados para a reflexão sobre as aplicações dos sistemas de recuperação da informação na e para as Ciências Humanas (RABEN e WIDMANN, 1972; RABEN e BURTON, 1981; TIBBO, 1991) bem como os desenvolvidos no âmbito da Informática Social (BISHOP e STARR, 1996; SAWYER, S.; ESCHENFELDER, 2002), podem contribuir com a não reprodução desse modelo tecnicista de perceber a tecnologia no campo da CI.

REFERÊNCIAS

- ATHERTON, P.; GREER, R.. Professional aspects of information science and technology. **Annual Review of Information Science and Technology** – ARIST, v.3, p. 329-355, 1968.
- Bawden, David. Alas poor ARIST: reviewing the information sciences. **Journal of Documentation**, v. 66, n. 5, p.625-626, 2010.
- BISHOP, A. P.; STARR, S. L. Social informatics of digital library use and infrastructure. **Annual Review of Information Science and Technology** – ARIST, v. 31, p.301-401, 1996.
- CAPURRO, R. Epistemologia e ciência da informação. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO - ENANCIB, 5., 2003. Belo Horizonte. **Anais...** Belo Horizonte: UFMG, 2003.
- COSTA, M. U. P.; Moura, Maria Aparecida . A Representação da informação em contextos de comunicação científica: a elaboração de resumos e palavras-chave pelo pesquisador-autor. **Informação & Informação** (UEL. Online), v. 18, p. 45-67, 2013
- GIANOLLA, R. **Informática na educação:** representações sociais do cotidiano. 3ed. São Paulo: Cortez, 2006
- EKBIA, H. Fifty Years of Research in Artificial Intelligence. **Annual Review of Information Science and Technology** – ARIST, v. 44, 2010.
- HARVEY, J. F. Professional aspects of information science and technology. **Annual Review of Information Science and Technology** – ARIST, v. 2, p. 419-444, 1967.

HJORLAND, B.. Letter to the editor: relevance research, the missing perspective(s). **Journal of the American Society for Information Science**, v. 51, n. 7, p. -211, maio 2000.

HUNTER, J. Collaborative Semantic Tagging and Annotation Systems. **Annual Review of Information Science and Technology** – ARIST, v. 43, p. 187-239, 2009.

JACOB, E.; LOEHRLEIN, A.. Information Architecture. **Annual Review of Information Science and Technology** – ARIST, v. 43, p. 147-186, 2009.

LESK, M. **The seven ages of information retrieval**. Ottawa: IFLA, 1996. Disponível em: <<http://www.ftsm.ukm.my/ss/TP6082/seven%20ages%20ir.pdf>>. Acesso em: 3 mar. 2013.

LADEIRA, Ana Paula. **Processamento de linguagem natural**: caracterização da produção científica dos pesquisadores brasileiros. Tese (Doutorado). Ciência da Informação. Universidade Federal de Minas Gerais. Escola de Ciência da Infomação. 2010. 259f.

LIMA JR. A. S. de.. **Tecnologização do Currículo Escolar**: Um possível significado Salvador, Universidade Federal da Bahia, 2003.

MONARCH, I.A. Information science and information systems: converging or diverging? In: 28th Annual Conference, Canadian Association for Information Science, CAIS: 2000, Dimensions of a Global Information Science. **Proceedings...** 2000. Disponível em: <http://www.cais-acsi.ca/proceedings/2000/monarch_2000.pdf> Acesso em 01 jun. 2013.

MORAES, R. A. A.. A análise histórica da política de informática na educação brasileira e as principais experiências conduzidas na rede pública de ensino de 1º e 2º graus até 1995. (Tese de Doutorado). Campinas, FE/UNICAMP, 1996.

MORAES, A. F. . Os pioneiros da Ciência da Informação nos EUA.. **Informação & Sociedade Estudos**, João Pessoa, PB, v. 12, n. 2, p. 101-124, 2002.

MORIN, E.. **Ciência com consciência**. Rio de Janeiro, Bertrand Brasil, 2002.

OLIVEIRA, M. (Org.). Ciência da Informação e Biblioteconomia: novos conteúdos e espaços de atuação. Belo Horizonte: Editora UFMG, 2005.

ORTEGA, Cristina Dotta. **Informática documentária**: estado da arte. Dissertação (Mestrado Ciências da Comunicação) - Universidade de São Paulo – Escola de Comunicação e Artes, São Paulo, 2002. 259f.

PINHEIRO, Lena Vania Ribeiro. Gênese da Ciência da Informação ou sinais anunciantes da nova área. In: AQUINO, Miriam de Albuquerque de (Org.). **O campo de Ciência de Informação: gênese, conexões e especificidades**. João Pessoa: Editora Universitária, 2002. p. 61-86

PINHEIRO, Lena Vania Ribeiro. Evolução e tendências da ciência da informação, no exterior e no brasil: quadro comparativo a partir de pesquisas históricas e empíricas. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO (ENANCIB), 6., 2005, 3, Florianópolis, SC. **Anais...** Florianópolis, SC, 2005.

PINHEIRO, Lena Vania Ribeiro. Ciência da Informação: desdobramentos disciplinares, interdisciplinaridade e transdisciplinaridade. In: GONZÁLEZ DE GÓMEZ, Maria Nélida; ORRICO, Evelyn Goyannes Dill. (Org.). **Políticas de memória e informação**: reflexos na

organização do conhecimento. Natal: Editora Universitária da UFRN/EDUFRN, 2006, v. , p. 111-141.

RABEN, J.; WIDMANN, R. L. Information systems applications in the humanities. **Annual Review of Information Science and Technology** – ARIST, v. 7, p. 439-469, 1972.

RABEN, J.; BURTON, S. K. Information systems and services in the arts and humanities. **Annual Review of Information Science and Technology** – ARIST, v. 16, p. 247-266, 1981.

ROBERTSON, S. P.; VATRAPU, R.. Digital government. **Annual Review of Information Science and Technology** – ARIST, v.44, p. 317-364, 2010.

SANTOS, P. L. V. A. *et al.* Mapeamento do termo tecnologia em periódicos da ci no escopo do GT- Informação e Tecnologia. In: XIV Encontro Nacional de Pesquisa em Ciência da Informação - ENANCIB 2013, 2013, Florianópolis. **Anais ...** Florianópolis: Universidade Federal de Santa Catarina - UFSC; ANCIB, 2013. v. 14. p. 1-19

SARACEVIC, Tefko. Ciência da informação: origem, evolução e relações. **Perspec. Ci. Inf.**, Belo Horizonte, v. 1, n. 1, p. 41-62, jan./jun. 1996. Disponível em: <<http://portaldeperiodicos.eci.ufmg.br/index.php/pci/article/view/235/22>>. Acesso em: 03 fev. 2012.

SAWYER, S.; ESCHENFELDER, K. R. Social informatics: Perspectives, examples, and trends social informatics. **Annual Review of Information Science and Technology** – ARIST, v.36, p. 427-465, 2002.

SHERA, J. H.; MCFARLAND, A. Professional aspects of information science and technology. **Annual Review of Information Science and Technology** – ARIST, v. 4, p. 439-471, 1969.

SILVA, Armando Malheiro da. Ciência da Informação e Sistemas de Informação: (re)exame de uma relação disciplinar. **Prisma.com**, n. 11, p. 2-30, jul., 2010. Disponível em: <http://prisma.cetac.up.pt/2_Ciencia_da_Informacao_Sistemas_de_Informacao_reexame_de uma_relacao_disciplinar_Armando_Malheiro.pdf>. Acesso em 14 mai. 2012.

TAYLOR, R. S. Professional aspects of information science and technology. **Annual Review of Information Science and Technology** – ARIST, v. 1, p. 15-40, 1966.

TIBBO, H. R. Information systems, services, and technology for the humanities humanities. **Annual Review of Information Science and Technology** – ARIST, v. 26, p. 287-346, 1991.

VIDOTTI, Silvana Aparecida Borsetti Gregorio ; OLIVEIRA, H. P. C. ; LIMA, I. F. Ensino de Tecnologias de Informação e Comunicação: um diagnóstico nos cursos de graduação em Biblioteconomia, Arquivologia e Museologia do Brasil. In: XIV Encontro Nacional de Pesquisa em Ciência da Informação - ENANCIB 2013, 2013, Florianópolis. **Anais...** Florianópolis: Universidade Federal de Santa Catarina - UFSC; ANCIB, 2013. v. 14. p. 1-19.

IMAGEM & TECNOLOGIA EM CATÁLOGOS DE MUSEUS DE ARTE: INTERSEMOIOSE EM QUESTÃO

IMAGE & TECHNOLOGY IN ART MUSEUM CATALOGS: INTERSEMIOSIS IN QUESTION

Fábio Rogério Batista Lima
Plácida Leopoldina Ventura Amorim da Costa Santos

Resumo: A digitalização e a transposição das obras de arte para catálogos impressos e digitais nos *webmuseus* criam simulacros das obras originais, e sua forma material e orgânica transforma-se em linguagem computacional de *bits* e *bytes*. Identificam-se as características que indicam as transformações da estrutura de apresentação da obra de arte na sua passagem para ambientes virtuais quando digitalizadas, uma vez que a mudança de um sistema semiótico para outro ocasiona rupturas e mutações na qualidade estética da obra de arte produzida sobre um suporte e transferida para outro. Com uso do método bibliográfico descritivo, verificaram-se as características dos catálogos de obras de arte, tanto os impressos quanto os da *Web*, demonstrando suas semelhanças e diferenças na apresentação das informações das obras. Os resultados da análise revelaram que a digitalização das obras de arte e sua transposição para ambientes digitais na construção de catálogos *on-line* e impressos acarretam perdas informacionais referentes aos elementos essenciais da composição plástica do recurso. A transformação mais evidente é a diferença de cores ocorrida tanto nos catálogos impressos e a obra (levando em consideração os tipos de materiais utilizados) quanto nos catálogos impressos e na *Web*, pois a cor luz é diferente da cor pigmento. Ao tratar das obras tridimensionais, como as obras escultóricas, a digitalização causa a perda de uma dimensão, o que deixa a imagem achatada e bidimensional. Os catálogos digitais *on-line* perdem em intensidade, mas ganham em extensividade, pois aproximam as pessoas, ao recontextualizarem as obras e servirem como guia e tradutor do acervo, levando o indivíduo a percorrer a exposição com um conjunto de informações que atuam como apoio didático-pedagógico no processo de aquisição de repertório e ampliação da base de conhecimento.

Palavras-chave: Imagem. Informação e Tecnologia. Intersemioses Digitais. Catálogos de museu de arte.

Abstract: The Digitization and transposition of artwork for print and digital catalogs in webmuseums create simulacra of the original works, and its material and organic form becomes computational language of bits and bytes. The characteristics that indicate the transformation in the structure of the presentation of the artwork in its passage to virtual environments when digitization is performed are identified. It is justified since the change from one semiotic system to another causes disruptions and mutations in the aesthetic quality of the artwork produced on a support and transferred to another. With a descriptive bibliographic method, the characteristics of artwork in catalogs is verified, both print and on the web, disclosing their similarities and differences in the presentation of artwork information. The results of the analysis revealed that the digitization of artwork and its transposition into digital environments in the construction of on-line and printed catalogs lead to informational losses related to the essential elements of artistic composition of the resource. The most obvious transformation is the color difference occurred both in printed catalogs and the artwork (taking into consideration the types of materials used) both in printed catalogs and on the web, as the light color is different from the color pigment. When dealing with three-dimensional artwork such as sculptural works, digitization causes the loss of a dimension, which turns the artwork into a flat, two-dimensional image. Digital

online catalogs lose in intensity, but gain in extensiveness, as bring people closer; they recontextualize the artworks and serve as translator and guide to the collection, leading the individual to go through the exhibition with a collection of information that serve as didactic support to the acquisition process of repertoire and broadening of knowledge base.

Keywords: Image. Information and Technology. Digital intersemioses. Art Museum Catalogs.

1 INTRODUÇÃO

Refletir sobre a atualidade é também enfatizar questões importantes referentes à arte, à ciência e à tecnologia, o que nos remete a questões como globalização, crescimento informacional, velocidade e virtualidade, situações relacionadas à evolução do homem e sua *Teknè*. O homem, enquanto agente dessa evolução, tem como uma de suas atividades a arte, uma manifestação de ordem estética resultante da percepção, das emoções e das ideias, a partir da utilização de uma grande variedade de meios e materiais, como a pintura, a escrita, a música, a dança, a fotografia, a escultura, entre outros.

A arte influenciou a ciência e foi por ela também muito influenciada. Entretanto, na contemporaneidade, é a ciência, juntamente com a tecnologia, que vem imprimindo significativamente suas marcas sobre a arte. Os “[...] materiais e métodos podem ser diferentes, mas as estimulações cruzadas são frequentes e férteis para a arte, a ciência e a tecnologia.” (CZEGLEDY, 2003, p. 126).

Desde o surgimento das principais sociedades primitivas, passando pelas complexas cidades pós-industriais até os dias atuais, o homem descobriu como fazer o fogo, cultivou a terra, domesticou os animais, construiu cidades, dominou os recursos energéticos naturais, implementou indústrias, conquistou o espaço cósmico, viajou pelos confins do espaço-tempo, e “[...] durante esse trajeto a tecnologia ganhou significações e representações diversas em um movimento de vai e vem com a vida social” (LEMOS, 2008, p. 25).

De acordo com Pinho (2008, p. 8), tudo o que é humano é ao mesmo tempo psíquico, técnico, histórico etc. e se integra com o que o autor chama de “equipamentos coletivos de subjetivação”, como, por exemplo, as novas tecnologias informáticas e a Internet. Com esses objetos tecnológicos, a história da comunicação entre os homens, bem como as construções de imagens e suas particularidades históricas e culturais, mudaram e continuam em constante transformação.

Nesse contexto, a proposta deste estudo foi a de identificar as características que indicam as transformações da estrutura de apresentação da obra de arte na sua transposição para ambientes virtuais quando digitalizadas, uma vez que a mudança de um sistema semiótico para outro ocasiona rupturas e mutações na qualidade estética da obra de arte –

cor, dimensão, textura etc. O objetivo foi demonstrar, com o uso do método bibliográfico descritivo, as semelhanças e as diferenças na apresentação das informações das obras e verificar as características dos catálogos.

2 IMAGEM: UMA SÍNTESE HISTÓRICA

Desde o princípio, o ser humano tem a necessidade de se comunicar com os seus semelhantes, e os meios utilizados para esse processo têm variado ao longo dos séculos. A linguagem, definida como produção e entendimento da fala, se desenvolveu com a evolução da espécie humana. A fala humana poderia ser comparada à comunicação entre outros animais, porém a linguagem humana tem elementos que a distinguem da linguagem dos animais, na medida em que possui criatividade e interpretação.

No início, quando os primeiros homens, *Homo Herectus* (que viveram entre 1,8 milhões de anos e 300.000 anos atrás), habitavam a Terra, o centro de sua fala não era muito desenvolvido e a comunicação se dava por meio de grunhidos, rosnadas e gestos. Na época dos *Neandertais* (há cerca de 300.000 anos até aproximadamente 29.000 anos), acredita-se que o hemisfério esquerdo do cérebro desenvolveu-se um pouco mais. Com o *Homo Sapiens* (há 200.000 anos), o cérebro se desenvolveu e propiciou o desenvolvimento da fala. Foi um período caracterizado pela linguagem. No processo de comunicação, o papel da linguagem é primordial, pois “[...] ela garante o intercâmbio da significação, mediando as relações entre pessoas, flexibilizando o pensamento e ampliando a capacidade conceitual e proposicional de cada indivíduo” (BARRETO, 2008, p. 114).

Acredita-se que esse desenvolvimento se deu pelo convívio simultâneo das duas culturas diferentes, *Neandertais* e *Homo Sapiens*, e pela necessidade da comunicação entre eles. Assim, outras formas de linguagens surgiram como, por exemplo, as imagens, que eram

[...] sempre modeladas por estruturas profundas, ligadas ao exercício de uma linguagem, assim como a vinculação a uma organização simbólica (a uma cultura, a uma sociedade); mas a imagem também é um meio de comunicação e de representação do mundo, que tem seu lugar em todas as sociedades humanas (AUMONT, 1993, p. 131).

Tais imagens, encontradas em sua maioria em paredes de cavernas, como as de Altamira e Lascaux, conhecidas como pinturas rupestres, eram feitas com o uso de sangue de animais, argila e seiva de plantas. Acredita-se que as sociedades primitivas se expressavam principalmente por meio dessas pinturas.

Os meios técnicos que suportam a informação têm uma relação com o espaço e com o tempo da vida social. Segundo (BARRETO, 1999),

Os homens, ao produzirem as teias de significações para si mesmos, empregam meios técnicos/suportes materiais para transmitir as formas simbólicas, fundamentando, assim, sua vida social nos aspectos da produção, armazenamento e circulação da informação e do conteúdo simbólico.

A grande variedade de pinturas rupestres, feitas por povos diferentes e em locais diferentes, mantém algumas características comuns na representação do cotidiano, de ferramentas, de armas, ou figuras que representavam situações específicas, como a caça. Embora muitas vezes as cenas de caça representadas em cavernas não descrevessem uma situação vivida pelo grupo, elas possuíam um caráter mágico, preparando o grupo para essa tarefa que lhes garantiria a sobrevivência.

Com o passar do tempo, a produção de imagens começou a ser feita por meio da utilização de variados tipos de materiais e suportes, mas ainda com o mesmo intuito: comunicar, representar e informar. Na modernidade, os modos de produção de imagem romperam definitivamente com os modos anteriores.

O desenvolvimento de aparelhos técnicos e químicos para a produção de imagem possibilitou a captura de imagens do mundo real de modo muito mais rápido. Exemplo disso é a fotografia, definida por Aumont (1993, p. 164) como “[...] ação da luz sobre certas substâncias que assim levadas a reagir quimicamente, são chamadas fotossensíveis. Uma superfície fotossensível exposta à luz será transformada provisória ou permanentemente [...]”.

A fotografia impulsionou o crescimento do uso da imagem para variados fins. Diferentemente da pintura, a fotografia passou a servir como documento comprobatório e indicial. A maneira de produzir imagens por máquinas fotográficas evoluiu com o passar do tempo e hoje temos à disposição, de modo muito acessível, máquinas digitais para a captura e produção instantânea de imagens.

Tal perspectiva nos leva a concordar com as reflexões do cientista norte-americano Negroponte (1995), quando diz que vivemos em um meio onde a tendência de nossas atividades de trabalho, nossos entretenimentos, nossas histórias e relações com outras pessoas tornam-se cada vez mais dependentes das tecnologias digitais. As imagens produzidas atualmente, na maioria das vezes, ilustram essa reflexão, pois percebemos que elas estão sendo produzidas cada vez mais com uso de equipamentos eletrônicos, seja por câmeras digitais, por aparelhos celulares, por *tablets* ou mesmo por computadores portáteis com câmeras embutidas.

O QUADRO 1 ilustra a evolução da imagem ao longo do tempo e o efeito que ela exerce sobre a sociedade e a cultura a que pertence.

QUADRO 1: Natureza da imagem, papel do receptor

NATUREZA DA IMAGEM		
PRÉ	FOTOGRÁFICA	PÓS
Figurar o visível e o invisível	Registrar o visível	Visualizar o modelável
Figurar por imitação	Capturar por conexão	Simular por variações de parâmetros
Imagen espelho	Imagen documento	Imagen matriz

PAPEL DO RECEPTOR		
PRÉ	FOTOGRÁFICA	PÓS
Contemplação	Observação	Interação
Nostalgia	Reconhecimento	Imersão
Aura	Identificação	Navegação

Fonte: Adaptação de Santaella (2001, p. 172-178)

As imagens hoje estão em todos os lugares: em placas de trânsito, *outdoors*, vitrines, bancas de jornal, logomarcas de empresas ou em livros, revistas e massivamente na rede Internet. De acordo com Aumont (1993, p. 314),

As imagens, isso é inegável, há mais de 100 anos multiplicam-se quantitativamente em proporções impressionantes e sempre crescentes. Além disso, percebemos que essas imagens invadem nossa vida cotidiana, que seu fluxo não pode ser contido.

Isso se dá devido às facilidades que temos na conexão entre computador e *scanners*, com os instrumentos de captura e restituição da imagem, e no progresso das interfaces de uso. Com a fotografia, a videoarte, o cinema e, principalmente, com a inserção de equipamentos informáticos na construção de obras digitais e também nos ambientes virtuais, como nos *webmuseus*, foi possível confeccionar uma nova proposta plástica recorrendo aos programas computacionais.

É no ambiente *Web* que está ocorrendo a fusão de linguagens e onde a arte agora está imersa. Segundo Eirin de Rapp (2000, p. 44), essa fusão “[...] se reflete nas páginas de arte, nas quais a informação estética é complementada por música, animação, conexões para outras páginas [...]”, num fluxo intersemiótico que aguça novas percepções cognitivas.

As percepções cognitivas, estimuladas pelas possibilidades oferecidas no ciberespaço, nos encaminharam para uma reflexão sobre as intersemioses digitais. Entretanto, não há como falar de intersemioses digitais sem antes nos remetermos à semiótica e à apresentação de seu conceito.

O estudo semiótico tem por objeto os signos e investiga todas as linguagens possíveis, como, por exemplo, a linguagem oral (fala), a linguagem textual (textos), a

linguagem corporal (gestos), a linguagem visual (imagens), dentre outras, "no exame dos modos de constituição de todo e qualquer fenômeno de produção de significação e sentido" (SANTAELLA, 1983, p. 13).

A semiótica, do grego *semeiotiké*, é a ciência que estuda o processo de articulação e produção de sentido entre os diversos tipos de códigos e linguagens. Tem por objetivo "estabelecer como devem ser todos os signos para uma inteligência capaz de apreender através da experiência [...]" (SILVEIRA, 2007, p. 38). Ocupa-se do estudo do processo de significação ou representação, na natureza e na cultura, dos conceitos e das ideias. Em uma acepção geral a "semiótica é a teoria de todos os tipos de signos, códigos, sinais e linguagens. Portanto, ela nos permite compreender palavras, imagens, sons em todas as suas dimensões e tipos de manifestações" (SANTAELA, 2010, p.59).

A intersemiose, por sua vez, consiste

[...] na interpretação dos signos verbais por meio de sistemas de signos não verbais', ou 'de um sistema de signo para outro, por exemplo, da arte verbal para a música, a dança, o cinema ou a pintura, ou vice-versa [...] (PLAZA, 2008, p. XI).

O processo tradutor intersemiótico, entretanto, é influenciado tanto pelos variados tipos de linguagem (sonora, visual, digital etc.), quanto pelo meio empregado na tradução de uma linguagem/mídia para a outra, por exemplo, por meio da transformação da fotografia em filme (meio cinematográfico), ocorrendo mudanças na forma de linguagem e no suporte. No entanto, ambos deixam a marca da história em seus processos e solicitam outra sintaxe e outro modo de formar sentido.

O espaço de exposição das obras na tela do computador via Internet permite que o usuário navegue contemplando as obras – muitas vezes com imagens em movimento. Ao mesmo tempo em que ouve um fundo musical, visualiza os catálogos das obras expostas e a biografia dos artistas. Esse espaço, com textos e imagens, possibilita ao visitante assistir a vídeos com as entrevistas dos curadores comentando as exposições, ou até mesmo com os próprios artistas falando sobre as obras ou a corrente artística de que fazem parte.

As tecnologias de manipulação de imagens e os novos ambientes virtuais para sua hospedagem, como os *webmuseus* e as galerias virtuais, acarretaram mudanças em nosso sistema cognitivo de percepção e recepção de informações imagéticas, "[...] criando quadros de memórias, fornecendo os símbolos e o ambiente mental, requeridos para determinados momentos de ruptura decisivos para a reestruturação dos sistemas da cultura" (JORENTE, 2009, p. 105).

Segundo Grau (2007, p.16), “[...] arte e ciência, mais uma vez, aliam-se a serviço dos métodos mais complexos de produção de imagens [...].” O mundo imagético ao nosso redor vem se transformando muito rapidamente. Ao mesmo tempo, presenciamos as várias formas de imagens e suas formas de produção pelas novas mídias, que exercem forte impacto sobre a teoria da imagem, teoria da cor e da arte como um todo, influenciando também os ambientes onde as obras são expostas, como o caso dos museus.

3 MUSEUS E SEUS CATÁLOGOS

A análise dos catálogos e das obras de arte foi realizada na Pinacoteca do Estado de São Paulo e no Museu de Arte Moderna de São Paulo – MAM-SP.

Fundada em 1905 pelo Governo do Estado de São Paulo, a Pinacoteca do Estado é um museu de artes visuais, com ênfase na produção brasileira do século XIX até a contemporaneidade.

O MAM-SP foi criado em 1948, por iniciativa do empresário Francisco Mattarazzo Sobrinho e de um grupo de artistas e intelectuais, entre eles Sergio Milliet, Tarsila do Amaral, Villanova Artigas, Antonio Cândido e Oswald de Andrade.

O estudo aqui apresentado foi realizado com a análise dos catálogos impressos e *online* de algumas obras dessas duas instituições, com a finalidade de fazer um levantamento das principais transformações estético-informacionais ocorridas nas imagens das obras transpostas em diferentes tipos de suportes, o catálogo impresso e o catálogo na *Web*.

3.1 Análise de catálogos de museu de arte

Diferentemente dos *folders*² de mão, ou também conhecidos como folhetos impressos publicitários, com poucas páginas, que servem para promover as exposições com dados informacionais sobre as obras e os artistas, o catálogo de obras de arte pode ser definido a partir da descrição de catálogo dada por Mey (1995, p.9):

[...] um canal de comunicação estruturado, que veicula mensagens contidas nos itens, e sobre os itens, de um ou vários acervos, apresentando-as sob forma codificada e organizada, agrupadas por semelhanças, aos usuários desse(s) acervo (s).

Dessa forma, o catálogo de museu é um produto informacional estruturado que atua como ligação entre o acervo e os usuários e “representa um exemplo de lista prática, que se refere a objetos existentes em um lugar determinado” (CRIPPA, 2010, p. 30). Os catálogos podem ser específicos, de determinados artistas, com fotos de suas obras, comentários feitos

² Os *folders* são impressos publicitários com diversas dobras e formatos, usados para a divulgação de eventos ou exposições, no caso dos museus. Ver imagens no apêndice.

por curadores e críticos de arte sobre suas exposições, ou ainda gerais, temáticos, cronológicos, geográficos, entre outros.

Os catálogos impressos oferecem não só a facilidade no manuseio, mas também a visualização das obras seguida de sua ficha técnica e seu contexto histórico. Eles servem como um guia e tradutor do acervo, levando o indivíduo a percorrer a exposição com um conjunto de informações que atuam como apoio didático-pedagógico no processo de aquisição de repertório e ampliação da base de conhecimento.

Entretanto, a estrutura visual e informacional de catálogos de museu de arte, impressos ou na *Web*, varia muito, e o fator financeiro tem seu peso nessa diferença, pois alguns são feitos com materiais de alta qualidade, luxuosos, e outros, não. Com os catálogos digitais na *Web*, perde-se em intensidade, na medida em que se precisa de menor esforço para mantê-los e menos dinheiro para construí-los, e ganha-se em extensividade, pois eles aproximam as pessoas, recontextualizando as obras. Alguns catálogos impressos dispõem de uma miniatura da obra – monocromática – junto com sua ficha técnica. Outros utilizam imagens maiores policromáticas, com ficha técnica e um pequeno comentário sobre a obra, como ilustra a FIG. 1.

FIGURA 1 - Catálogos impressos – MAM/SP e Pinacoteca do Estado de São Paulo



Fonte: Autor

Normalmente, as fichas técnicas das obras nos catálogos de museus de arte impressas seguem um padrão estrutural para a apresentação organizada das informações.

As fichas descritivas oferecem ao usuário informações sobre o artista da obra, o nome da obra – quando houver –, a data em que foi produzida, a técnica utilizada pelo artista, a dimensão, e também, de onde ela veio até ser registrada no museu. São dados importantes, pois é a partir dessas informações descritivas que a obra se torna um objeto único em uma instituição.

A imagem nos catálogos de museus de arte é um instrumento informacional fundamental e de extrema importância, pois é a imagem que representa a obra em sua

ausência, por isso deve ser o mais fiel possível. No entanto, nem sempre isso acontece, como ilustra a FIG. 2.

FIGURA 2 - Análise comparativa entre dois catálogos impressos³



Fonte: Autor

No exemplo da FIG. 2, percebe-se a diferença nas tonalidades entre dois catálogos impressos na representação da mesma obra. Na imagem à esquerda, a cor está próxima da tonalidade ocre, já na imagem à direita a cor está voltada para o azul violeta.

Muitas vezes, quando uma obra é fotografada para servir de representação em catálogos *on-line*, ocorrem mutações em suas cores. Isso acontece devido ao fato de que a cor luz ‘Newtoniana’ difere da cor pigmento. Nos catálogos digitais *on-line*, as cores das obras se convertem em modelos de cores *RGB* (*vermelho-Red, verde-Green, azul-Blue*).

O propósito principal do sistema *RGB* é a reprodução de cores em dispositivos eletrônicos, como monitores de TV e computadores. Esse sistema de cores segue o modelo baseado na teoria de visão colorida tricromática, de *Young-Helmholtz*.

No caso de obras tridimensionais, como em esculturas, fazendo também referência à FIG. 2, quando fotografadas sofrem perdas em sua composição espacial, pois a obra escultórica, na foto, perde uma dimensão, ficando bidimensional e achataada.

Na análise comparativa entre dois catálogos, impresso e *on-line*, da Pinacoteca do Estado de São Paulo, em que a imagem ilustrada representa a mesma obra, conforme figura 3 - *Cabeça de índio* - do pintor Henrique Bernadelli, houve diferenças na qualidade estética da imagem em relação às suas dimensões:

³ AMARAL, A. (Org.). Museu de Arte Contemporânea da Universidade de São Paulo: Perfil de um Acervo. São Paulo: Techint. 1988. 391p. MAC USP 40 anos: interfaces contemporâneas. Apresentação: Adolpho José Melfi. Introdução: Elza Adjzenberg. Texto: Elza Adjzenberg. Helouise Costa. Cristina Freire. Gabriela Wilder. Daisy Peccinini. Kátia Canton. São Paulo (SP): Museu de Arte Contemporânea – MAC, São Paulo. 2003.

FIGURA 3: catálogos impresso e *on-line* – Pinacoteca



Fonte: Autor

No catálogo *on-line*, a imagem é menor em relação à mesma imagem no catálogo impresso. Em muitos casos, quando a imagem ilustrativa da obra em catálogos é muito pequena, há o comprometimento da visibilidade dos detalhes e do conjunto plástico da obra, ofuscando muitas vezes informações visuais que são essenciais para o conjunto da obra.

O acervo *on-line* da Pinacoteca se apresenta por meio de uma lista alfabética dos artistas com nome, seguido de data de nascimento/morte e número de obras disponíveis na base de dados.

A análise semiótica de algumas obras é apresentada a seguir, como uma possibilidade metodológica para a averiguação das transformações ocorridas na passagem da reprodução da obra original para os suportes impressos e digitais.

3.2 Análise semiótica de obras musealizadas

Diante de uma obra de arte, a primeira coisa que o observador costuma fazer em sua leitura visual é enxergar os índices, ou seja, ele comprehende que se trata de uma mulher, com vestimentas negras, segurando uma carta na mão esquerda, encostada em uma janela aberta. Não existe caminho mais fácil para perder rapidamente quali-signos como esse, pois, agindo dessa forma retraímos nossas sensibilidades para com as qualidades plásticas da obra.

Para evitar esse risco, temos que olhar as obras sempre com novos olhares e tentar retardar as nomeações tipo: “isso é aquilo”, “aquilo é aquilo” etc. A recomendação diante

de uma obra de arte é que nos deixemos agir como a criança, que ainda não é capaz de reconhecer figuras, para que possamos perceber as qualidades plásticas da pintura.

Para a pintura, como objeto único, suas qualidades (quali-signos) são essenciais. Por isso a importância, quando se trabalha com análise qualitativa de obras de arte, do contato direto com a original, pois é na original que percebemos as cores, as texturas, o gesto da mão do pintor ao riscar a tela com a tinta.

A análise consiste na descrição de seus aspectos compostivos, levando em consideração: **Estrutura, Forma, Função e Conteúdo**, tendo como base teórica a semiótica aplicada.

Esses aspectos são analisados com o objetivo de estabelecer parâmetros de comparação entre a obra e suas versões impressa e digital na *Web*, apesar de concordarmos com a notificação do autor Walter Benjamin (1936), que diz a importância de estarmos presente em frente da obra original “*hic et nunc*”, para notarmos sua qualidade plástica e sua potência aurática.

Tendo isso em mente e, tendo como parâmetro a obra original, apresenta-se a análise comparativa com o intuito de pontuar as principais transformações sofridas pela obra na passagem de um suporte para o outro, como ilustra a FIG. 4, tendo como foco a representação da obra nos catálogos.

FIGURA 4 - Análise comparativa da obra Saudade - original, impressa, digital na *Web*



Fonte: Autor

Quando estamos dentro de um museu presencial (paredes), sentimos a presença do público, dos espaços arquiteturais, da iluminação, dos ruídos, da distância entre um quadro e outro, da variedade de tipos de moldura que cada quadro possui. Essas características são as que se perdem com a qualidade plástica da obra na apresentação em ambientes virtuais de exposição. Porém, são perdas compensadas, muitas vezes, com ganhos de outras ‘qualidades’, que não seriam possíveis nos ambientes tradicionais. Para ilustrar de forma sintética essas transformações, apresenta-se o QUADRO 2.

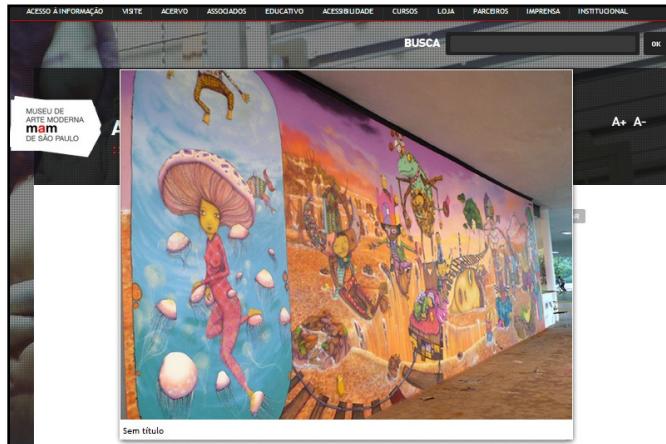
QUADRO 2 - Transformações do quadro Saudade após a digitalização e transposição para diferentes suportes

	OBRA ORIGINAL	REPRODUÇÃO IMPRESSA	REPRODUÇÃO DIGITAL – WEB
Estrutura	Tela com moldura fixada em parede	Papel (vários tipos de papéis e formatos)	Plataforma Web; suporte digital (<i>bits e bytes</i>)
Forma	Presença gestual do pintor como (sinais das pinceladas); linhas, pontos, cores, texturas, formas geométricas; imagem feita com pigmento orgânico – tinta	Perda da qualidade gestual como (sinais das pinceladas); não há textura, mudança nas cores (formato CMYK – Cyan, Magenta, Yellow and Black); redução das formas e da nitidez dos detalhes figurativos da obra	Perda da qualidade gestual como sinais das pinceladas; não há textura, mudança nas cores (formato RGB – Red, Green and Blue); redução das formas e da nitidez dos detalhes figurativos da obra
Função	Contemplativa, comunicativa, aurática	Informativa/representativa	Informativa/simulativa/interativa
Conteúdo	Plaquetas informativas; curadores	Imagen; ficha técnica; análise comentada da obra (contexto histórico)	Imagen; ficha técnica; análise comentada da obra (contexto histórico); hiperlinks

Fonte: Autor

Algumas imagens não constam do catálogo *on-line* do MAM/SP, por falta de autorização dos responsáveis legais. Contudo, há exemplos, como o caso do painel dos grafiteiros OsGêmeos, que constam do catálogo *on-line* do museu e possuem imagens coloridas. No entanto, a imagem disponível no *site* comprometeu a boa visualização das figuras que constam do painel e que não podem ser vistas pela imagem disponibilizada pelo catálogo *on-line*, como se nota na FIG. 5.

FIGURA 5 - Painel d'Os Gêmeos: catálogo *on-line* do MAM/SP



Fonte: http://www.mam.org.br/acervo_online

Partindo de uma breve análise visual e utilizando a semiótica aplicada de Santaella (2010), percebemos que partes da obra que são vistas apenas na presença da obra original se perdem na exposição disponibilizada pelo catálogo na *Web*, como mostra a FIG. 6.

FIGURA 6: Parte da obra sem nitidez na exposição pelo catálogo na *Web*

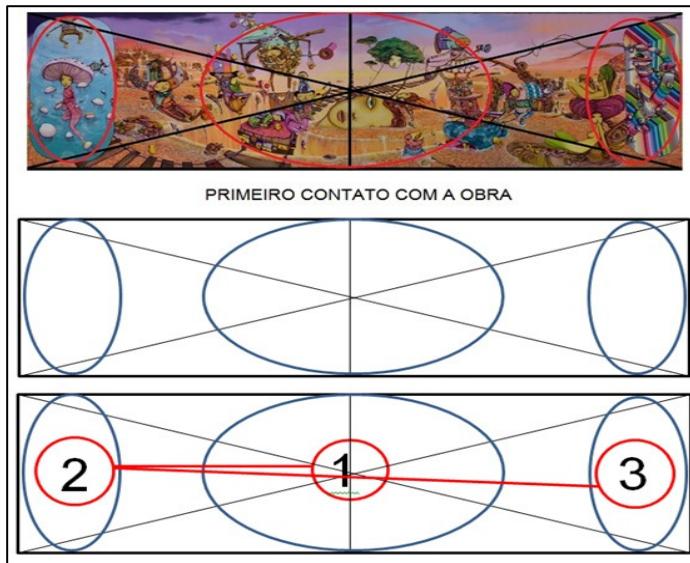


Fonte: Autor

Quando nos deparamos com uma obra de arte, nosso olhar se direciona para determinada parte dela, algo que chama a atenção, somos direcionados para um ponto da figura, de modo intuitivo ou não. Essa era uma técnica utilizada por pintores renascentistas, como pode ser visto nas obras do pintor Jan van Eyck. As obras renascentistas eram pensadas sobre estruturas visuais, feitas simetricamente no espaço da tela com a intenção de dar um equilíbrio e ao mesmo tempo direcionar o olhar do espectador para certos pontos da imagem.

Fazendo uma análise do painel d'OsGemeos, percebemos certas semelhanças com esses elementos compositivos. As extremidades, por serem simétricas, dão equilíbrio à obra. Mas isso só se percebe com a visão da obra por inteiro, ao se fazer uma leitura linear em panorâmica.

FIGURA 7: leitura visual do painel d'OsGemeos – visão panorâmica



Fonte: Autor

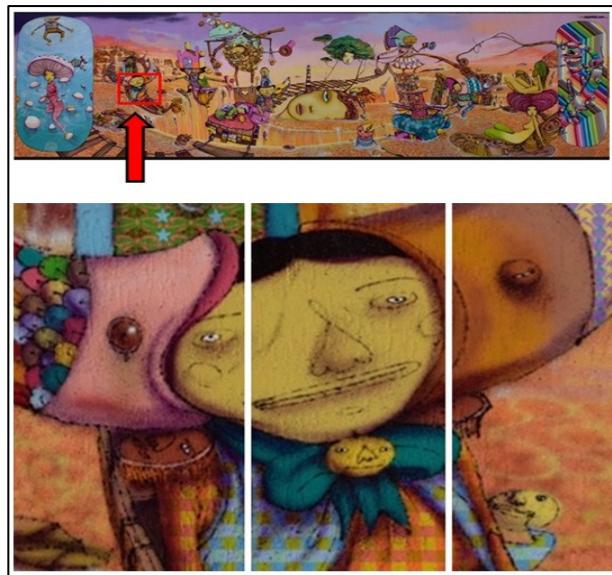
Assim como o pintor holandês Hieronymus Bosch (1450-1516), os grafiteiros fazem uso nesse painel de muitos elementos visuais (muitas figuras, detalhes, efeitos de cores, texturas etc.). Limitaremos nosso estudo, para fins didáticos, à análise sintática da imagem e faremos um pequeno recorte, decompondo-a, para análise desses elementos.

No primeiro quadrante, percebemos a riqueza nos detalhes, no uso de cores intensas e variadas: um peixe na cor rosa com a boca aberta, como se estivesse tentando engolir a cabeça de uma figura masculina, ao mesmo tempo em que nos remete a uma imagem de um ornamento, como se fosse um chapéu, usado por essa figura masculina. Na extremidade superior direita, veem-se detalhes do desenho com figuras pequenas em azul, de estrelas feitas por técnicas de estêncil (máscaras/moldes), assim como na parte inferior do quadrante percebem-se os detalhes coloridos de formas geométricas na vestimenta da figura.

No segundo quadrante, nota-se a face da figura feita na cor amarela (marca característica dos grafiteiros) e a presença de várias carinhas amarelas, que estão presentes em várias partes do quadro, em roupas, instrumentos musicais, objetos etc.

No terceiro quadrante, a presença do efeito de água e seu movimento com cores e transparências. O traço fino da ponta do *spray* que, simulando uma fina pincelada, também é presente nas obras dos grafiteiros.

FIGURA 8: Decompondo o painel para análise semiótica



Fonte: Autor

Essas são algumas das possíveis observações visuais da obra, que demonstram a marca gestual do pintor, e que se perdem na passagem para os catálogos digitais na Web. Como exemplo, observe-se a figura no catálogo do MAM-SP:

FIGURA 9: Análise comparativa entre a obra original e a imagem no catálogo na Web



Fonte: Autor

Nessa imagem comparativa, percebemos que os elementos visuais que compõem a obra, assim como a linearidade da leitura da imagem, se perdem. Isso ocorre devido à posição em que foi tirada a fotografia do painel. Dessa forma, a imagem disponibilizada no catálogo na *Web* não nos permite ver a obra como um todo, pois a visualização da parte direita do painel foi prejudicada.

No QUADRO 3, é apresentada a síntese das transformações qualitativas da obra analisada em sua transposição para o catálogo digital.

QUADRO 3 - Síntese das transformações qualitativas do painel d'OsGêmeos e sua reprodução

	OBRA ORIGINAL	REPRODUÇÃO DIGITAL –WEB
Estrutura	Parede, pigmentos orgânicos (tintas)	Plataforma <i>Web</i> ; suporte digital (<i>bits e bytes</i>).
Forma	Presença gestual do pintor, como efeitos visuais (efeitos com máscaras e sprays – estêncil), linhas, pontos, cores, texturas, formas geométricas	Perda da qualidade gestual do pintor, como efeitos com máscaras e sprays – estêncil, distorções nas formas e figuras, mudança nas cores (formato RGB – Red, Green and Blue); redução das formas e da nitidez dos detalhes figurativos da obra
Função	Contemplativa, comunicativa, aurática	Informativa/simulativa/interativa
Conteúdo	Curadores	Imagen; ficha técnica, <i>hyperlink</i>

Fonte: Autor

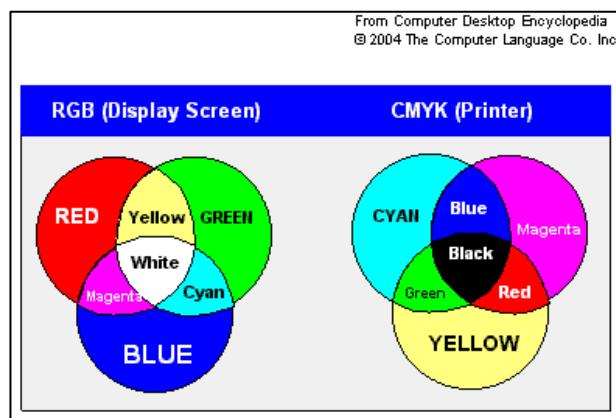
Diante da análise dos resultados, conclui-se que a digitalização das obras analisadas e sua transposição para outros tipos de suporte fez com que ocorressem mudanças, tanto estruturais e formais, quanto na função e no conteúdo da obra. Pois, os diferentes tipos de suportes, como o impresso (papel) e o digital na *Web*, possuem particularidades que afetam as qualidades da obra original. Ao fazermos uma análise das particularidades de cada tipo de suporte, levando em consideração a questão que mais afetou a obra (mudança na cor) concluímos que a cor luz difere da cor pigmento.

Ao digitalizarmos uma obra, transportando sua imagem para o ambiente da Internet em forma de catálogos *on-line*, para ser vista por meio de dispositivos eletrônicos (computador, *tablet*, telefone móvel etc.) sua estrutura muda de pigmento orgânico (tinta,

tecido, parede etc.) para a linguagem computacional (pixel). Essa semiose afeta tanto a obra quanto seu significante.

Quando uma obra é digitalizada para servir de catálogo impresso, as cores que serão aplicadas para representar a obra constituem-se de quatro canais de cores: ciano, magenta, amarelo e preto – sistema de cores chamado de (CMYK). No entanto, se a mesma obra for digitalizada e transposta para o ambiente *Web*, ocorrem transformações também relacionadas à mudança na cor, pois ela passa para a linguagem (RGB), que funciona sob três canais de cores: vermelho, verde e azul, que também dará origem a outras variáveis, gerando uma composição diferente da linguagem (CMYK), como mostra a FIG. 10.

FIGURA 10: Sistema de cores



Fonte: <http://www.answers.com/topic/rgb-2>

Dessa maneira, o papel do profissional da informação, em conjunto com o profissional da área da informática (designer gráfico), e o da museologia, entre outros, deverá ser o de buscar alternativas para o estabelecimento de uma adequada visualização das obras de arte nos variados tipos de suportes e formas de apresentação, no sentido de manter a integridade das informações representadas, levando em consideração a relevância dos diferentes tipos de ambientes informacionais na aquisição e apropriação do conhecimento.

4 CONSIDERAÇÕES FINAIS

Percebemos que, por meio das reproduções digitais das obras de arte, houve uma quebra de paradigma, passando de imagem contemplação para imagem representação, até culminar em imagem simulação (SANTAELLA, 2001). Nesse contexto, um número considerável de obras de grandes mestres tem sido disponibilizado, pelo menos potencialmente. Embora a socialização aproxime o patrimônio cultural da humanidade de

uma grande parcela da população, as obras em si, ou melhor, seus simulacros, deixam a desejar, pois, na passagem de uma linguagem semiótica para outra ocorrem perdas informacionais relativas à estrutura plástica e que enriquecem a obra e caracterizam sua essência.

A transformação mais evidente observada pela análise das obras originais e sua versão digitalizada em forma de catálogos, tanto impressos quanto os da *Web* foi: a diferença de cores entre o catálogo impresso e a obra, levando em consideração os tipos de papel utilizados no catálogo impresso, e a diferença de cores entre a obra e o catálogo *on-line*, pois a cor luz difere da cor pigmento. Essas diferenças refletem no resultado final das imagens dos catálogos (impresso e *on-line*). Nas obras tridimensionais, como as obras escultóricas, há perdas em uma dimensão, ficando a imagem achatada e bidimensional. Com os catálogos digitais *on-line*, há esforço e custo menores no processo de atualização do acervo, o que os torna mais acessíveis ao visitante, ganhando em extensividade.

A herança artística da humanidade está agora ao alcance de todos os interessados em arte e para quem possua acesso à rede Internet, não importando sua localização geográfica ou sua classe social.

REFERÊNCIAS

- AUMONT, J. **A imagem**. Campinas: Papirus, 1993.
- BARRETO, A. A. Os destinos da ciência da informação: entre o cristal e a chama. **DataGramazero**, Rio de Janeiro, n. 0, dez. 1999. Disponível em: http://www.datagramazero.org.br/dez99/Art_03.htm. Acesso em: 17 nov. 2011.
- BARRETO, A. A. Uma quase história da Ciência da Informação. **DataGramazero**, Rio de Janeiro, v. 9, n. 2, abr. 2008. Disponível em: <http://dgz.org.br/ago07/Ind_com.htm>. Acesso em: 28 ago. 2011.
- BENJAMIN, W. A obra de arte na era de sua reproduibilidade técnica. In: _____. **Magia e técnica, arte e política**: ensaios sobre literatura e história da cultura. São Paulo: Brasiliense, 1994. p. 165-196. (Obras Escolhidas, v. 1).
- COUCHOT, E. **A tecnologia na arte**: da fotografia a realidade virtual. Porto Alegre: Ed. UFRS, 2003.
- CRIPPA, G. A faceta humanística da ciência da informação: ordem e memória do/no museu. In: SEMINÁRIO SERVIÇO DE INFORMAÇÃO EM MUSEU, 1., 2010, São Paulo. **Anais...** São Paulo: Pinacoteca do Estado de São Paulo, 2010. p. 23-51.
- CZEGLEDY, N. Arte como ciência: ciência como arte. In: DOMINGUES, D. (Org.). **Arte e vida no século XXI**: tecnologia, ciência e criatividade. São Paulo: Ed. UNESP, 2003.

EIRIN DE RAPP, M. N. Web-paginas de arte: informação X arte. In: PINHEIRO, L. V. R.; GONZÁLEZ DE GÓMEZ, M. N. (Org.). **Interdiscursos da Ciência da Informação: arte, museu e imagem.** Rio de Janeiro; Brasília, DF: IBICT/DEP/DDI, 2000. p. 43-58

FARTHING, S. **Tudo sobre arte.** Rio de Janeiro: Sextante, 2011.

GRAU, O. **Arte virtual:** da ilusão a imersão. Tradução Cristina Pescador *et al.* São Paulo: Ed. Unesp: Ed. Senac São Paulo, 2007.

JORENTE, M. J. V. **Tecnologías, mídias, criação e hipertextualidade na transformação da informação em conhecimento interativo.** 2009. 244f. Tese (Doutorado em Ciência da informação) – Faculdade de Filosofia e Ciência, Universidade Estadual Paulista, Marília, 2009.

LEMOS, A. **Cibercultura:** tecnologia e vida social na cultura contemporânea. Porto Alegre: Sulina, 2008.

MEY, E. S. A. **Introdução à catalogação.** Brasília DF: Briquet de Lemos, 1995.

NEGROPONTE, N. **A vida digital.** Tradução de Sérgio Tellaroli. São Paulo: Companhia das Letras, 1995.

PINHO, J. A. S. As novas tecnologias da informação e comunicação diante da transversalidade entre natureza e cultura. **Culturas Midiáticas**, Paraíba, Ano I, n. 1, jul./dez., 2008. Disponível em:
<http://www.cchla.ufpb.br/ppgc/smartgc/uploads/arquivos/b29c8438c920100526091245.pdf>. Acesso em: 24 jun. 2011.

PLAZA GONZALES, J. As imagens de terceira geração, tecno-poéticas. In: PARENTE, A. (Org.). **Imagen máquina:** a era das tecnologias do virtual. Rio de Janeiro: Ed. 34, 1993. p. 72-88.

PLAZA GONZALES, J. **Tradução intersemiótica.** São Paulo: Perspectiva, 2008.

SANTAELLA, M. L. NOTH, W. **Imagen:** cognição, semiótica e mídia. São Paulo: Iluminuras, 2001.

SANTAELLA, M. L. **O que é Semiótica.** São Paulo: Brasiliense, 1983.

SANTAELLA, M. L. **Semiótica aplicada.** São Paulo: Cengage Learning. 2010.

SILVEIRA, L. F. B. **Curso de Semiótica Geral.** São Paulo: Quartier Latin, 2007.

UTILIZANDO O GOOGLE ANALYTICS PARA MONITORAR A AUDIÊNCIA DE BLOGS

USING GOOGLE ANALYTICS TO MONITOR THE AUDIENCE OF BLOGS

Célio Andrade Santana Júnior

Camila Oliveira Lima

Diego Andres Salcedo

Andréia Cândida de Souza

Resumo: Realizar as atividades de monitoramento de mídias sociais se mostra como um trabalho complexo devido ao comportamento passivo que a maioria dos usuários (*lurkers*) exercem em sites de redes sociais e não deixam rastros perceptíveis da participação dos mesmos nestes espaços. Para realizar esse monitoramento, foram adotadas atividades de web *analytics* que podem auxiliar os administradores destas mídias sociais a avaliar a audiência “invisível” do seu canal de comunicação. Este trabalho tem como objetivo apresentar a ferramenta Google *Analytics*, como ela pode ser utilizada para monitorar a audiência de blogs e que tipos de indicadores são coletados. Para avaliar a ferramenta, esta foi aplicada ao blog *vernashsunhas.blogspot.com* por um período de 8 meses para coletar informação sobre a audiência do blog. A coleta dos dados da audiência foi realizada pela própria ferramenta enquanto a análise se deu a partir uma análise temática baseada nos tipos de métricas encontrados. Foi percebido que a ferramenta auxilia apenas em parte no trabalho do monitoramento dos dados relativos à audiência uma vez que ela se mostra capaz de coletar os dados relativos às métricas frias que refletem apenas a participação do usuário nos blogs não se mostrando adequada para coletar dados relativos às métricas mornas e quentes relativas a engajamento e fidelização. Desta forma, o Google *Analytics* não pode ser utilizado como única fonte de informação para um monitoramento eficiente da audiência de um blog.

Palavras-chave: Blogs. Google *Analytics*. *Web Analytics*.

Abstract: Perform activities of monitoring social media appears as a complex task due to passive behavior that most users (*lurkers*) exert on social networking sites and leave no discernible trace of the participation of these same spaces. To perform this monitoring, were adopted from web analytics that can assist administrators of these social media to review the "invisible" hearing your channel of communication activities. This paper aims to present Google Analytics tool, as it can be used to monitor the hearing of blogs and what kind of information it collects indicators. To evaluate the tool, it was applied to the "vernashsunhas.blogspot.com" blog for a period of 8 months to collect information on the hearing of the blog. The collection of audience data was performed by the tool while the analysis was done from one based thematic analysis on the types of metrics found. It was noticed that the tool helps only partly the work of monitoring data relating to the audience once she proves able to collect data on the cold metrics that reflect the user participation in blogs not showing adequate to collect data warm and hot metrics related to engagement and loyalty. Thus, Google Analytics can't be used as the sole source of information for effective monitoring of the hearing of a blog.

Keywords: Blogs. Google *Analytics*. *Web Analytics*.

1 INTRODUÇÃO

Atualmente vivemos em um mundo no qual os crescentes avanços da tecnologia e da comunicação vêm exercendo uma forte influência sobre a sociedade (SCHONS, 2007, p.

1) Dentre os avanços na comunicação, o meio que mais se destaca é a internet, devido a sua amplitude global e facilidade de acesso (JARDIM, 2010, p. 6).

A Internet se mostra cada vez mais integrada à vida cotidiana em questões relativas ao uso das informações neste espaço estão mudando do “quem está logado” para uma análise mais aprofundada do que as pessoas fazem online. Entre estas considerações, existe um crescente reconhecimento de que a Internet está estimulando conexões e forjando novas ligações entre pessoas e organizações (HAYTHORNTHWAITE, 2005).

Com o advento da Web 2.0, termo cunhado pela empresa norte-americana O'Reilly Media para designar a segunda geração da Web. Diversas ferramentas de Tecnologia de Informação e Comunicação (TIC) foram construídas criando assim novos espaços para construção do conhecimento, além de proporcionar a iteração entre usuários em tempo real bem como a participação destes usuários de forma mais atuante (ALEXANDER, 2006).

Novos meios de comunicação se tornaram cada vez mais familiares ao público em geral e o seu uso foi sendo adaptado a partir de convenções comuns. Estes novos canais se mostram essenciais para manter o trabalho e conexões sociais da vida cotidiana atravessando os diversos mundos sociais do indivíduo tais como: trabalho, casa, e qualquer tipo de geografia (WELLMAN; HAYTHORNTHWAITE, 2008).

Esta nova geração da Internet foi caracterizada pelo surgimento das mídias sociais que logo se tornaram populares reunindo um número considerável de usuários em um mesmo espaço. Neste contexto, podemos considerar que a primeira mídia social a atingir um número massivo de usuários foram os Blogs e, por este motivo, provavelmente, é a ferramenta da Web 2.0 mais conhecida e utilizada pelos usuários (JARDIM, 2010).

Bruns (2008) define blogs como uma página na internet que, se pressupõe, ser atualizada com frequência através da publicação de mensagens – chamados de posts – constituídas por imagens e/ou textos normalmente de pequenas dimensões, muitas vezes incluindo hiperlinks para outras páginas de interesse e/ou comentários e pensamentos pessoais do autor, e apresentadas de forma cronológica, sendo as mensagens mais recentes normalmente apresentadas em primeiro lugar.

Barros (2004) sugere que uma das vantagens em se utilizar dos blogs, é que qualquer usuário que tenha acesso a Internet pode criar uma página deste tipo, uma vez que a tecnologia empregada é bastante simples. E que o sucesso no uso dos blogs se dá devido à facilidade de se publicar conteúdo online. Neste contexto, os blogs são parte de uma crescente conjunção de comunicação pessoal e ferramentas de gerenciamento de

informação que fornecem um mar infinito de histórias para leitores que compartilham dos mesmos interesses.

Scoble e Israel (2006) sugerem que com a popularização dos blogs, diversos fenômenos relativos ao comportamento dos usuários na Internet foram observados. O primeiro deles é a que os usuários passam não somente a consumir informações, mas também, produzir conteúdo neste espaço. Esse novo perfil de usuários foi denominado prosumers, expressão oriunda da língua inglesa a partir da junção das palavras Producer e Consumer. Este neologismo foi registrado por Alvin Toffler (1980) na obra “A Terceira Onda”, onde ele sugere o novo papel do consumidor da informação e bens culturais na sociedade pós-moderna. Nielsen (2006) observa outro fenômeno que é a massiva presença de usuários que não geram nenhum tipo de conteúdo em um blog, estes usuários são chamados de Lurkers.

Nielsen (2006) sugere que 90% do conteúdo publicado em mídias sociais são produzidos por apenas 1% dos usuários, enquanto 9% dos usuários geram os outros 10% e 90% dos usuários são lurkers. Isso implica que para cada usuário que comenta, realiza uma marcação, curte ou compartilha conteúdo de um blog, outros nove visitaram e consumiram aquele conteúdo.

O uso de blogs no meio empresarial fica evidenciado no estudo state of blogosphere (2011). Este sugere que uma em cada oito empresas mantém um blog ativo. A pesquisa também indicou que 57% das empresas com blogs adquirem pelo menos um novo cliente/consumidor através deste canal de comunicação. E por fim aponta 61% dos consumidores que compram em lojas virtuais nos Estados Unidos fizeram compras baseada em recomendações de blogs. Cerca de 46% dos blogueiros entrevistados se consideram probloggers, ou seja, blogueiros que se tornaram referência para determinados assuntos de interesse e acabaram por transformar os seus blogs em um meio rentável caracterizando, portanto, a profissionalização destes. (AKRITIDIS; KATSAROS; BOZANIS, 2009).

Devido ao caráter comercial alcançado pelos blogs, se tornou uma prática comum o monitoramento do conteúdo por parceiros ou concorrentes promovido nestes canais (HONSCHA, 2009) ou até mesmo monitorar a própria audiência com o intuito de conhecer, e se comunicar, melhor com os usuários “consumidores” (PIKAS, 2005).

Blogs são considerados fontes de informação “potencialmente útil” e devido à sua numerosa quantidade, elaborar uma lista destes e monitorá-los é considerada uma tarefa exaustiva (NANNO ET. AL, 2004). Por movimentarem uma quantidade muito grande de dados, em um espaço tão curto de tempo, a tarefa de monitorar as informações que circulam

nestas mídias seria impossível sem o auxílio de ferramentas técnicas adequadas (CARVALHO, 2012).

Uma empresa com uma visão estratégica online bem estabelecida não pode mais ignorar o potencial dos blogs como ferramenta de comunicação. Tornou-se vital Acompanhar as conversações que ocorrem na rede para conseguir agir rapidamente às movimentações do mercado (CIPRIANI, 2006, p.113). Por isso, mais importante para as empresas do que avaliar o conteúdo das mídias sociais relativos aos concorrentes, é monitorar o público que consome o conteúdo criado em suas próprias mídias sociais (HONSCHA, 2009).

A motivação deste trabalho vem do reconhecimento da complexidade em monitorar a audiência dos próprios visitantes em blogs. Então perguntas do tipo: como perceber a audiência de um blog específico? Como coletar essas informações? As informações disponibilizadas são suficientes para determinar a satisfação dos usuários com um blog?

O presente trabalho tem como objetivo geral apresentar a ferramenta Google Analytics e utilizá-la para realizar o monitoramento de um blog real com o intuito de verificar quais informações sobre os visitantes são coletados pela ferramenta e se é possível avaliar a audiência do blog a partir destas informações. Este estudo justifica-se pela análise das métricas fornecidas pela ferramenta de web analytics e determinar se ela é, efetivamente, satisfatória no processo de monitoramento de blogs.

2 LITERATURA PERTINENTE

2.1 Blogs

Enquanto fenômeno emergente e persistente da cibercultura, os *blogs* demonstram um sintoma (e consequência) do desejo das pessoas de manter-se conectadas e em comunicação (LEMOS, 2009). Inicialmente os blogs foram considerados “diários *online*”, contudo, com o desenvolvimento desta mídia social, houve uma evolução em termos das ferramentas em si e em tendências temáticas até atingir a profissionalização (HINERASKY, 2012).

O termo “*weblog*” foi dado pelo norte-americano Jorn Barger para referir-se ao seu jornal *online* e este se deriva da junção das palavras “*web*” (rede) e “*log*” (diário de navegação) e foi utilizada para designar uma página que permite a construção de um diário pessoal da rede (HINERASKY, 2012). A abreviação “*blog*” teria surgido apenas em maio de 1999, quando Peter Merholz divide o termo em “*we-blog*” (nós blogamos) ou apenas “*blog*”, criando ao mesmo tempo o verbo *blogar* e o sujeito *blogueiro* (MALINI, 2008).

Ordunã (2007) sugere que *blogs* são páginas pessoais na Internet que, devido à semelhança de diários *online*, tornam possíveis a todos publicar na rede. Por ser uma publicação focada no usuário e no conteúdo, e não na programação ou no design gráfico, os *blogs* multiplicaram o leque de ações dos usuários de Internet de levar para a rede conteúdo próprio e atualizado, sem intermediários (MALINI, 2008; ORDUNÃ, 2007).

Sendo assim, podemos considerar os *blogs* como espaços democráticos onde os usuários são livres para opinar sobre qualquer assunto e discutir a respeito de qualquer tema de seu interesse, compartilhando suas ideias com os usuários da rede, permitindo a disseminação de informações que eram, muitas vezes, subestimadas pelas mídias tradicionais. Essa maior possibilidade de interconexões promovida pela internet permite que os *blogueiros* acessem o conteúdo uns dos outros, por meio de referências, formando assim uma comunidade ou rede social (SPYER, 2007; ORDUNÃ, 2007).

Ordunã (2007) sugere que o principal elemento de um *blog* são as postagens (*posts*), ordenadas segundo a cronologia inversa (com as publicações mais recentes primeiro), em que cada uma possui um endereço URL permanente que facilita sua conexão a partir de sites externos. As histórias podem ser arquivadas e apresentadas, ou cronologicamente (por meses e anos) ou por temática (categorias/*tags*). A maior parte dos *blogs* traz uma seleção de conexões (*blogroll*) que reúne os *websites* lidos ou pelo menos recomendados pelo autor e alguma referência pessoal que, com o título e a descrição do *blog* que ajudam o leitor a situá-lo.

Esta interconexão entre todos os tipos de *blogs* por meio das *blogrolls* e suas hiperligações configura um fenômeno social, um “ente coletivo”, denominado por Willian Quick, em 2001, como blogosfera (MALINI, 2008), caracterizado pelo espírito comunal que distingue preferências, popularidades de estilos e produtos (HINERASKY, 2012). A blogosfera é um espaço no qual são estabelecidas milhares de conversações paralelas e muitas delas podem possuir informações úteis para as organizações e até mesmo para os indivíduos (ORDUNÃ, 2007).

Hinerasky (2012, p. 43) afirma que é possível diferenciar, na prática, as etapas que envolvem a conexão entre autores e leitores: a) tomada de conhecimento do *blog* pelo leitor; b) descoberta de um interesse comum; c) trocas individuais e trocas em grupo; d) confiança e construção da autoridade. Estas etapas levam aos *blogueiros* à conquista da audiência e a possibilidade de notoriedade e legitimidade entre os pares, leitores e empresas.

Neste sentido, a audiência de um *blog* se constitui, basicamente, das relações estabelecidas entre os interagentes e é medida de duas formas, (i) por captação e (ii) por interação (HINERASKY, 2012).

2.2 Web analytics e monitoramento de mídias sociais

Segundo o Aberdeen Group (2000), *web analytics* é o monitoramento e comunicação de uso de uma página na internet de modo que empresas possam entender melhor a complexa relação entre as ações dos visitantes e uma página em si com o intuito de otimizar o sítio para aumentar as vendas e a lealdade do cliente. Waisberg e Kaushik (2009) apontam que a *web analytics* é a ciência de aprimorar *websites* com o intuito de aumentar a sua rentabilidade através da promoção de uma melhor experiência do usuário quando este o visita. É uma ciência porque a *web analytics* utiliza estatísticas, técnicas de mineração de dados, e um processo metodológico.

Inan (2002) aponta que o surgimento da *web analytics* se deu por entender que a internet representa um canal de negócios em si mesmo, ao invés de ser apenas uma ferramenta de publicidade, e como tal, necessita de uma abordagem holística para a medição de elementos na rede.

Quando tratamos de redes sociais em si as técnicas de *web analytics* são conhecidas como monitoramento de mídias sociais. E neste sentido, esta surge como uma nova área de conhecimento que visa “captar, agregar, analisar e auxiliar as organizações na criação de ações e estratégias para o negócio de forma rápida e com base nas informações geradas de forma espontânea pelos consumidores nesses canais” (E-LIFE, 2012). O papel do monitoramento é ser um radar da empresa no “burburinho” online, que remete não só as opiniões e avaliações dos consumidores, mas também o comportamento do mercado com relação à marca e seus concorrentes. Os depoimentos espontâneos disponíveis nas redes sociais são também “fonte para novas ideias e possibilidades para produtos e serviços” (LIFE, e. 2013b).

A empresa SocialMedia Monitor (2013) afirma que a complexidade deste monitoramento parte da nova realidade uma vez que “[...] as marcas já tiveram controle absoluto sobre o conteúdo exposto de sua imagem enquanto hoje, com a internet e, mais especificamente, com as redes sociais, são os consumidores que criam o conteúdo”. Neste contexto, o consumidor, que também é usuário das redes sociais, assume também outro importante perfil: o de influenciador (SOCIALMEDIA MONITOR, 2013).

Silva (2010) define monitoramento de mídias sociais como:

Coleta, armazenamento, classificação, categorização, adição de informações e análise de menções online públicas a determinado(s) termo(s) previamente definido(s) e seus emissores, com os objetivos de: (a) identificar e analisar reações, sentimentos e desejos relativos a produtos, entidades e campanhas; (b) conhecer melhor os públicos

pertinentes; e (c) realizar ações reativas e proativas para alcançar os objetivos da organização ou pessoa (SILVA, 2010).

Quando tratamos especificamente sobre o monitoramento de *blogs*, Orduña (2007, p. 179) elucida:

A opinião dos consumidores está ali, esperando ser ouvida pelas empresas, e pode ser consultada gratuitamente. Informações sobre referencias, idéias, propostas, necessidades, sonhos e aspirações de indivíduos, que até então somente eram obtidas mediante pesquisa de mercado cara, agora estão disponíveis e ao alcance de todo mundo, inclusive dos concorrentes (ORDUÑA, 2007, p.179).

A E-Life (2013) aponta que algumas vantagens diferenciam o monitoramento de mídias sociais das demais metodologias de investigação do mercado e do comportamento do consumidor dentre as quais se destacam: (i) Velocidade e custo, (ii) Informações sem filtro, (iii) Informações espontâneas e contextualizadas, (iv) dados qualitativos e quantitativos, (v) Informações vão além das perguntas.

Existem duas maneiras de monitoramento de mídias sociais (i) Reativo e (ii) Proativo. No monitoramento reativo, a marca responde às citações diretas do consumidor, seja positiva ou negativa. Já no monitoramento proativo, a marca deve monitorar as empresas concorrentes e a sua própria marca a fim de captar potenciais clientes (MEDIA, 2013).

2.3 Ferramentas para o monitoramento de mídias sociais

Para auxiliar no processo de coleta de dados para o monitoramento da audiência das mídias sociais existem inúmeras ferramentas. Laine e Fruhwirth (2010) ressaltam que:

Ferramentas para monitoramento de mídias sociais são (na maioria das vezes) serviços de software oferecidos através da Internet para filtrar e analisar o conteúdo textual produzido por e na mídia social. As ferramentas encontram conteúdo baseados nas palavras-chave definidas pelos usuários. As ferramentas incorporam múltiplas funcionalidades, como análise de volume, fonte, autor, palavra-chave, região e sentimento, e reportam estas análises convenientemente de modo gráfico (LAINE, M.; FRUHWIRTH, C., 2010).

Na concepção de Faustino (2013) a importância em medir o impacto social de uma publicação específica ou do próprio *blog* como um todo tem como objetivo se adequar as preferências do usuário. Medir este impacto social possibilita a compreensão de quais os tipos de conteúdo os leitores preferem. Desta forma, o *blogueiro* pode criar novos *posts*, direcionados, com o intuito atrair mais leitores para o *blog*.

Por isso, algumas ferramentas de monitoramento de audiência já são disponibilizadas pelas próprias ferramentas de *blogs*. Para o Wordpress⁴ existe o *plugin* Social Metrics que permite uma análise sobre os *posts* mais curtidos e divulgados. Já o *Blogger*⁵ fornece ao administrador uma “visão geral” das métricas, indicadores e medição de audiência que englobam as visualizações por página, hora, dia, semana, mês e total de visualizações do *blog* (FAUSTINO, 2013).

Neste sentido, cabe citar a ferramenta ADDThis⁶, responsável por mostrar números relacionados ao engajamento dos visitantes, mostrando como eles estão compartilhando o conteúdo do *blog* – seja copiando a URL e compartilhando com outras pessoas, quantos cliques a URL recebe, qual rede social é mais influente para o um *blog*, qual o conteúdo tem maior “efeito viral”. Com esses dados é possível entender o comportamento do usuário dentro da mídia social, a partir a visualização dos conteúdos que são mais acessados e a forma como são acessados (EIS, 2011).

O Google disponibiliza um serviço gratuito de análise de audiência de qualquer página na Internet chamado Google *Analytics*. A ferramenta é adicionada a um *blog* através de um código HTML. O Google *Analytics* gera relatórios de medição de audiência do *blog* e é capaz de identificar além das tradicionais métricas de taxa de exibição, “hit” de uma página, localização geográfica do visitante, forma com a qual chegou à página seja através de *links* de outros sites, buscador e quais as palavras chaves utilizadas (KRISHNAMURTHY; WILLS, 2009). Para o monitoramento de *blogs*, especificamente, algumas ferramentas são listadas por Silva (2012) tais como: Radian6, Scup, Trendrr, Alterian SM2, Seekr, Livebuzz, PostX, SocialMetrix, ScoutLabs.

2.4 Classificações de métricas para monitoramento

Diferente do mundo “não conectado” onde é difícil precisar números e estatísticas sobre como as pessoas interagem com as marcas que elas consomem, na *web* isso acontece de forma muito natural. A Internet é repleta de dados e números disponíveis chamadas de métricas e se faz necessário utilizar os métodos adequados para o monitoramento adequado nesse meio. (DINAMIZE, 2011).

Uma pesquisa intitulada “Mídias Sociais nas Empresas” realizada pela Deloitte Touche Tohmatsu (2010) afirmou que as empresas mensuram as iniciativas de mídias

⁴ www.wordpress.com

⁵ www.blogger.com

⁶ <http://www.addthis.com/>

sociais através, principalmente, das métricas e indicadores. Silva (2012) sugere que o uso de métricas voltadas para mídias sociais tem pontos positivos no que tange a persistência e “*buscabilidade*” dos dados. As informações são públicas e disponíveis e existem diversos níveis de mensuração. Contudo, vale ressaltar que esta medição varia de acordo com a mídia social pesquisada e assim, o profissional tem que lidar com as diversas terminologias diferentes presentes na rede.

Tuten (2008) destaca que medidas como número de visitantes únicos, páginas vistas, frequência de visitas, duração de visitas e taxas de cliques são inapropriados para captar as informações realmente relevantes para campanhas realizadas em mídia social. O aspecto quantitativo na mensuração de número de pessoas presentes em uma comunidade virtual, por exemplo, é útil, mas insuficiente para entender a essência da interação que os consumidores estabeleceram com a marca e o grau de engajamento.

Silva (2012) aponta que o monitoramento de mídias sociais se dá em quatro níveis: (i) Engajamento, (ii) Alcance, (iii) Adequação e (iv) Influência que são definidos como:

Engajamento: Se refere ao grau de participação e envolvimento de determinado perfil ou grupo de pessoas em relação a um tema ou assunto;

Alcance: Se refere ao grau efetivo de disseminação de determinado conteúdo ou grau de disseminação potencial que um perfil possui;

Influência: Se refere ao grau de atenção e mobilização que um determinado perfil pode gerar em outros;

Adequação: Se refere ao grau de proximidade que determinado conteúdo tem com as características e valores desejados.

Outra classificação bastante utilizada é apresentada por Albee (2014) que são as métricas *Hard* e *Soft*. As métricas “*Hard*” são relacionadas às vendas, importando-se somente com os resultados relativos à receita. Já as métricas *Soft* são focadas em iniciativas que impulsionam o engajamento e a interação com o usuário. Segundo a mesma autora, apesar da dicotomia, a utilização de ambos os tipos são igualmente importantes para uma organização. Não se consegue obter receita sem iniciativas que impulsionem o comportamento do usuário na rede.

Broek (2011) sugere outra classificação que engloba três tipos de métricas: as (i) frias, (ii) mornas e (iii) quentes. As métricas frias se referem aos dados quantitativos que medem o comportamento da audiência em geral tais como alcance, frequência e visualizações. As métricas mornas são ligadas ao engajamento dos usuários com a mídia social em questão e avalia comportamentos específicos como análise de sentimento e

avaliação de fatores virais. As métricas quentes avaliam comportamentos relativos a conversão, ou seja, mede como o negócio é alavancado a partir da mídia social em análise e para isso pode ser usado avaliações de *leads* ou mesmo uma avaliação do ecossistema. Medina (2013) afirma que as métricas frias têm por finalidade a avaliação do valor de mercado, faturamento, *Market Share*, lucro e preço. As métricas mornas objetivam a medição de opiniões, vínculo emocional, relevância, sustentabilidade e percepção.

Levando em conta a tipologia das métricas, Silva (2013) afirma que existem três grupos de métricas sociais que necessitam de um cuidado maior no monitoramento: (i) Alcance, (ii) Engajamento, (iii) Influência. As métricas de Alcance e impressões se referem à audiência real ou potencial de uma mensagem, através das visualizações, seguidores, membros, entre outros. O Engajamento permite analisar qual parcela da audiência interage de algum modo com o perfil e o conteúdo. As métricas de influência e relevância permitem analisar o quanto a marca são percebidas pelos consumidores.

3 METODOLOGIA

Esta pesquisa utilizou de abordagens tanto quantitativas, tanto qualitativas para a sua realização. A parte quantitativa se refere à avaliação do desempenho do *blog* e da validação da regra 90-9-1 proposta por Nielsen (2006). A parte qualitativa da pesquisa refere-se à avaliação daquilo que o *google analytics* se propõe a medir de todo o universo identificado no monitoramento de mídias sociais.

Para tanto, utilizou-se o *google analytics* para monitorar o *blog* Ver nas Unhas⁷ durante o período de Julho de 2013 a Fevereiro de 2014 com o intuito de coletar informações sobre a audiência da página e avaliar quais métricas foram, de fato, foram coletadas pela ferramenta. A escolha do *blog* se deu por duas razões: (i) pelo fato do *blog* possuir certa influência no nicho de mercado de beleza e (ii) os administradores do *blog* permitiram o monitoramento do mesmo e compartilharam todos os dados coletados.

Esta pesquisa é classificada como uma pesquisa empírica quanto aos meios, sem o controle ou interferência do pesquisador, e a busca pelas respostas é dada pela observação do fenômeno. Quanto aos fins esta pesquisa é classificada como uma pesquisa descritiva pelo objetivo de verificar o fenômeno e inferir relações com o meio a partir da observação (Michel, 2009).

O método de pesquisa utilizado foi o comparativo dos dados coletados com a teoria 90-9-1 de Nielsen (2006) para a parte quantitativa e das métricas coletadas pelo *google*

⁷ <HTTP://vernasanhas.blogspot.com>

analyitcs com as categorias propostas por Broek (2011) e Silva (2013). A técnica de coleta de dados foi a observação intensiva direta a partir da própria ferramenta *google analytics*.

Os passos realizados para esta pesquisa estão listados na FIGURA 1 a seguir:

FIGURA 1 - Fluxo de execução da pesquisa



Fonte: Os Autores (2014)

Inicialmente foi realizada uma pesquisa bibliográfica a respeito do monitoramento de *blogs* e de sua investigação na Ciência da Informação. Parte considerável do material encontrado não são artigos de conferências e periódicos e sim publicações em meios especializados e livros. Outro fato relevante identificado é que grande parte do *corpus* relativo aos artigos acadêmicos são de áreas humanas como Marketing e Administração o que sugere uma lacuna de pesquisa do tema em Ciência da Informação.

A seguir foi necessário escolher um *blog* que possuísse certo volume de postagens para que pudéssemos avaliar dados mais próximos daqueles *blogs* que possuem grande reputação. Além disso, todas as ferramentas gratuitas para o monitoramento exigem um acesso privado, o que foi gentilmente atendido pelos administradores do *blog* Ver nas Unhas. A escolha da ferramenta *Google Analytics* se deu por, além de ser gratuita, possuir uma comunidade de prática relevante, sendo uma das mais populares para realizar o monitoramento de *blogs*.

Uma vez selecionada as ferramentas, foi necessário escolher, as métricas que seriam utilizadas nesta pesquisa. Foram consideradas aquelas que captaram pelo menos um evento no *blog* entre Julho de 2013 e Fevereiro de 2014. Após a escolha das métricas foi iniciado o processo de coleta de dados que foi realizada pela própria ferramenta.

A análise dos dados foi realizada a partir das informações capturadas pela ferramenta. Para a parte quantitativa foi coletado o total de visitantes, e no *blog* foram

coletados os comentários, curtidas e compartilhamentos para outras mídias sociais. Nesta etapa se fez necessário excluir os “visitantes” referentes a outras ferramentas de monitoramento tais como Scoop e Social Media Monitor que também monitoravam o *blog*. Para a parte qualitativa foi necessário identificar as métricas que tinham coletado no mínimo uma ocorrência do evento e classificá-la em uma das categorias de Broek (2011) ou Silva (2013).

4 RESULTADOS

Os dados demográficos do *blog* nos nove meses de coleta estão apresentados no QUADRO 1 a seguir.

QUADRO 1 – Dados Demográficos do *Blog*

Métrica	Quantidade
Número de Visitantes	3183
Número de Visitantes Únicos	401
Ferramentas de Monitoramento	4
Usuários que Comentaram, Compartilharam ou Curtiram.	32
Total de <i>Lurkers</i>	365
Número de Postagens	88
Tempo médio por Visualização	2 Min. e 15 Seg.
Visitas por Visitante	9,3
Visualizações de Páginas por Visitante	3,2
Visualização de Páginas	9328

Fonte: Os Autores (2014)

Olhando para estes dados podemos perceber que a regra 90-9-1 se aplica a este *blog* ficando com a seguinte distribuição: Produtor de Conteúdo (0,2%), Pessoas que Comentam, Curtem e Compartilham (8%) e *Lurkers* (91,7%). Vale ressaltar que esse resultado não tem como objetivo generalizar como verdadeira a regra de Nielsen, mas apenas constatar que a mesma se mostra verdadeira nos 8 meses em que este *blog* foi avaliado.

As métricas consideradas por este estudo estão listadas no Quadro 2 a seguir

QUADRO 2 – Métricas identificadas pelo *Google Analytics*

Métrica	Descrição	Classificação
Visualização da Página	Quantas vezes a página foi visualizada	Métrica Fria
Visualização da Página por usuários	Quantos usuários visualizaram uma página	Métrica Fria
Visitas	Quantas visitas uma página obteve. Uma visita é definida por um intervalo de 3º minutos, se um mesmo usuário visualiza a mesma página mais de uma vez em 30 minutos essa métrica só será incrementada uma vez.	Métrica Fria
Visitantes Únicos	Quantos visitantes o site obteve, excluindo os retornos (visitantes repetidos).	Métrica Fria
Página de Entrada	Qual página do blog os usuários costumam acessar primeira (chegada ao blog)	Métrica Fria
Página de Saída	Qual página do blog os usuários costumam acessar por último (saída no blog)	Métrica Fria
Duração da Visita	Quanto tempo dura a visita	Métrica Fria
Página Referência	Qual página o usuário estava imediatamente antes de chegar ao blog (quem indicou o blog)	Alcance
Taxa de Cliques	Quantos cliques, em um determinado hiperlink, foram dados por um usuário.	Métrica Fria
Próxima Página	Qual a página o usuário vai imediatamente após sair do blog.	Influência/Alcance
Segunda Página	Quais páginas o usuário visitou dentro do blog	Métrica Fria
Novos Visitantes	Quantos visitantes são novos no blog	Métrica Fria
Visitantes Antigos	Quantos visitantes retornaram ao blog	Métrica Fria
Visitas por Visitante	Vistas por Visitante	Engajamento/Métrica Morna
Tempo para Retorno	Quanto tempo um visitante demora a retornar	Engajamento
Frequência	Quantas vezes uma ação foi realizada por um visitante	Engajamento
Taxa de Saída de Página	Quantidade de Saídas do Blog (a partir de uma determinada página) pela quantidade de vezes em que ela foi visualizada.	Métrica Fria
Taxa de Rejeição	Porcentagem de Visitas que só visualizaram uma página.	Métrica Fria
Visualização por Visita	Quantas visualizações de uma página foram realizadas em uma visita.	Métrica Fria
Tempo na Página	Quanto tempo uma visualização de uma página durou	Engajamento
Conversão	Taxa de usuários que se tornam assíduos no blog	Métrica Quente
Páginas visitadas a partir de buscas	Quantas vezes uma página foi visitada a partir de uma busca interna no blog.	
Buscas usadas	Textos de buscas utilizados	Métrica Morna/ Engajamento
Categorias de Busca	Quais categorias de busca foram utilizadas	Métrica Morna/ Engajamento
Buscas Únicas	Quantas dessas buscas foram únicas	Métrica Morna/ Engajamento

Fonte: Os Autores (2014)

Existem outras métricas que não puderam ser avaliadas neste blog por se tratar de indicadores relativos a comércio eletrônico. Como o blog não promove vendas ou mesmo

campanhas, não foi gerado eventos que acionassem essas métricas. Outras métricas referentes à estrutura do blog, como por exemplo, quantas páginas têm um blog, ou mesmo dados demográficos do usuário tais como localidade, tipo de navegador e etc., não se encaixaram em nenhuma das categorias selecionadas e foram desconsideradas.

5 CONSIDERAÇÕES FINAIS

A Internet vem recebendo uma crescente atenção por parte das empresas por se apresentar como mais um veículo de comunicação de massa, onde novas formas de divulgação se mostram necessárias devido a como este espaço é percebido pelos usuários. Neste contexto, as mídias sociais surgem como espaços de interação onde pessoas revelam, mesmo que de forma fragmentada, suas necessidades e anseios.

Toda e qualquer ação realizada na Internet pode ser registrada e transformada em informação. Que, devidamente trabalhada, pode desvendar o comportamento e preferências dos usuários e assim, auxiliar as empresas, a identificar meios para se aproximar de potenciais clientes. Esse novo posicionamento, por parte das organizações, extrapola o conceito de relacionamento com o cliente e passa a modificar toda a estratégia de marketing e vendas exigindo uma gestão da informação também desses canais.

O trabalho de *web analytics*, e especialmente o monitoramento de mídias sociais, está diretamente ligado ao entendimento das organizações relativas à sua presença *online*. Para tanto, se perfaz a necessidade de controlar também, em um fluxo de informação coerente, as informações da audiência e da reputação das empresas na Internet.

Ao avaliar, de forma exploratória, o *blog* ver nas unhas utilizando a ferramenta *google analytics*, foi possível perceber, primeiramente, que a teoria de Nielsen, que sugere que cerca de 90% dos visitantes de uma página são “invisíveis”, também era um fato real naquele espaço. Muito mais do que certificar do fato, este número revela a importância de desvendar o comportamento de um visitante que praticamente não interage com o *blog*.

Esse fato reforça a necessidade do monitoramento do *blog* enquanto espaço de interações com esses usuários que naquele momento se mostram invisíveis, mas que podem ser influenciados pelo conteúdo do *blog*, ou até mesmo se tornarem consumidores daqueles produtos e serviços.

Nisso a ferramenta *google analytics* se mostra deficiente, uma vez que as métricas apresentadas são meramente quantitativas, desprovidas de significados e principalmente sem nenhuma conotação de contexto. O monitoramento de mídias sociais hoje está voltado para buscar cinco “tipos” de informação entre a interação com os usuários: (i) Informações quantitativas relativas à interação do usuário com a página (métricas frias), (ii)

Engajamento dos usuários com a marca, (iii) Influência da marca entre os usuários, (iv) alcance do canal de comunicação e (v) Capacidade de convencer os usuários a consumir a marca (conversão). O *google analytics* auxilia apenas na obtenção das métricas quantitativas não devendo assim ser usado como única fonte de informação sobre a interação de usuários e canais *online*.

Não há dúvidas que as métricas frias sejam relevantes, pois elas indicam cenários, tendências e anomalias que são identificadas e auxiliam organizações a modificar as suas estratégias. No entanto, ainda há uma visão, de certa forma enganosa, em que se acredita que apenas as métricas quantitativas são capazes de demonstrar fielmente a realidade da popularidade e consumo de uma página, sendo assim, creem que quanto maior a quantidade de números (visualização de páginas, visitantes únicos), mais bem colocada está a presença *online* da página analisada. Isso pode ser visto em artigos científicos onde bibliotecas que avaliam sua presença online baseada apenas no *google analytics* e diversos artigos da indústria apontando casos de sucesso na utilização da Internet.

Trabalhar com monitoramento de mídias sociais, não se resume apenas em utilizar uma ferramenta de gerenciamento de tráfego para saber a quantidade de visitas e usuários únicos que entram em uma página, mas sim procurar formas de efetuar análises comparativas, qualificar o tráfego no ponto de vista da conversão e métricas quentes, entender os principais caminhos utilizados pelos usuários dentro de seus sites/*blogs* e principalmente saber o propósito pelo qual sua instituição está na se posicionando na Internet.

O monitoramento de *blogs* não é capaz, sozinho, de apontar o posicionamento *online* de uma empresa ou marca. Entender este posicionamento se torna importante para uma companhia por que uma parte considerável de seu mercado consumidor está emitindo opiniões espontâneas sobre a marca/produto/serviço que podem influenciar outros potenciais consumidores e dirigir ações de marketing neste nicho pode ser mais efetivo do que ações realizadas em outros meios de comunicação. Então este monitoramento de *blogs* não consegue capturar o sentimento dos usuários e, em muitos casos, o quanto as suas opiniões estão circulando na Internet. Por isso, outras técnicas/ferramentas que são capazes de analisar as métricas quentes e mornas devem ser utilizadas em conjunto do monitoramento do *blog*.

Este nicho de mercado abre mais uma porta para profissionais da informação, que necessita identificar quais informações são necessárias para entender o *status* atual de uma empresa na Internet, identificar as fontes de informação que capturem estes dados, elaborar

meios para transformar essa informação desestruturada e distribuída em conhecimento com detalhes o suficiente para o apoio a tomada de decisão, decidir políticas e fluxos de informação do relacionamento online com o cliente e tantas outras competências que devem fazer parte da formação de profissionais da informação.

5.1 Trabalhos futuros

Esta pesquisa, por se tratar de uma pesquisa ainda embrionária, possui um conjunto de lacunas que podem ser elucidadas a partir de pesquisas futuras. O primeiro trabalho futuro é utilizar outras ferramentas que coletam informações referentes às outras atividades relativas ao monitoramento de mídias sociais, e a partir disso, sugerir um portfólio de ferramentas que supra todas as necessidades informacionais relativas ao monitoramento de mídias sociais.

Outro trabalho futuro sugerido é elaboração de referências para relacionar os dados coletados para finalidades diferentes no monitoramento. Por exemplo, podemos descobrir que por uma análise de sentimento (engajamento) os usuários estão insatisfeitos com a marca e com isso o número de visitas cai (métrica fria), ou por causa de um evento negativo (engajamento) a marca atinja usuários que não tinham acesso à mesma (alcance). Então relacionar estas métricas se mostra como outro desafio da área.

Outro trabalho futuro é identificar quais competências o profissional da informação deve possuir para trabalhar com este monitoramento de mídias sociais uma vez que este mercado é amplamente ocupado por profissionais de marketing e comunicação.

Um último trabalho futuro sugerido é mensurar o quanto este tipo de monitoramento traz de vantagem competitiva para empresas. É uma avaliação do retorno de investimento não das ações *online* em si, mas sim, de conhecer estas informações e avaliar o quanto benéfico é interagir na internet com potenciais consumidores.

REFERÊNCIAS

ABERDEEN GROUP. **Web Analytics**: Translating Clicks into Business, Aberdeen Group, Boston, MA. 2000.

ALBEE, A. Marketing Metrics: The Hard and The Soft. Disponível em:
[<http://socialmediatoday.com/ardathalbee/556522/marketing-metrics-hard-and-soft>](http://socialmediatoday.com/ardathalbee/556522/marketing-metrics-hard-and-soft).
Acesso em: 03 jan. 2014.

ALEXANDER, B. Web 2.0: A new wave of innovation for teaching and learning?
Educause Review, California, EUA. v. 41, n. 2, p. 32–44. 2006.

BARROS, M. A. Blogs e bibliotecários. Encontro Nacional dos Estudantes de Biblioteconomia, Documentação, Ciência e Gestão da informação, 27., 2004, Recife. Anais... Recife: [s.n], 2004.

BROEK, M. D. Discussion: Is Return on Influence the new ROI. 2011. Disponível em: <<http://www.theconversationmanager.com/2011/09/30/is-return-on-influence-the-new-roi/>>. Acesso em: 3 jan. 2014.

BRUNS, A. Blogs, Wikipedia, Second Life, and beyond: From production to produsage. Nova York: Peter Lang, 2008. 418 p.

CARVALHO, N. Dados, Informação e Inteligência Competitiva. Em: SILVA, T. **Para entender o Monitoramento de Mídias Sociais**. 164 p. 2012.

CIPRIANI, F. Blog Corporativo. 2006. Disponível em: <http://wiki.blogcorporativo.net/index.php>Listas/UniversityBlogs>. Acesso em: 26 nov. 2013.

DINAMIZE. Métricas para Mídias Sociais: Um Mar de Possibilidades, 2011. Disponível em: <http://www.dinamize.com/site/ebooks/ebookmidiaassociais.pdf> Acesso em: 26 abr. 2014.

E-LIFE. Blogueiros e Orkutianos: O perfil da Blogosfera brasileira. 2013b. Disponível em: <<http://www.elife.com.br/cadastropapers/?paper=blogueiros-e-orkutianos&id=1201>>. Acesso em: 27 nov. 2013.

E-LIFE. Técnicas de Coleta e Classificação em Redes Sociais. 2013. Disponível em: <<http://www.elife.com.br/cadastropapers/?paper=tecnicas-de-coleta-e-classificacao-em-redes-sociais&id=2577>>. Acesso em: 28 nov. 2013.

EIS, D. AddThis: analisando as mídias sociais do seu site. 2011. Disponível em: <<http://tableless.com.br/addthis-analisando-as-midias-sociais-do-seu-site/>>. Acesso em: 28 nov. 2013.

FAUSTINO, Paulo. Analise o impacto das mídias sociais no seu site WordPress. 2013. Disponível em: <<http://imasters.com.br/artigo/21551/wordpress/analise-o-impacto-das-midias-sociais-no-seu-site-wordpress/#section-topics>>. Acesso em: 28 nov. 2013.

HINERASKY, D. A. **O fenômeno dos blogs street-style**: do flâneur ao “star blogger”. 2012. 289 p. Tese (Doutorado em Comunicação Social) - Pontifícia Universidade Católica do Rio Grande do Sul – PUC/RS, Porto Alegre, 2012.

HONSCHA, G. L. **A profissionalização dos blogs brasileiros**: um estudo sobre as dinâmicas promocionais na blogosfera. 2009. 145 p. Dissertação (Mestrado em Biblioteconomia) - Universidade Federal do Rio Grande do Sul. 2009.

INAN, H. **Measuring the success of your Website**. New Jersey. Longman Publishing Group, 2002

JARDIM, A. D. **Aplicações de Modelos Semânticos em Redes Sociais**. 2010. 106 p. Dissertação (Mestrado em Ciências da Computação) – Universidade Católica de Pelotas UCPEL. 2010.

KRISHNAMURTHY, B.; WILLS, C. E. On the leakage of personally identifiable information via online social networks. In: **Proceedings of the 2nd ACM workshop on Online social networks**. ACM, 2009. p. 7-12.

LAINÉ, M.; FRUHWIRTH, C. Monitoring Social Media: Tools, Characteristics and Implications. In: TYRVÄINEN, P. ; JANSEN, S.; CUSUMANO, M.A. (Eds.): **Lecture Notes in Business Information Processing**. v. 51, p. 193-198. 2010.

LEMOS, A. A Cultura da Mobilidade. **Revista FAMECOS**, Porto Alegre. n. 40, dez., 2009.

MALINI, F. **Por uma Genealogia da Blogosfera**: considerações históricas (1997-2001). GT – Cibercultura e Tecnologias da Comunicação, do Inovcom, XIII Congresso de Ciências da Comunicação na Região Sudeste. **Anais...** São Paulo, 2008.

MEDIA, Monitoring Social. Social Media monitoring and the Rise of Social Customer Service. 2013. Disponível em: <<http://www.monitoring-social-media.com/social-media-monitoring-and-the-rise-of-social-customer-service>>. Acesso em: 27 nov. 2013.

MEDINA, B. A reinvenção dos meios e dos conteúdos. 2013. Disponível em: <<http://pt.slideshare.net/brauliomedina/a-reinveno-dos-meios-e-dos-contedos>>. Acesso em: 03 jan. 2014.

MICHEL, M. H. **Metodologia e Pesquisa Científica em Ciências Sociais**: um guia prático para acompanhamento da disciplina e elaboração de trabalhos monográficos. São Paulo: Atlas, 2009.

NANNO, T.; SUZUKI, Y.; FUJIKI, T.; OKUMURA, M. Automatic Collection and Monitoring of Japanese Weblogs. **Transactions of the Japanese Society for Artificial Intelligence**, v. 19, n. 6, p. 511-520. 2004.

NIELSEN, J. Participation inequality: Encouraging more users to contribute. 2006. Disponível em: <http://www.nngroup.com/articles/participation-inequality/>. Acesso em: 26 mai. 2014.

ORDUNÃ, O. **Blogs**: revolucionando os meios de comunicação. São Paulo: Thomson Learning, 2007.

PIKAS, C. K. Blog searching for competitive intelligence, brand image, and reputation management. **Online**, New York. v. 29, n. 4, p. 16-21, jul./ago. 2005.

SCHONS, C. H. O volume de informações na internet e sua desorganização: reflexões e perspectivas. **Informação e Informação**, Londrina, v.12, n.1, jan./jun.2007.

SCOBLE, R.; ISRAEL, S. Naked conversations: how blogs are changing the way businesses talk with customers. New Jersey. John Wiley & Sons, 2006. 246 p.

SILVA, T. Monitoramento de Marcas e Conversações: alguns pontos para discussão. In: DOURADO, D.; SILVA, T.; CERQUEIRA, R.; AYRES, M. (Orgs.). **#MidiasSociais: Perspectivas, Tendências e Reflexões**. Florianópolis: Bookess, 2010.

SILVA, T. Monitoramento de Mídias Sociais. In: SILVA, T. **Para entender o Monitoramento de Mídias Sociais**. 2012. p. 41.

SILVA, T. **Métricas e monitoramento de mídias sociais**. 2013. Disponível em: <<http://pt.slideshare.net/tarushijio/monitoramento-de-mdias-sociais-imerso-digital>>. Acesso em: 01 jan. 2014.

SOCIALMEDIA Monitor. Por que monitorar as Redes Sociais. 2013. Disponível em: <<http://www.socialmediamonitor.com.br/por-que-monitorar-redes-sociais>>. Acesso em: 27 nov. 2013.

SPYER, J. **Conectado**: o que a internet fez com você e o que você pode fazer com ela. Rio de Janeiro: Jorge Zahar Ed, 2007.

STATE OF BLOGOSPHERE. State of Blogosphere: Introduction and Methodology. 2011. Disponível em: <<http://technorati.com/social-media/article/state-of-the-blogosphere2011introduction/>>. Acesso em: 06 jun. 2014.

TOHMATSU, Deloitte Touche. Mídias sociais nas empresas. 2010. Disponível em: <http://www.deloitte.com/assets/DcomBrazil/LocalAssets/Documents/Estudosepesquisas/Apresentacao_MidiasSociais.pdf>. Acesso em: 25 jan. 2014.

TOFFLER, A. **A Terceira Onda**. 4. ed. Rio de Janeiro, RJ : Record, 1980.

TUTEN, T. **Advertising 2.0**: social media marketing in a web 2.0 World. Westport, Connecticut: Praeger, 2008.

WAISBERG, D.; KAUSHIK, A. Web Analytics 2.0: empowering customer centricity. **The original Search Engine Marketing Journal**, v. 2, n. 1, p. 5-11, 2009.

WELLMAN, B. HAYTHORNTHWAITE, C. **The Internet in everyday life**. Malden, Wiley-Blackwell, 2008. 588 p.

BASES METODOLÓGICAS PARA ARQUITETURA DA INFORMAÇÃO PERVERSIVA⁸

METHODOLOGICAL BASES FOR PERVASIVE INFORMATION ARCHITECTURE

Henry Poncio Cruz de Oliveira
 Silvana Aparecida Borsetti Gregorio Vidotti
 Virginia Bentes Pinto

Resumo: Atualmente a informação e os produtos tecnológicos ocupam lugar fundamental nas relações humanas e sociais. Há um cenário que resulta dos avanços ocorridos na técnica e na tecnologia permitindo uma larga produção de artefatos tecnológicos cada vez mais integrados ao cotidiano das pessoas. Emerge um movimento de integração de ambientes físicos aos ambientes de informação digital, de construção de ambientes tecnológicos híbridos e de delineamento de experiências *cross-channel*, potencializando o debate sobre a Arquitetura da Informação Pervasiva. Este estudo objetiva apresentar uma base metodológica para a Arquitetura da Informação Pervasiva, como contributo para a pesquisa e projeto de ecologias informacionais complexas. A pesquisa, bibliográfica e metodológica, utiliza abordagem qualitativa de análise e se apoia epistemologicamente na fenomenologia e no sistemismo. Apresenta como resultado uma metodologia iterativa para Arquitetura da Informação Pervasiva, baseada em sete etapas: fenomenologia do contexto, mapeamento de essências, relacionamentos complexos, definição da ecologia, padrões da ecologia, implementação e avaliação. A base metodológica apresentada é generalista e contextualista, podendo ser adaptada para ser aplicada em contextos específicos.

Palavras-chave: Arquitetura da Informação. Arquitetura da Informação Pervasiva. Metodologia. Ecologias informacionais complexas. Informação e Tecnologia.

Abstract: Nowadays, information and technological artifacts occupy key place in human and social relations. There is one scenario that results from advances in the technique and technology, enabling a wide production of technological artifacts increasingly integrated into everyday life. A movement of integration of physical environments to digital information environments, construction of hybrid technology environments and design of cross-channel experiences emerges, stimulating the debate on Pervasive Information Architecture. This study objective is to present a methodological foundation for Pervasive Information Architecture as a contribution to the research and design of complex informational ecologies. The research can be classified as literature and methodological, uses qualitative approach of analysis and is epistemologically based on phenomenology and systemism. Presents as result an iterative methodology for Pervasive Information Architecture, based in seven steps: phenomenology of context, mapping essences, complex relationships, ecology definition, ecology patterns, implementation and evaluation. The methodology presented is general and contextual, can be adapted and applied in specific contexts.

Keywords: Information Architecture. Pervasive Information Architecture. Metodology. Complex informational ecologies. Technology and Information.

⁸ Financiamentos da pesquisa: CAPES DINTER; CNPq PQ – Processo: 308443/2010-1; CNPq CHSSA – Processo: 407149/2012-0; CNPq Universal – Processo: 486147/2011-8.

1 INTRODUÇÃO

As contínuas mudanças tecnológicas intensificadas após a revolução industrial têm modificado os setores da sociedade e construído um panorama em que a informação e os artefatos tecnológicos ocupam lugar fundamental nas relações humanas e sociais. Trata-se de um cenário que resulta dos avanços ocorridos na técnica e na tecnologia, permitindo uma larga produção de artefatos tecnológicos cada vez mais integrados ao cotidiano das pessoas.

A técnica, para Agazzi (1997), diz respeito a um conjunto de conhecimentos eficazes que o homem tem expandido ao longo dos séculos para melhorar seu modo de vida, sendo tão antiga quanto a própria humanidade. Segundo Oliveira (2014), a tecnologia diz respeito às construções teóricas realizadas sobre a técnica enquanto qualidade do ato produtivo, é ciência e ao mesmo tempo ideologia da técnica, atua sobre desdobramentos maquínicos na sociedade, “é um fenômeno que engloba a técnica, os sujeitos e os produtos tecnológicos como componentes essenciais” (OLIVEIRA, 2014, p. 54).

Os avanços na técnica e na tecnologia têm gerado movimento que busca integrar os ambientes físicos aos ambientes de informação digital, por meio da criação de camadas informacionais que os intersecciona e facilita a experiência de acesso e uso destes ambientes. Há ainda os ambientes híbridos que *per se*, coligam elementos físicos e digitais estrutural e funcionalmente. Nesta linha, Oliveira (2014) sinaliza a intensificação de um fenômeno ecológico em que uma mesma informação necessita ser acessada em um ambiente físico, em um site web, por meio de aplicativo para tablets ou aplicativo para smartphone, numa televisão digital, ou em outros artefatos tecnológicos. Este fenômeno demanda reflexões arquiteturais que consideram a organização, a representação, o armazenamento e a recuperação da informação em ambientes diferentes e integrados ecologicamente (OLIVEIRA, 2014).

Porém, nos ambientes físicos, nos ambientes de informação digital ou nos ambientes híbridos, nem sempre é possível recuperar a informação que se deseja ou realizar uma atividade pretendida em função de problemas na Arquitetura da Informação (AI), fato que pode comprometer a experiência de interação. Os problemas supracitados têm sido tratados nos últimos 40 anos pela Arquitetura da Informação, usada como um contributo teórico e prático para o melhoramento do acesso e do uso da informação nos diversos tipos de ambientes. Atualmente, as reflexões teóricas e práticas concernentes à Arquitetura da Informação têm se deslocado para as experiências informacionais pervasivas em ambientes físicos, digitais ou híbridos. Trata-se do que Resmini e Rosati (2011) designam de Arquitetura da Informação Pervasiva (AIP).

Pioneiros no debate sobre a Arquitetura da Informação Pervasiva e no projeto de experiências *cross-channel*, Andrea Resmini e Luca Rosati argumentam que os problemas informacionais e tecnológicos que temos hoje precisam ser tratados por outra forma de se pensar e de se fazer Arquitetura da Informação (RESMINI; ROSATI, 2011).

Oliveira (2014) assevera que a Arquitetura da Informação Pervasiva é uma temática fronteiriça e emergente, por isso necessita de aprofundamentos teóricos e metodológicos, mas reconhece a obra *Pervasive information architecture: designing cross-channel user experiences*, publicada por Resmini e Rosati (2011), como indispensável no debate teórico, no delineamento de práticas, de metodologias e de procedimentos em Arquitetura da Informação Pervasiva. Porém adverte que no livro de Resmini e Rosati (2011) é marcadamente pragmático e não aprofunda a Arquitetura da Informação Pervasiva nos planos conceitual e metodológico.

Considera-se a Arquitetura da Informação Pervasiva um objeto/fenômeno de investigação de caráter social, histórico, marcado ideologicamente pela técnica e pela tecnologia. Baseando-se nas ideias científicas da pós-modernidade, o presente texto é resultado da pesquisa que teve com objetivo: Apresentar uma base metodológica para Arquitetura da Informação Pervasiva.

2 ARQUITETURA DA INFORMAÇÃO PERVASIVA

Tangenciando a científicidade da Arquitetura da Informação, Macedo (2005, p. 143) assevera que a Arquitetura da Informação tem características

[...] de uma disciplina que se estabelece no contexto da ciência pós-moderna. A ciência moderna pressupunha a necessidade de existência de uniformidades básicas e regularidades empíricas acerca do fenômeno que engloba o objeto de estudo de uma área, bem como a utilização de um método científico rigoroso para investigá-lo. Na pós-modernidade, como foi visto, esses pressupostos têm sido repensados. Os valores de universalidade e certeza têm dado lugar a pluralidade e a complexidade. E é nessa conjuntura que a área de Arquitetura da Informação se estabelece. Portanto, sua natureza é inherentemente interdisciplinar, e seus métodos, modelos e teorias são derivados de outras disciplinas.

Tratando a AI na perspectiva da pós-modernidade, Macedo (2005) destaca as características da interdisciplinaridade, da complexidade e do diálogo com métodos e teorias de outras disciplinas. Fundamentando a AI como disciplina científica Albuquerque (2010, p. 133) argumenta que

Como Disciplina, Arquitetura da Informação é um programa de investigação epistemológica da realidade que busca identificar e relacionar, no mínimo e necessariamente, quatro dimensões no problema tratado: Forma, Contexto, Manifestação e Significado [...].

Na contribuição de Albuquerque há um reconhecimento do caráter epistemológico da Arquitetura da Informação. O autor ainda argumenta que a AI pode adotar duas abordagens: orientada ao objeto, quando o propósito é examinar o fenômeno; ou orientada ao produto, quando o objetivo é criar/construir um produto que atue sobre o fenômeno. Expandindo a reflexão epistêmica sobre a Arquitetura da Informação Albuquerque e Lima-Marques (2011, p. 68) afirmam que

Como Disciplina, o termo Arquitetura da Informação refere-se a um esforço sistemático de identificação de padrões e criação de metodologias para a definição de espaços de informação, cujo propósito é a representação e manipulação de informações; bem como a criação de relacionamentos entre entidades linguísticas para a definição desses espaços de informação.

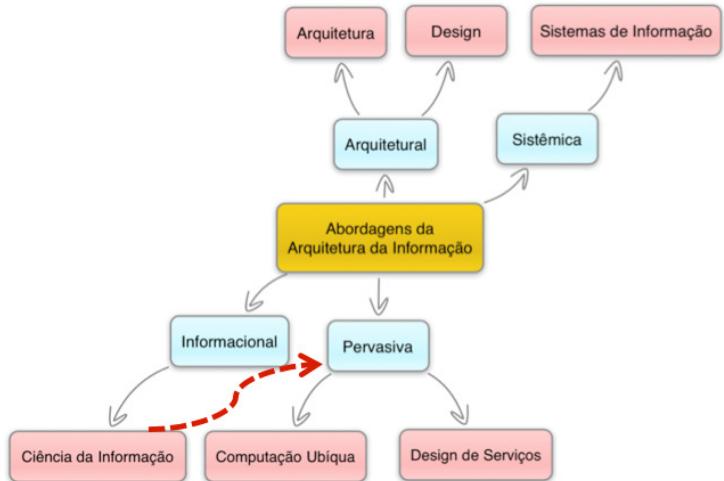
Os argumentos de Albuquerque e Lima-Marques (2011) sobre a AI, coadunam com o objetivo desta pesquisa, visto que a disciplina AI se dedica a identificação de padrões e criação de metodologias utilizáveis no projeto e na pesquisa em ambientes de informação.

Vidotti, Cusin e Corradi (2008) defendem que

Arquitetura da Informação enfoca a organização de conteúdos informacionais e as formas de armazenamento e preservação (sistemas de organização), representação, descrição e classificação (sistema de rotulagem, metadados, tesauro e vocabulário controlado), recuperação (sistema de busca), objetivando a criação de um sistema de interação (sistema de navegação) no qual o usuário deve interagir facilmente (usabilidade) com autonomia no acesso e uso do conteúdo (acessibilidade) no ambiente hipermídia informacional digital. (VIDOTTI; CUSIN; CORRADI, 2008, p.182).

A Arquitetura da Informação tem se constituído historicamente como uma práxis profissional e como um setor de estudos preocupado na solução de problemas arrolados ao acesso e uso do vasto quantitativo de informações (RESMINI; ROSATI, 2011). Os autores León (2008), Resmini e Rosati (2011) e Oliveira (2014), além de reconhecer a característica interdisciplinar da AI, identificam visões ou abordagens regulares que orientaram as pesquisas científicas e as práticas em AI desde seu estabelecimento. Oliveira (2014, p. 77) adverte que “a ação teórica e prática da Arquitetura da Informação não se cristalizou em uma perspectiva única, não se fundamentou na influência de apenas um ramo disciplinar e epistemológico”. Dialogando com León (2008) e Resmini e Rosati (2011), Oliveira destaca que no transcurso da AI houve contribuições importantes da Arquitetura e Design de Informação, dos Sistemas de Informação, da Ciência da Informação e, mais recentemente, da Computação Ubíqua e do Design de Serviço, conforme apresentado na FIGURA 1.

FIGURA 1 – Abordagens da Arquitetura da Informação



Fonte: Oliveira (2014, p. 83)

A FIGURA 1 indica que existem processos de estabelecimento, de manutenção e de ruptura de abordagens que norteiam as atividades científicas e práticas na AI (OLIVEIRA, 2014). León (2008), Resmini e Rosati (2011) e Oliveira (2014) perfilham a preponderância de campos disciplinares como Arquitetura, Design, Ciência da Informação, Sistemas de Informação, Computação Ubíqua e Design de Serviços na história da AI, mas a interdisciplinaridade fez a AI dialogar com setores disciplinares como a Ergonomia, a Usabilidade, a Psicologia Organizacional, a Ciência da Computação, as Ciências Cognitivas, o Desenho Industrial, o Design Gráfico, a Sociologia, a Antropologia, entre outras (MACEDO, 2005; DILLON, 2003).

Oliveira (2014, p. 83) reconhece que Arquitetura, Design, Ciência da Informação, Sistemas de Informação, Computação Ubíqua e Design de Serviços “quebram as fronteiras disciplinares e trazem a AI uma fluidez identitária” típica na pós-modernidade.

A abordagem arquitetural, fortemente vinculada aos primeiros anos da AI, foi influenciada pela Arquitetura e Design, sendo Wurman (1996) o representante desta abordagem. A abordagem Sistêmica firmou-se por meio da Teoria Geral dos Sistemas e das demandas de criação de sistemas de informação para os contextos organizacionais.

A Abordagem Informacional tem como representante Morville e Rosenfeld (2006) que se esforçaram para impregnar na AI os estudos clássicos da Biblioteconomia e da Ciência da Informação, esta abordagem tem sido largamente aplicada em ambientes de informação digital da web até os dias atuais. Em diálogo teórico com Resmini e Rosati (2011), Oliveira (2014) nomeou a abordagem mais recente da AI de Pervasiva, em função

das demandas informacionais e tecnológicas que articulam de forma integrada distintos espaços, ambientes, tecnologias e pessoas.

O autor supracitado esclarece que

A percepção de novos e desafiadores problemas relacionados à Arquitetura da Informação suscitou uma revisão do que já estava posto sobre a AI. Nesse contexto, uma nova forma de enxergar a AI tem sido cultivada de forma significativa pelos italianos Andrea Resmini e Luca Rosati, com destaque para livro *Pervasive information architecture: designing cross-channel user experiences*, publicado pelos referidos autores em 2011 (OLIVEIRA, 2014, p. 108).

O QUADRO 1, elaborado com base nas contribuições de Resmini e Rosati (2011), sinaliza uma série de mudanças que potencializaram o surgimento da Arquitetura da Informação Pervasiva.

QUADRO1 – Fatos relacionados à emergência da Arquitetura da Informação Pervasiva

Arquiteturas da informação tornam-se ecossistemas	Quando diferentes mídias e contextos estão entrelaçados com força, nenhum artefato pode ficar como uma entidade isolada. Cada artefato torna-se um elemento de um ecossistema maior. Todos esses artefatos têm vários <i>links</i> ou relacionamentos uns com os outros e têm de ser concebidos como parte de um processo único e contínuo de experiência do usuário.
Os usuários se tornam mediadores	Os usuários são mediadores nos ecossistemas e produzemativamente novos conteúdos ou corrigem o conteúdo existente. A tradicional distinção entre autores e leitores, ou produtores e consumidores, torna-se sutil, a ponto de ser inútil e vazia de significado.
Estático se torna dinâmico	O papel ativo desempenhado pelos usuários/mediadores torna os conteúdos eternamente inacabados, em constante mudança, perpetuamente abertos ao aperfeiçoamento e manipulação.
Dinâmico se torna híbrido	Essas novas arquiteturas abraçam diferentes domínios (análogicos, digitais e híbridos), diferentes tipos de entidades (dados, itens físicos e pessoas) e os diferentes meios de comunicação. Todas as experiências são experiências <i>bridge</i> ou <i>cross-media</i> , abrangendo diferentes ambientes.
Horizontal prevalece sobre vertical	Em arquiteturas abertas e em constante mudança, os modelos hierárquicos (<i>top down</i>) são difíceis de manter e de dar suporte, como os usuários empurram as arquiteturas em direção à espontaneidade, às estruturas efêmeras ou temporárias de significado e à mudança constante.
Design de produto se torna o design da experiência	Quando cada artefato, seja o conteúdo, produto ou serviço, é parte de um ecossistema maior, o foco muda de como criar itens únicos para como criar experiências de processo.
Experiências se tornam experiências cross-media	Múltiplas experiências em ponte conectadas a mídias, a ambientes e ecologia ubíqua, um processo único onde todas as partes contribuem para uma experiência de usuário global e sem emendas.

Fonte: Adaptado de Resmini e Rosati (2011, p. 52 a 54, tradução nossa)

As mudanças supracitadas, são vetores contribuintes para o surgimento da AIP, concebida por Oliveira (2014, p. 166) como sendo [...] uma abordagem teórico-prática da

disciplina científica pós-moderna Arquitetura da Informação, trata da pesquisa científica e do projeto de ecologias informacionais complexas. Busca manter o senso de localização do usuário na ecologia e o uso de espaços, ambientes e tecnologias de forma convergente e consistente. Promove a adaptação da ecologia à usuários e aos novos contextos, sugerindo conexões no interior da ecologia e com outras ecologias. Facilita a interação com conjuntos de dados e informações ao considerar os padrões interoperáveis, a acessibilidade, a usabilidade, as qualidades semânticas e a encontrabilidade da informação, portanto deve buscar bases na Ciência da Informação.

Adotamos o conceito de Arquitetura da Informação Pervasiva elaborado por Oliveira (2014) em função de sua amplitude e de suas bases na Ciência da Informação, tal elaboração abrange a) o status científico da AIP quando classifica-a como abordagem teórico-prática da disciplina científica pós-moderna Arquitetura da Informação; b) a definição do objeto da AIP, no caso as ecologias informacionais complexas; b) as ações que de pesquisa e projeto que a AIP realiza sobre seu objeto; c) o que a AIP objetiva e d) como a AIP se materializa em uma pesquisa ou projeto ecológico.

3 ASPECTOS METODOLÓGICOS DA PESQUISA

A pesquisa científica é, na compreensão de Demo (1995, p. 34), uma atividade do cotidiano, “uma atitude, um questionamento sistemático crítico e criativo, mais a intervenção competente na realidade, ou o diálogo crítico permanente com a realidade em sentido teórico e prático”. No caso desta pesquisa, a atitude crítica e questionamento sistemático foi realizada sobre a Arquitetura da Informação Pervasiva, no sentido de produzir uma intervenção que resulte numa base metodológica utilizável por Arquitetos da Informação e pesquisadores em Ciência da Informação.

Tendo como base o objetivo de apresentar um aparato metodológico para a AIP, a presente investigação pode ser classificada como uma pesquisa metodológica. Para Moresi (2003, p. 9), a “pesquisa metodológica é o estudo que se refere à elaboração de instrumentos de captação ou de manipulação da realidade. Está, portanto, associada a caminhos, formas, maneiras, procedimentos para atingir determinado fim”.

Como se trata de uma pesquisa metodológica, se faz necessário explicitar a compreensão de metodologia que norteou esta pesquisa. O termo metodologia é concebido entre os metodólogos com significados variados. Diversos são os termos empregados nas pesquisas para se referir ao processo de “preocupação instrumental” e das “formas de se fazer Ciência”. Dependendo do tipo de pesquisa e da ciência envolvida, podem ser

utilizados os termos metodologia, método, materiais e métodos, procedimentos, aspectos procedimentais, entre outros.

Recorrendo as raízes etimológicas, a metodologia diz respeito “ao estudo dos caminhos, dos meios de uma teoria” (BAQUERO, 2009, p, 22). A contribuição de Gomes (1998) é significativa para a presente pesquisa, pois este autor comprehende a metodologia como sendo o estudo analítico e crítico dos métodos de investigação e de prova, ou ainda como sendo a descrição, análise e avaliação crítica dos métodos científicos.

Lopes (2005, p. 93) diferencia método e metodologia a partir das expressões metodologia **da** pesquisa e metodologia **na** pesquisa: metodologia **da** pesquisa indica “a investigação ou teorização prática da pesquisa científica”, por outro lado, a metodologia **na** pesquisa indica “o trabalho com os métodos empregados”. De forma complementar, Baquero (2009) assevera que a metodologia trata da lógica interna da investigação, já os métodos constituem uma série de passos que o investigador segue no processo de construção de um conhecimento sobre um determinado objeto.

A presente pesquisa se faz via metodologia **da** e **na** pesquisa. Enquanto pesquisa executada no âmbito da Ciência da Informação, recorre ao trabalho com os métodos que devem ser empregados para atingir o objetivo da investigação, ou seja, metodologia **da** pesquisa. Mas também trata da metodologia **na** pesquisa, na medida que visa produzir um conhecimento metodológico aplicável aos processos de Arquitetura da Informação Pervasiva, considerando aspectos epistemológicos.

Retornando-se aos aspectos metodológicos adotados na presente pesquisa, os dados e informações coletadas para a construção da base metodológica para a AIP foram extraídos da literatura da área de Ciência da Informação, Tecnologia da Informação e Arquitetura da Informação e foram analisados numa perspectiva qualitativa. Para tanto, recorreu-se a livros e capítulos de livros, artigos de periódicos científicos, dissertações de mestrado e teses de doutorado, de modo que esta pesquisa pode ser classificada como bibliográfica.

Do ponto de vista procedural, a pesquisa foi dividida em duas etapas. Na primeira etapa se investigou a existência de conceitos e metodologias para a AIP por meio da revisão de literatura. Após a revisão detalhada da literatura, adotou-se a atividade de aprofundamento conceitual da AIP, feita por Oliveira (2014) como insumo para a elaboração de uma trilha metodológica que viabilize a pesquisa científica e o projeto em Arquitetura da Informação Pervasiva.

A construção da base metodológica se deu via desenho fenomenológico-sistêmico, ou seja, articulando-se a epistemologia fenomenológica com a epistemologia sistêmica, tal

escolha se alinha ao conceito de AIP proposto por Oliveira (2014), que defende a fenomenologia como adequada para tratar da AIP visto que fornece as relações entre os fenômenos e suas essências. Husserl é considerado o fundador da corrente fenomenológica (MARTINS; THEÓPHILO, 2009). Nesta corrente os fenômenos são “[...]objetos de estudo da fenomenologia, o instrumento é a intuição e busca-se compreender a essência dos fenômenos sem separar sujeito e objeto.” (MARTINS; THEÓPHILO, 2009, p. 44).

Em complemento a epistemologia fenomenológica, adotou-se a abordagem sistemática que segundo Martins e Theóphilo (2009), alimenta-se da Teoria Geral dos Sistemas de Ludwig Von Bertalanffy (BERTALANFFY, 1975). Sobre o sistemismo Demo (1985, p. 111) acrescenta:

[...] a teoria sistemica aproveitou tudo o que veio do campo da informática, na qualidade de instrumentos capazes de detecção de conflitos, de elaboração de respostas adequadas, de planejamento integrado, de controle de processos, de avaliação de projetos, e assim por diante.

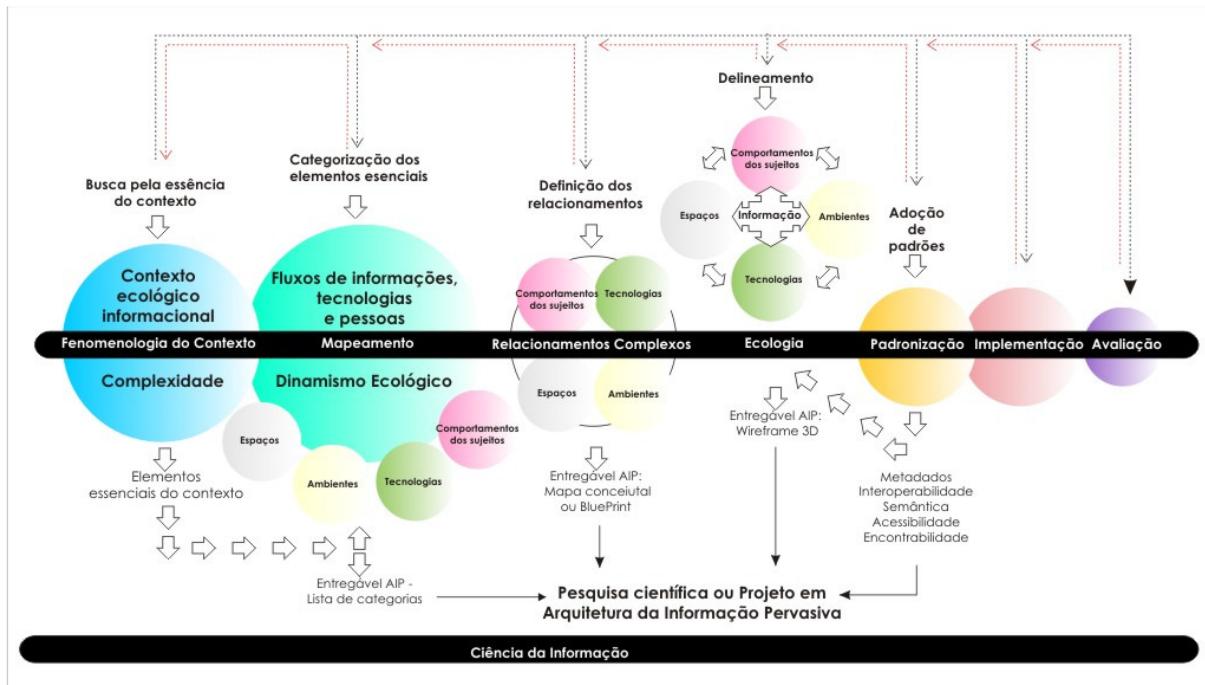
Adotou-se fenomenologia e sistemismo como um caminho articulado para se chegar ao objetivo da presente pesquisa que é apresentar um aparato metodológico utilizável em pesquisas científicas e no projeto de Arquiteturas Informacionais Pervasiva.

4 UMA BASE METODOLÓGICA PARA A ARQUITETURA DA INFORMAÇÃO PERVASIVA

No intuito de possibilitar a investigação científica em Arquitetura da Informação Pervasiva e a orientação aos arquitetos da informação na práxis de projetar, acompanhar e avaliar a Arquitetura da Informação Pervasiva em uma ecologia informacional complexa, elaborou-se uma base metodológica derivada do aprofundamento conceitual realizado por Oliveira (2014).

O que apresentamos é o produto da reflexão sobre critérios e métodos para construção e investigação de ecologias informacionais complexas pautadas pela Arquitetura da Informação Pervasiva, por meio de sete etapas iterativas que objetivam delinear a) a fenomenologia do contexto, b) o mapeamento de essências; c) os relacionamentos complexos; d) a definição da ecologia; e) os padrões da ecologia; f) a implementação e g) a avaliação. A FIGURA 2 é uma representação visual da base metodológica.

FIGURA 2 - Base metodológica para Arquitetura da Informação Pervasiva



Fonte: Oliveira (2014, p. 173)

A FIGURA 2 representa uma metodologia se vista como “[...] um conjunto de procedimentos metódicos de uma ou mais ciências acoplada a análise filosófica de tais procedimentos [...]” (ABBAGNANO, 2007, p. 650). Trata-se de um conjunto de procedimentos técnicos de averiguação ou verificação à disposição de determinada disciplina ou grupo de disciplinas viabiliza a prática de projetos, acompanhamento e avaliação de ecologias informacionais complexas, em termos arquiteturais. De forma complementar, trata da análise crítica das técnicas de investigação empregadas nos contextos científicos de pesquisa em Arquitetura da Informação Pervasiva.

As ecologias informacionais complexas são os objetos/fenômenos tratáveis pela abordagem Pervasiva da AI (OLIVEIRA, 2014). Por esta razão, a fenomenologia é adotada como um mecanismo epistêmico capaz de equilibrar complexidade e simplicidade para se chegar ao que é essencial e indispensável na ecologia informacional complexa. A investigação científica e o projeto de uma ecologia informacional complexa deve ser norteado por uma exaustiva busca das condições indispensáveis à existência e ao funcionamento da ecologia, sendo necessário chegar às categorias que, se ausentes, tornam a ecologia instável.

A fenomenologia do contexto, apresentada na Figura 2, deve ser realizada sobre o contexto ecológico complexo através dos procedimentos de redução fenomenológica e eidética até obter o *eidos*, ou elementos essenciais da ecologia.

Do ponto de vista procedural, esta etapa pode ser executada através de três passos conforme apresentados no QUADRO 2.

QUADRO 2 - Fenomenologia do Contexto

Passo	Descrição
Coleta de dados e informações sobre a ecologia informacional complexa	O pesquisador e/ou arquiteto da informação pode recorrer a relatórios técnicos e científicos que tratem da ecologia como um todo e também das partes da ecologia (espaços, ambientes, tecnologias, artefatos tecnológicos, necessidades e comportamentos dos sujeitos). Execução de <i>Briefing</i> . Aplicação de questionários. Realização de entrevistas semiestruturadas com amostra de sujeitos da ecologia. Execução de <i>Brainstorms</i> . Coleta em fontes de dados abertos, entre outros.
Tratamento dos dados e informações	A coleta dos dados e informações relacionadas à ecologia poderá gerar um grande volume de dados/informações. Assim, o pesquisador e/ou arquiteto da informação poderá realizar um procedimento de crítica dos dados, no sentido de selecionar no banco de dados/informações, os elementos de natureza conceitual, descritiva e relacional, para serem usados na etapa subsequente.
Redução fenomenológica e eidética	Nesse passo, o pesquisador e/ou arquiteto da informação poderá aplicar os princípios fenomenológicos de redução para atingir o <i>eidos</i> da ecologia informacional. Do ponto de vista procedural, é necessário abandonar juízos apriorísticos sobre a ecologia para percorrer os dados/informações e selecionar aquilo que for essencial e indispensável como: conceitos, características e possíveis relações entre as partes da ecologia. Esses elementos conceituais podem ser colocados em separado para serem categorizados na segunda etapa.

Fonte: Extraído de Oliveira (2014, p. 176)

A segunda etapa do processo iterativo trata do mapeamento de espaços, ambientes, aparatos tecnológicos e sujeitos. Após a fenomenologia do contexto aparecerão os elementos constituintes da ecologia informacional, compreendidos como essenciais ao funcionamento e à manutenção da ecologia. Na primeira etapa se gera um rol de categorias desmontadas e desconexas, que devem ser mapeadas, listadas e organizadas na segunda etapa. Tal categorização permite que o pesquisador e/ou arquiteto da informação tenha ciência de quantos e quais são os espaços informacionais, os ambientes informacionais, as tecnologias, os artefatos tecnológicos componentes da ecologia, o perfil e também os comportamentos dos sujeitos. Esta etapa deve produzir entregáveis sob a forma de listas e mapas de categorias.

Num terceiro momento são construídos os relacionamentos complexos entre partes da ecologia e ainda se gera possibilidades de conexão com outras ecologias. Se a segunda etapa gerou a categorização e os esclarecimentos sobre os espaços, os ambientes, as tecnologias e artefatos tecnológicos, sobre os sujeitos e seus comportamentos, a terceira etapa estabelecerá as relações entre os achados. Podem ser utilizados nesta etapa softwares para mapas conceituais, mapas mentais, construção de *blueprints*, entre outras possibilidades.

Executadas as etapas anteriores, se tem condições de, na quarta etapa, fazer o delineamento da ecologia, permitindo que ela seja visualizada em todas as suas dimensões. Será possível ainda representá-la através de conceitos, de visualidades, de mapas conceituais, ou outros recursos que mostrem ao pesquisador e/ou arquiteto da informação a amplitude e as exigências do projeto. Nesse ponto será possível visualizar a camada informacional responsável por conectar as partes da ecologia, gerando um todo complexo e dinâmico e sendo possível perceber a abordagem sistêmica. *Wireframes* são possíveis entregáveis para esta etapa, porém os softwares disponíveis no mercado para a construção de *wireframes* trabalham em duas dimensões, insuficientes para representar a ecologia informacional complexa em nível de relatório de pesquisa científica ou de projeto prático. Um *wireframe* gerado para esta etapa deve ser tridimensional para possibilitar a visualização e a navegação no *wireframe* enquanto representação da ecologia.

A quinta etapa diz respeito a adoção de Padrões que possibilitem o funcionamento da ecologia. Considerando a complexidade e funcionamento da ecologia, os espaços informacionais, os ambientes, as tecnologias e os artefatos tecnológicos da ecologia devem ser pensados em termos de padrões que efetivem os fluxos informacionais e de sujeitos na ecologia. Devem ser considerados os padrões de metadados, padrões de interoperabilidade, padrões semânticos e padrões de acessibilidade para que se efetive a encontrabilidade da informação na ecologia. A propósito, a encontrabilidade da informação é um “[...] elemento que se situa entre as funcionalidades de um ambiente informacional tradicional, digital ou híbrido e as características dos sujeitos [...]” (VECHIATO, 2013, p. 169).

Executados os passos anteriores, a sexta etapa trata da implementação e será responsável pela consolidação do funcionamento ecológico. Nesta etapa serão produzidos os softwares, hardwares, adaptações nos ambientes físicos no intuito de possibilitar o funcionamento da ecologia.

A sétima etapa apresentada nem sempre encerra o processo pois a metodologia é iterativa e permite o retorno a quaisquer das etapas anteriores. O projeto da ecologia ou a

investigação científica necessita ser avaliada, o pesquisador e/ou arquiteto da informação deve inferir a necessidade de finalizar sua atividade ou iniciar uma nova iteração, ou quantas forem necessárias, para refinar o processo e possibilitar o funcionamento pervasivo da ecologia informacional. Podem ser utilizados testes de software, testes de usabilidade, testes de acessibilidade, avaliação da recuperação da informação, avaliação da conexão entre partes heterogêneas da ecologia, entre outros. Por fim, destacamos a Ciência da Informação como campo capaz de nortear o processo iterativo aqui delineado como uma base metodológica para a Arquitetura da Informação Pervasiva.

5 CONSIDERAÇÕES FINAIS

Nesta pesquisa adotou-se uma articulação epistemológica entre fenomenologia e sistemismo como caminho para se chegar aos objetivos de uma pesquisa de natureza metodológica que trata da Arquitetura da Informação Pervasiva. A fenomenologia permitiu encontrar as categorias essenciais da base metodológica aqui apresentada. A ação efetiva da fenomenologia permitiu que ela fosse inserida nas etapas iniciais conforme apresentado na Figura 2. O sistemismo foi responsável por finalizar a elaboração da base metodológica para a AIP, tendo como base princípios de integração entre as partes da ecologia e o todo, gerando a possibilidade de integração de uma ecologia informacional complexa com outras.

Outras abordagens epistemológicas podem ser utilizadas para tratar de pesquisas relacionadas a Arquitetura da Informação Pervasiva, inclusive trazendo outros pontos de vista e enriquecendo o debate sobre Arquitetura da Informação.

O aparato metodológico apresentado neste estudo é de natureza teórica e necessita ser aplicado em contextos pragmáticos. Apoiando-se nos princípios da pós-modernidade, este aparato metodológico pretende ser generalista e contextualista ao mesmo tempo. É um aparato metodológico generalista na medida que articula categorias essenciais ao funcionamento ecológico: a) a fenomenologia do contexto, b) o mapeamento de essências; c) os relacionamentos complexos; d) a definição da ecologia; e) os padrões da ecologia; f) a implementação e g) a avaliação, o que permite sua aplicação em quaisquer contextos ecológicos complexos. Por outro lado, sua plasticidade permite que as etapas sejam alteradas ou que novas sejam inseridas em função dos contextos ecológicos específicos, o que permite classificar o aparato metodológico como contextualista.

Atualmente, a base metodológica para Arquitetura da Informação Pervasiva apresentada neste trabalho está sendo aplicada em uma pesquisa apoiada pelo CNPq que objetiva investigar a aplicação da Arquitetura da Informação Pervasiva no processo editorial de livros eletrônicos.

REFERÊNCIAS

- ABBAGNANO, Nicola. **Dicionário de Filosofia**. 5. ed. São Paulo: Martins Fontes, 2007.
- AGAZZI, Evandro. **El impacto epistemológico de la tecnología**. Génova: Universidad Génova, 1997. Disponível em: <<http://www.argumentos.us.es/numero1/agazzi.htm>>. Acesso 25 ago. 2013.
- ALBUQUERQUE, Alfram Roberto Rodrigues de. **Discurso sobre fundamentos de Arquitetura da Informação**. 2010. 241f. Tese (Doutorado em Ciência da Informação) – Faculdade de Ciência da Informação. Universidade de Brasília, Brasília, 2010.
- ALBUQUERQUE, Alfram Roberto Rodrigues de; LIMA-MARQUES, Mamede. Sobre os fundamentos da Arquitetura da Informação. **Perspectivas em Gestão & Conhecimento**, João Pessoa, v. 1, p. 60-72, 2011.
- BAQUERO, Marcello. **Pesquisa quantitativa nas Ciências Sociais**. Porto Alegre: UFRGS Editora, 2009.
- BERTALANFFY, Ludwig Von. **Teoria geral dos sistemas**. São Paulo: Vozes, 1975.
- DEMO, Pedro. **Introdução à metodologia da ciência**. São Paulo: Atlas, 1985.
- DEMO, Pedro. Metodologia Científica em Ciências Sociais. São Paulo: Atlas, 1995.
- DILLON, Andrew. **Information Architecture why, what & when?** 2003. Disponível em: <<http://www.asis.org/Conferences/Summit2000/dillon/>>. Acesso em: 26 nov. 2011.
- GOMES, Willinan. **Fenomenologia e Pesquisa em Psicologia**. Porto Alegre: Editora UFRGS, 1998.
- LEÓN, Rodrigo Ronda. Arquitectura de Información: análisis histórico-conceptual. **No sólo usabilidad Journal**, n. 7, Abr. 2008. Disponível em: <http://www.nosolousabilidad.com/articulos/ai_cc_informacion.htm>. Acesso em: 20 abr. 2013.
- LOPES, Maria Immacolata Vassallo de. **Pesquisa em Comunicação**. São Paulo: Edições Loyola, 2005.
- MACEDO, Flávia Lacerda Oliveira de. **Arquitetura da informação: aspectos epistemológicos, científicos e práticos**. 2005. 190 f. Dissertação (Mestrado em Ciência da Informação) – Faculdade de Ciência da Informação, Universidade de Brasília, Brasília, 2005.
- MARTINS, Gilberto de Andrade; THEÓPHILO, Carlos Renato. **Metodologia da investigação científica para Ciências Sociais Aplicadas**. São Paulo: Atlas, 2009.
- MORESI, Eduardo (Org.). **Metodologia da pesquisa**. Brasília: Universidade Católica de Brasília, 2003. Disponível em: <http://http://www.unisc.br/portal/upload/com_arquivo/metodologia_da_pesquisa.pdf>. Acesso em: 9 out. 2013.
- MORVILLE, Peter; ROSENFELD, Louis. **Information for Architecture for the Word Wide Web**. 3. ed. Sebastopol, CA: O'Reilly, 2006.

OLIVEIRA, Henry Poncio Cruz de. **Arquitetura da informação pervasiva: contribuições conceituais.** 2014. 202 f. Tese (Doutorado em Ciência da Informação) – Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília, 2014.

RESMINI, Andrea; ROSATI, Luca. **Pervasive information architecture:** designing cross-channel user experiences. Burlington: Elsevier, 2011.

VECHIATO, Fernando Luiz. **Encontrabilidade da informação: contributo para uma conceituação no campo da ciência da informação.** 2013. 206 f. Tese (Doutorado em Ciência da Informação) – Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília, 2013.

VIDOTTI, Silvana Aparecida Borsetti Gregório; CUSIN, César Augusto; CORRADI, Jiliane Adne Mesa. Acessibilidade digital sob o prisma da Arquitetura da Informação. In: GUIMARÃES, José Augusto Chaves; FUJITA, Mariângela Spotti Lopes. **Ensino e pesquisa em Biblioteconomia no Brasil:** a emergência de um novo olhar. São Paulo: Cultura Acadêmica, 2008.

WURMAN, Richard Saul. **Information Architects.** Zurich: Graphis Press Corp, 1996.

ENCONTRABILIDADE DA INFORMAÇÃO: ATRIBUTOS E RECOMENDAÇÕES PARA AMBIENTES INFORMACIONAIS DIGITAIS⁹

FINDABILITY OF INFORMATION: ATTRIBUTES AND GUIDELINES TO DIGITAL INFORMATIONAL ENVIRONMENTS

Fernando Luiz Vechiato

Silvana Aparecida Borsetti Gregorio Vidotti

Resumo: Em um primeiro momento, apresenta as bases epistemológicas e teóricas que sustentam a inserção de um conceito operatório para a encontrabilidade da informação no campo da Ciência da Informação. Em um segundo momento, define os principais atributos e recomendações para a encontrabilidade da informação, com vistas à sua compreensão sob uma perspectiva prática, a partir do respaldo epistemológico e teórico apresentado. Por meio de um estudo teórico, exploratório, bibliográfico e documental, os resultados revelam que epistemologicamente a encontrabilidade da informação se insere na Ciência da Informação no paradigma pós-custodial, que privilegia o acesso à informação e é significativamente influenciado pelas Tecnologias de Informação e Comunicação. As concepções teóricas desse conceito se concentram na evolução da World Wide Web e na perspectiva das dimensões da linguagem, em que se verifica a pragmática como tendência, visto que alia o desenvolvimento tecnológico às ações praticadas por sujeitos informacionais e institucionais em ambientes informacionais digitais. Tendo em vista que essas ações podem viabilizar a encontrabilidade da informação, esses sujeitos são entendidos como mediadores. Portanto, o conceito de mediação infocomunicacional é base para a encontrabilidade da informação. Para sua aplicação em ambientes informacionais digitais, foram definidos atributos e recomendações que devem ser considerados no contexto da Arquitetura da Informação. Os atributos caracterizam o conceito e as recomendações sinalizam ações práticas para o projeto e para a avaliação da encontrabilidade da informação. As concepções epistemológicas, teóricas e práticas abordadas alicerçam o referido conceito operatório na Ciência da Informação, e deve ser mais explorado neste campo científico porque o objetivo de qualquer ambiente informacional é ampliar as possibilidades de encontro da informação pelos sujeitos informacionais.

Palavras-chave: Encontrabilidade da Informação. Arquitetura da Informação. Mediação Infocomunicacional. Ambientes Informacionais Digitais. Informação e Tecnologia.

Abstract: At first, this paper presents the epistemological and theoretical bases that support the inclusion of an operational concept for the findability of information in the field of Information Science. In a second moment, it defines the key attributes and recommendations for the findability of information, with a view to their understanding in a practical perspective, from the epistemological and theoretical support presented. From a theoretical, exploratory, bibliographic and documentary study, the results reveal that epistemologically the findability of information fits in Information Science in post-custodial paradigm that focuses on access to information and is significantly influenced by Information and Communication Technologies. The theoretical conclusions of this concept focus on the evolution of the World Wide Web and from the perspective of the dimensions

⁹ Financiamentos da pesquisa: CAPES PDSE – Processo: 18914-12-8; CAPES DS; CNPq PQ – Processo: 308443/2010-1; CNPq CHSSA – Processo: 407149/2012-0; CNPq Universal – Processo: 486147/2011-8.

of language, where there is as pragmatic trend, as technological development that combines the actions taken by informational and institutional individuals in digital information environments. Given that these actions turn enable or not the findability of information, these individuals can be understood as mediators. Therefore, the concept of infocommunicational mediation is the basis for the findability of information. For its application in digital information environments, were defined attributes and guidelines which should be considered in the context of Information Architecture. The attributes characterizing the concept and guidelines point at practical actions for the design and for evaluation of the findability of information. Epistemological, theoretical and practical concepts addressed underpin the called operative concept in information science and should be further explored in this scientific field, because the purpose of any information environment is to expand the possibilities of finding the information by the informational individuals.

Keywords: Findability of Information. Information Architecture. Infocommunicational Mediation. Digital Informational Environments. Information and Technology.

1 INTRODUÇÃO

O contexto sociotécnico e pragmático hodierno resulta da evolução das Tecnologias de Informação e Comunicação (TIC) e das ações infocomunicacionais empreendidas por diferentes mediadores em ambientes informacionais digitais, sejam estes sujeitos informacionais¹⁰ e/ou institucionais¹¹.

Ambos os aspectos contribuem significativamente para ampliar as possibilidades de organização, de disseminação, de recuperação, de acesso e de uso da informação. Nesse sentido, há uma contribuição mútua entre humanos e máquinas/tecnologias, tendo em vista que um transforma o outro no decorrer dos processos que compõem o fluxo infocomunicacional.

A partir dessas premissas, entendemos que a busca pelo efetivo diálogo entre sistemas de informação e sujeitos informacionais potencializa as possibilidades de encontro da informação em ambientes informacionais digitais.

O termo encontrabilidade, traduzido do inglês *findability*, foi definido preliminarmente por Peter Morville (2005a) em uma perspectiva mais técnica que

10 Neste texto, utilizamos o termo ‘sujeitos informacionais’ em substituição ao termo ‘usuários’, considerando a complexidade que envolve o sujeito e sua interação com recursos informacionais em sistemas e ambientes de informação. Partimos da concepção de Assis e Moura (2013) para caracterizar os ‘sujeitos informacionais’: “Por sujeito informacional entende-se um sujeito social que manifesta a sua subjetividade através do estabelecimento de identidades e percursos informacionais na web. Ele é visto como um sujeito social pragmático, uma vez que constrói suas relações pela via da linguagem e do compartilhamento de significados. Tal fenômeno marca a passagem de um usuário passivo em busca de recursos que atendam às suas necessidades de informação para um sujeito ativo e dinamizador dos fluxos informacionais [...].” (ASSIS; MOURA, 2013, p.86).

11 Entendemos por ‘sujeitos institucionais’ aqueles que atuam nos contextos institucionais como os profissionais de informática e os profissionais da informação.

científica. Considerando seu diálogo com os estudos relacionados à recuperação da informação na Ciência da Informação e sua relevância para os ambientes informacionais, mormente os digitais, buscamos sustentar um conceito operatório para a encontrabilidade a fim de ser incorporado no bojo do referido campo científico.

Por meio de estudo teórico, exploratório, bibliográfico e documental, este trabalho teve como objetivo: em um primeiro momento, apresentar as bases epistemológicas e teóricas que alicerçem a incorporação do conceito de encontrabilidade da informação na Ciência da Informação, construídas tendo como ponto de partida as concepções técnicas de Morville (2005a); e, em um segundo momento, definir os principais atributos e recomendações para a encontrabilidade da informação, com vistas à sua compreensão sob uma perspectiva prática.

2 ENCONTRABILIDADE (*FINDABILITY*) À LUZ DE PETER MORVILLE: UMA PERSPECTIVA TÉCNICA

Nos últimos anos, os avanços tecnológicos ampliaram as possibilidades de recuperação, de acesso, de uso e de apropriação da informação a partir da ligação e do relacionamento entre dados e informações em diferentes dispositivos e sistemas, além da colaboração dos sujeitos informacionais na produção e na organização da informação e do conhecimento em ambientes informacionais digitais.

Diante desse cenário sociotécnico, Peter Morville apresentou o termo *findability* em seu livro *Ambient Findability* no ano 2005, em que demonstra a facilidade que temos de encontrar qualquer informação a respeito de algo ou alguém a partir de qualquer local e a qualquer momento. (MARCOS, 2007).

A tradução do termo *findability* para a língua portuguesa não é consensual entre os autores. Passaremos nesse momento a utilizar o termo em português ‘encontrabilidade’, a exemplo de Landshoff (2011), com vistas à sua sustentação e consolidação nas pesquisas desenvolvidas no Brasil.

Na referida publicação, Morville (2005a, p.4, tradução nossa) conceitua encontrabilidade como:

- a) A qualidade de ser localizável ou navegável;
- b) O grau no qual um determinado objeto é facilmente descoberto ou localizado;
- c) O grau no qual um sistema ou ambiente suporta a navegação e recuperação.

A partir dos conceitos de Morville (2005a), percebemos que a encontrabilidade ocorre a partir da busca de informação por meio da navegação em um *web site* ou por meio das estratégias de pesquisa lançadas em um mecanismo de busca (*search engine*).

Em relação aos mecanismos de busca, alguns *web sites* são mais facilmente encontrados a partir de um mecanismo de busca externo do que a partir de seus próprios recursos de pesquisa (SPAGNOLO *et al.*, 2010). Ademais, os sujeitos nem sempre iniciam suas buscas a partir da *home page* de um *web site* específico e é preciso, portanto, tornar todas as suas páginas encontráveis a partir de qualquer mecanismo de pesquisa externo. (MORVILLE, 2005b). Nesse sentido, Morville (2005b) sugere investir em estratégias de *Search Engine Optimization* (SEO).

Outro aspecto que precisa ser considerado em relação à busca é que nem sempre a informação encontrada é derivada de uma busca prévia, podendo ocorrer a descoberta acidental de informação, visto que os sujeitos podem encontrar accidentalmente algo sem estar necessariamente buscando no momento da navegação ou da busca, fato este que modifica seu comportamento.

Em algumas aplicações *Web*, a pesquisa pode não ser o suficiente, sendo necessário investir na navegação, considerando o entendimento do conteúdo pelo sujeito informacional quando não está a procurar por um item específico. Por meio da navegação, é possível sugerir caminhos e *affordances* (pistas) aos sujeitos com vistas à exploração de conteúdos que eles não sabem que existem (SPAGNOLO *et al.*, 2010), o que pode propiciar a descoberta acidental da informação, conforme mencionado anteriormente.

A expressão ‘*ambient findability*’ que empresta o título ao livro de Morville (2005a) alia a internet e a computação ubíqua. A informação está “nas nuvens” e isso modifica nossas mentes fisicamente. E é nesse ambiente que testamos nossa capacidade de selecionar as fontes pertinentes que satisfaçam nossas necessidades (MORVILLE, 2005a; MORVILLE; SULLENGER, 2010).

Destarte, entendemos que a definição de encontrabilidade, além da navegação e da busca em sistemas e ambientes, bem como dos aspectos que delineiam as características dos sujeitos informacionais, alia também mobilidade, convergência e ubiquidade provenientes do desenvolvimento tecnológico, considerando as ações humanas para a busca do conhecimento em um determinado ambiente que possui características analógicas e digitais.

Vale mencionar que Peter Morville é um dos autores do livro *Information Architecture for the World Wide Web* que, junto a Louis Rosenfeld, propõe uma metodologia para o projeto da Arquitetura da Informação (AI) de ambientes informacionais

digitais, além de um conjunto de sistemas que formam a anatomia da AI, considerando as dimensões contexto, conteúdo e usuários (MORVILLE; ROSENFELD, 2006).

O desenvolvimento da AI para o ambiente *Web* é resultado dos esforços em considerar os estudos da Biblioteconomia na era da Internet e divulgar seus princípios e suas práticas no âmbito da *User eXperience (UX)* e do *Web Design*, conforme salienta Morville (2005b). Para o autor a encontrabilidade é o problema de grande parte dos *web sites* e a Arquitetura da Informação é a solução.

Logo, entendemos que a Arquitetura da Informação é o caminho para a encontrabilidade e esta, por sua vez, não está apenas associada ao projeto de sistemas e ambientes informacionais, mas sim à capacidade que esses sistemas conferem em prover a informação adequada aos sujeitos, considerando as características, as limitações e as competências que eles trazem consigo no processo de busca de informação.

Batley (2007) considera que os profissionais da informação já aplicam a AI em sua prática profissional, mormente no tocante à encontrabilidade. Argumenta que o foco da capacitação desses profissionais é aumentar a possibilidade dos sujeitos informacionais em encontrar a informação por meio da representação e da organização da informação e do fornecimento de ferramentas e assistência para promover sua encontrabilidade.

A encontrabilidade atua tanto no âmbito de um *web site* específico, sendo a AI a principal responsável pela encontrabilidade via navegação, quanto via mecanismos de busca, o que possibilita às páginas do *web site* serem encontradas a partir desses mecanismos, o que sugere, conforme mencionado, a utilização de estratégias de SEO. Essas estratégias estão mormente relacionadas ao uso de metadados, ligações e recursos que facilitem a indexação das páginas disponíveis pelos robôs de busca.

Embora estes sejam alguns dos caminhos para a encontrabilidade, destacamos neste trabalho aspectos que podem subsidiar a encontrabilidade em ambientes informacionais, especialmente os digitais. Para enfocar esses aspectos, que se traduzem em atributos e recomendações, se faz necessário apresentar um conceito operatório para encontrabilidade no contexto da Ciência da Informação, conforme será evidenciado na próxima seção, alicerçado em bases epistemológicas e teóricas que possam sustentar sua inserção no âmbito desta Ciência.

3 ENCONTRABILIDADE DA INFORMAÇÃO: UMA PERSPECTIVA CIENTÍFICA

3.1 Bases epistemológicas

Revisitando o percurso histórico-evolutivo da Ciência da Informação, observamos que, embora sua formação esteja atrelada ao contexto científico e tecnológico do período pós-Segunda Guerra Mundial, os conceitos-chave que a fundamentam e os elementos da prática profissional surgem anteriormente a esse momento. Este campo científico se desenvolveu a partir de características que revelam dois paradigmas: o primeiro custodial, patrimonialista, historicista e tecnicista, que está a ser gradualmente substituído por um novo paradigma, pós-custodial, informacional e científico (SILVA; RIBEIRO, 2002).

Essa perspectiva paradigmática foi construída pelos pesquisadores portugueses Armando Malheiro da Silva e Fernanda Ribeiro a partir de um resgate histórico, considerando desde os primórdios da escrita até as possibilidades colaborativas que nos permitem os ambientes informacionais digitais, perpassando pela prática profissional no âmbito dos arquivos e bibliotecas e pelos eventos que sinalizaram a constituição da Ciência da Informação, reconhecendo, portanto, os antecedentes históricos, teóricos e práticos que sustentaram a formação deste campo científico e que abarcam os dois paradigmas.

Podemos sintetizar como principais características do paradigma custodial: a ênfase na guarda, conservação e restauro do suporte informacional, sendo base da *práxis* de arquivistas e bibliotecários; a preservação da cultura erudita e da memória; o aumento da importância do acesso aos conteúdos informacionais por meio de instrumentos de pesquisa; e a autonomia equivocada das disciplinas que constituem os fazeres dos arquivistas e bibliotecários (SILVA; RIBEIRO, 2002).

O movimento que vem culminando na crise do paradigma custodial se inicia com as realizações dos advogados belgas Paul Otlet e Henri La Fontaine, se fortalece com a explosão informacional nos anos 1950 e se efetiva com o advento da internet, especialmente da *Web* nos anos 1990. A partir desses eventos, percebemos que a evolução tecnológica que permeia a humanidade contribui para a evolução dos recursos, serviços, sistemas e ambientes informacionais e potencializam as possibilidades de acesso à informação.

Dessa forma, um novo paradigma passa a ser desenhado no decorrer do século XX, um paradigma pós-custodial, informacional e científico, que possui como características: a ênfase na informação como fenômeno humano e social, independente do suporte informacional; o dinamismo informacional em antinomia ao imobilismo documental; a

prioridade no acesso à informação, visto que “[...] só o acesso público justifica e legitima a custódia e a preservação” (SILVA; RIBEIRO, 2010, p.41); e a ascensão de um objeto científico apropriado a esse paradigma – a informação social, que transforma as bases teóricas e empíricas tratadas no âmbito de um campo científico social, bem como a *práxis* dos profissionais da informação (SILVA; RIBEIRO, 2002).

O paradigma pós-custodial não negligencia a custódia, a memória e a preservação que caracterizam seu paradigma antecessor. Tem como premissa a ênfase no acesso e, portanto, os sujeitos e seus comportamentos, competências e habilidades passam a ter a importância necessária e esperada para o projeto de sistemas e ambientes informacionais.

Nessa perspectiva, as TIC trouxeram inúmeros desafios que precisam ser discutidos recorrentemente no campo da Ciência da Informação. O acesso e o uso da informação devem ser enfatizados, visto que o usuário da *Web* é um ser difuso e, portanto, se torna difícil delinear um perfil. Os sujeitos informacionais, nesse momento, passam a ser determinantes a partir de suas habilidades para buscar, selecionar e avaliar os recursos informacionais disponíveis, trazendo preocupações mais relacionadas à *Information Literacy* (SILVA; RIBEIRO, 2010), sendo este um dos aspectos também abordados por Morville (2005a) em seu livro.

Diante dessa perspectiva epistemológica e paradigmática, o delineamento de um conceito para a encontrabilidade da informação na Ciência da Informação se insere no escopo do paradigma pós-custodial. A convivência entre humanos e máquinas faz emergir nesse contexto uma concepção sociotécnica que enriquece esse novo paradigma.

A noção de agenciamentos maquínicos sustenta filosoficamente essa convivência. É contextualizada na Ciência da Informação pela pesquisadora brasileira Silvana Drumond Monteiro para a compreensão das relações e o alinhamento das TIC, dos sistemas de signos e da cognição por meio da conexão e da hibridização entre homem (e/ou sociedade) e máquina (e/ou técnica), contribuindo para uma perspectiva não esvaziadora da tecnologia (MONTEIRO, 2007; MONTEIRO, 2012; MONTEIRO, ABREU, 2009) que ainda acomete este campo científico.

Um conceito operatório para a encontrabilidade da informação se enquadra nessas bases epistemológicas e se insere na seara dos estudos referentes às TIC, que devem ser consideradas como um dos objetos de estudo da Ciência da Informação, conforme sugerem Santos e Vidotti (2009), com o mesmo grau de relevância em relação, por exemplo, à área de Organização da Informação e do Conhecimento, devido à necessidade de construção de

seu status ontológico no bojo desse campo científico na busca por concepções teóricas e metodológicas próprias e apropriadas.

3.2 Bases teóricas

A encontrabilidade da informação, enquanto conceito, emerge da evolução tecnológica, mormente da *Web*. Não apenas as tecnologias e as práticas informacionais subsidiaram a evolução da *Web*, mas também suas possibilidades de linguagem.

Destarte, a evolução tecnológica da *Web*, tendo como respaldo os estudos da linguagem abordados por alguns autores da Ciência da Informação, possibilita a percepção a respeito da linguagem e de como ela influenciou as ações dos sujeitos que impulsionaram tal desenvolvimento.

Retomando a noção de agenciamentos maquínicos, Monteiro (2012) a comprehende a partir do conceito de dobra de Leibniz, ressignificado por Deleuze (1991). Em *A dobra: Leibniz e o barroco*, Gilles Deleuze explica que a dobra é uma das características do barroco e que este possui uma função operatória, não remetendo apenas a uma essência. O barroco “[...] não pára de fazer dobras. [...] ele curva e recurva as dobras, leva-as ao infinito, dobra sobre dobra, dobra conforme dobra. O traço do barroco é a dobra que vai ao infinito” (DELEUZE, 1991, p.13).

Em termos de linguagem, a dobra é o sentido que permanece entre as coisas e a linguagem, é a conexão entre homem, máquina e linguagem, por quanto o próprio signo é uma dobra, pois pode se dobrar, se desdobrar e se redobrar em várias semióticas e tecnologias, considerando que a dobra mais simples de um signo é seu significante/significado (MONTEIRO, 2012). A *Web* possui várias dobras semióticas, visto que ela é uma só mas se desdobra em diferentes momentos tendo em vista suas possibilidades de linguagem, compondo a cartografia do ciberespaço.

Para caracterizar as dobras semióticas da *Web*, estudamos sua evolução a partir das dimensões da linguagem, quais sejam a sintaxe, a semântica e a pragmática, conforme explica Jorente (2012, p.106):

A sintaxe, no sentido semiótico do termo, lida com as propriedades formais dos signos e símbolos, e a semântica lida com a relação entre os signos e suas designações. A pragmática lida com os aspectos psicológicos, biológicos e sociológicos que ocorrem no funcionamento dos signos.

Nessa perspectiva, delineiam-se as seguintes dobras da *Web*: a *Web Sintática*, a *Web Semântica* e a *Web Pragmática*.

A *Web Sintática* agrupa as tecnologias que caracterizam a estrutura estática e organizacional da *Web 1.0* com as tecnologias e práticas colaborativas da *Web 2.0* (ou *Web Social*). Para encontrarmos determinada informação são utilizadas as palavras/termos que constam nas páginas, independente de serem organizadas e/ou disponibilizadas de forma *top-down* (*Web 1.0*) ou *bottom-up* (*Web 2.0*), e sem consultar as respectivas descrições que interpretam os significados das palavras. A *Web Sintática* compreende até hoje os ambientes informacionais digitais disponíveis, mormente em relação à sua estrutura, ou seja, páginas *Web* e recursos inter-relacionados via hiperlinks e possibilidades de colaboração (KOO, 2011; SANTAELLA, 2012).

A *Web Semântica* nasceu como possibilidade de trazer significado aos recursos disponíveis na *Web* por meio de tecnologias semânticas, metadados, ontologias e agentes inteligentes. É também denominada *Web 3.0* e *Web de Dados*. O uso deste último termo parece mais apropriado, pois, conforme argumenta Andrade (2012), a semântica utilizada para designar a *Web Semântica* está mais para a semântica formal, ligada à sintaxe e à lógica, a qual difere da semântica da linguística.

Também denominada *Web Social Semântica* e *Web 4.0*, a *Web Pragmática* alia as tecnologias semânticas à colaboração social, o que possibilita agregar contexto às informações de acordo com o uso social da linguagem, considerando, portanto, as características e o comportamento dos sujeitos informacionais (ANDRADE, 2012; KOO, 2011).

Percebemos, a partir das reflexões apresentadas, que a *Web* evoluiu condicionada ao desenvolvimento tecnológico, mas também ao uso social da linguagem, cuja máxima ocorre hodiernamente com a mediação praticada pelos sujeitos informacionais no ciberespaço.

Dessa forma, os sujeitos informacionais na rede, junto aos profissionais da informação e projetistas/desenvolvedores de sistemas (sujeitos institucionais), podem ser considerados mediadores (MALHEIRO; RIBEIRO, 2011), porque contribuem para promover a encontrabilidade da informação nos ambientes colaborativos que se edificam no ciberespaço.

A mediação no âmbito da Ciência da Informação pode apoiar um conceito operatório para a encontrabilidade da informação, considerando o paradigma pós-custodial, envolto em concepções sociotécnicas e pragmáticas.

Portanto, a mediação na Ciência da Informação, em uma ênfase pós-custodial, abrange todos os processos informacionais, desde a produção até a apropriação da informação, com destaque para uma mediação cumulativa (MALHEIRO; RIBEIRO, 2011),

consequência do desenvolvimento tecnológico e das possibilidades de interação, de colaboração e de compartilhamento no ciberespaço, onde coexistem diferentes mediadores, como os profissionais da informação, os profissionais de informática e os sujeitos informacionais.

Todavia, sugerimos que, quanto houvesse a necessidade em se direcionar a mediação, *a priori*, para o objeto informação, a mediação sempre está envolvida em um processo de comunicação, ou como Martins (2010) formula, ‘nos espaços híbridos da comunicação’, derivando daí, uma ‘mediação infocomunicacional’.

Este termo nos parece apropriado para compreender a mediação no âmbito da Ciência da Informação e todos os seus componentes interdisciplinares provenientes da área de Comunicação e que podem contribuir para os estudos em TIC.

3.3 Encontrabilidade da informação: conceito operatório

O conceito de encontrabilidade apresentado por Morville (2005a) está relacionado a uma abordagem mais técnica que científica, tornando-se necessário reconfigurá-lo para ser discutido e incorporado na Ciência da Informação.

Destarte, a partir das premissas epistemológicas e teóricas discutidas, e com base em Vechiato (2013), entendemos que: ‘Encontrabilidade da informação’ sustenta-se fundamentalmente nas funcionalidades de um ambiente informacional e nas características dos sujeitos psico-sociais.

Relacionada aos processos que compõem o fluxo infocomunicacional, desde a produção até a apropriação da informação, a encontrabilidade da informação deriva-se dos princípios da Arquitetura da Informação e da Mediação Infocomunicacional e tem como elemento fundamental a Intencionalidade dos sujeitos nas ações informacionais empreendidas durante o processo de comunicação que, inclusive, subsidiam a elaboração de técnicas e de tecnologias para a organização, representação da informação e recuperação da informação.

As bases epistemológicas e teóricas apresentadas para a inserção do conceito ‘encontrabilidade da informação’ na Ciência da Informação, junto à perspectiva técnica da encontrabilidade (*findability*) abordada por Peter Morville, forneceram subsídios para a definição de atributos e recomendações de encontrabilidade da informação que são apresentados na próxima seção, os quais estão diretamente relacionados ao referido conceito operatório.

4 ATRIBUTOS E RECOMENDAÇÕES PARA A ENCONTRABILIDADE DA INFORMAÇÃO EM AMBIENTES INFORMACIONAIS DIGITAIS

4.1 Atributos de Encontrabilidade da Informação

Os atributos de encontrabilidade da informação, apresentados nas subseções que seguem, são entendidos como características que potencializam as possibilidades de encontro da informação pelos sujeitos num sistema ou ambiente informacional.

4.1.1 Taxonomias Navegacionais

As taxonomias navegacionais estão associadas à organização da informação que propicia ao sujeito encontrar a informação por meio da navegação. Desse modo, auxiliam na descoberta de informações. São utilizadas em estruturas de organização *top-down*.

Aquino, Carlan e Brascher (2009) problematizam que as taxonomias elaboradas em ambientes informacionais digitais nem sempre se preocupam com a lógica na organização da informação, o que pode dificultar a encontrabilidade da informação, pois compromete a busca e a navegação.

A elaboração das taxonomias para a navegação em ambientes informacionais digitais deve ser apoiada nos seguintes aspectos:

Categorização: relaciona-se ao estabelecimento de categorias gerais e suas respectivas subcategorias baseadas em definições consistentes e de fácil entendimento, para que possam ser rapidamente compreendidas pelos usuários.

Controle terminológico: diz respeito à escolha dos termos adequados para representar os conceitos, de forma objetiva, evitando problemas como imprecisão e ambigüidade. Serão consideradas situações de sinonímia, polissemia, emprego de siglas, abreviaturas, e termos em outros idiomas, uma vez que podem comprometer a comunicabilidade das taxonomias.

Relacionamento entre os termos: enfoca a hierarquização, a qual assume grande relevância, já que esse é o principal elemento responsável pela navegação do usuário e é a base de qualquer sistema classificatório. Assim, a estrutura da taxonomia deve demonstrar claramente a subordinação entre os níveis hierárquicos. Uma outra forma de relacionamento entre os termos são as referências cruzadas que normalmente ocorrem, no ambiente *web*, por meio da utilização de links.

Multidimensionalidade: orienta-se à análise da capacidade da taxonomia permitir que um termo possa estar em mais de uma categoria, de acordo com o contexto (AQUINO; CARLAN; BRASCHER, 2009, p.207-208).

Para auxiliar na construção de taxonomias em ambientes informacionais digitais, o método mais adequado é o *card sorting*. É empregado para descobrir a representação ou o modelo mental que os sujeitos informacionais elaboram sobre um conjunto de itens de informação pretendidos para um determinado ambiente, visando sua categorização. Os itens

são descritos em fichas de papel e espalhados sobre uma mesa. Cada participante organiza as fichas em grupos, segundo suas próprias perspectivas. Em seguida, as combinações são avaliadas e discutidas, podendo gerar análises quantitativas e qualitativas (FARIA, 2010).

4.1.2 Folksonomias

As *folksonomias* estão relacionadas à organização social da informação que propicia ao sujeito a classificação de recursos informacionais, bem como encontrar a informação por meio da navegação (uma nuvem de *tags*, por exemplo) ou dos mecanismos de busca, ampliando as possibilidades de acesso. São utilizadas em estruturas de organização *bottom-up*. Quando associadas aos vocabulários controlados e às tecnologias semânticas, potencializam as possibilidades de constituição de uma *Web Pragmática*.

A *folksonomia*, por sua vez, é considerada como uma classificação social (MORVILLE; ROSENFIELD, 2006) e tem como uma de suas características o não controle terminológico. Todos os instrumentos utilizados para organização da informação, como taxonomias, tesouros e ontologias, se alicerçam nos termos utilizados por uma determinada comunidade ou domínio, mas partem da linguagem natural para se constituírem como linguagem controlada, eliminando as ambiguidades, por exemplo. Todavia, na *folksonomia*, os termos atribuídos pelos sujeitos permanecem atrelados à linguagem natural, o que interfere significativamente na recuperação e na encontrabilidade da informação.

Embora houvesse preocupações a respeito da dificuldade na recuperação da informação a partir das *folksonomias*, já nos é claro que as *tags* atribuídas pelos sujeitos estão auxiliando inclusive na elaboração dos instrumentos tradicionais para organização da informação.

4.1.3 Wayfinding

O termo ‘*wayfinding*’ foi utilizado primeiramente pelo arquiteto Kevin Lynch, em 1960, em seu livro “*The image of the city*” para descrever o conceito de legibilidade ambiental, ou seja, os elementos do ambiente construído como mapas, identificação das ruas, sinais direcionais e outros dispositivos que nos permitem “navegar” em espaços complexos como as cidades (MORVILLE, 2005a; LYNCH; HORTON, 2009). O conceito tem sido investigado por biólogos, antropólogos e psicólogos para descrever o comportamento de homens e animais em ambientes naturais e artificiais (MORVILLE, 2005a).

O *wayfinding* possui quatro componentes principais: orientação (*orientation*), que indica o posicionamento no ambiente; decisões de rota (*route decisions*), que indicam os

possíveis lugares que podem ser explorados; mapeamento mental (*mental mapping*), relacionado às experiências de lugares já explorados que permitem mapear o espaço e tomar decisões de lugares a explorar; e encerramento (*closure*), que permite reconhecer se o lugar de destino é realmente o lugar certo (MORVILLE, 2005a; LYNCH, HORTON, 2009).

É possível relacionar o *wayfinding* dos ambientes físicos com a navegação na *Web*, entretanto esta se constitui um tipo especial de ambiente que pode não fornecer as pistas espaciais e uma navegação concreta comparados ao caminhar por uma cidade. Isto é, enquanto a navegação na *Web* possua muitas semelhanças com o mundo físico, a experiência do sujeito não pode contar com paisagens que sirvam como marcos no seu caminhar entre as páginas (LYNCH, HORTON, 2009).

No decorrer da navegação, os caminhos possíveis são elaborados mentalmente pelo sujeito. Nesse momento, o princípio da consistência, também presente como heurística da Arquitetura da Informação Pervasiva (RESMINI; ROSATI, 2011), se torna relevante para que ele consiga se movimentar no ambiente. A estrutura hierárquica de um *web site*, por exemplo, é familiar ao sujeito. Elementos explícitos como uma trilha de navegação o permite compreender qual o caminho percorrido (LYNCH, HORTON, 2009).

Lynch e Horton (2009) propõem recomendações para a orientação espacial em *web sites* a partir dos elementos elencados por Kevin Lynch:

- Caminhos: elaborar caminhos de navegação consistentes;
- Regiões: criar uma identidade única, consistente, mas considerando a especificidade de cada local (ou categoria de assunto);
- Nós: não confundir o sujeito com muitas opções na *home page* e no menu;
- Marcos: utilizar marcos consistentes na navegação visando a orientação do sujeito no espaço.

Lynch e Horton (2009) também atentam que as formas de encontrar informação em um *web site* ocorre ou por meio da navegação ou via recurso de busca no próprio *site*. Além disso, o *web site* é encontrado por meio dos mecanismos de busca, como o *Google*, mas como não se sabe ao certo qual página será recuperada, é preciso criar marcos para que o sujeito se oriente dentro dele, independentemente do nível de profundidade em que estiver. Como os sujeitos têm utilizado mecanismos de busca como portais para encontrar informação, a quantidade de sujeitos que acessam preliminarmente as *home pages* está diminuindo cada vez mais.

Percebemos que a concepção de *wayfinding* aplicada à *Web* é redutora em relação à complexidade do estudo em ambientes físicos. Destarte, o *wayfinding* se torna um caminho profícuo para os estudos da AI Pervasiva.

4.1.4 *Affordances*

As *affordances* caracterizam um importante atributo da interface com o sujeito. Em trabalho anterior, conforme estudos realizados, definimos *affordance*¹² como um princípio de usabilidade, relacionado aos incentivos e pistas atribuídos ao sistema que proporcionam aos sujeitos a realização de determinadas ações (VECHIATO; VIDOTTI, 2012).

Todavia, a partir dos estudos de encontrabilidade da informação, entendemos que as *affordances* também estão associadas à encontrabilidade, visto que fornecem subsídios para o encontro da informação.

Na navegação, as *affordances* auxiliam na orientação espacial (*wayfinding*), podendo ser utilizados: metáforas, trilhas de navegação, priorização da informação mais significativa, utilização adequada de elementos estéticos entre outros. Nos mecanismos de busca, conforme a anatomia da busca de Morville e Callender (2010), as *affordances* também auxiliam na elaboração da *query* (como no recurso autocomplete / autosugestão, por exemplo) e na escolha pela informação adequada nos resultados de busca.

4.1.5 *Descoberta de informações*

A descoberta de informações está condicionada aos demais atributos de encontrabilidade da informação no que diz respeito às facilidades que a interface (navegação e/ou mecanismos de busca) oferece para encontrar a informação adequada às necessidades informacionais do sujeito.

A descoberta accidental de informação, por sua vez, se refere às necessidades que estão em segundo plano, mas que, por intermédio da interação, podem ser priorizadas.

¹² O psicólogo americano James Gibson elaborou a Teoria das Affordances a partir de seus estudos no campo da percepção visual. Para Gibson, “[...] affordances são possibilidades de ação que o ambiente [ou objeto] oferece ao agente. Apesar de parecer uma definição simples, há características do conceito envolvidas nesta definição, que necessitam ser destacadas. Uma delas está baseada na idéia de que durante a interação com o ambiente, o agente percebe as possibilidades de ação e não as qualidades do ambiente. Além disso, a captação de tais possibilidades depende da escala corporal e das capacidades de ação do agente” (OLIVEIRA, 2005, p. 90-91).

4.1.6 Acessibilidade e Usabilidade

Com a aplicação da Arquitetura da Informação na Ciência da Informação, foi necessário aliar os estudos de usabilidade e de acessibilidade para garantir que os sistemas e ambientes informacionais digitais proporcionassem o acesso equitativo da informação (acessibilidade) no âmbito do público-alvo estabelecido nos meandros de um projeto com facilidades inerentes ao uso da interface e da informação disponibilizada (usabilidade), viabilizando, portanto, uma experiência do sujeito satisfatória.

A facilidade de acesso e de uso das interfaces nos mais diversos tipos de sistemas e ambientes informacionais propicia a encontrabilidade da informação.

4.1.7 Mediadores

A mediação dos informáticos está associada ao desenvolvimento de sistemas, dispositivos, bancos de dados e interfaces com utilização de linguagens computacionais, com vistas à gestão e à recuperação da informação.

A mediação dos profissionais da informação, por sua vez, está associada a todos os processos informacionais em todos os tipos de ambientes informacionais. Um bibliotecário, por exemplo, atua na seleção, na organização, na representação, no armazenamento/preservação e na disseminação da informação, bem como promove ações para facilitar a encontrabilidade da informação e a interação dos sujeitos com diferentes sistemas e ambientes informacionais, conforme corrobora Batley (2007).

Por fim, a mediação dos sujeitos informacionais está relacionada às ações infocomunicacionais que empreendem em quaisquer sistemas e ambientes informacionais, por exemplo no que diz respeito à produção e à organização da informação e do conhecimento em ambientes colaborativos, gerados a partir de seus conhecimentos, comportamento e competências que caracterizam sua Intencionalidade, conforme será abordado na próxima subseção.

4.1.8 Intencionalidade

No âmbito da Ciência da Informação, destacamos o estudo da pesquisadora brasileira Mágory Miranda (2010) que abordou a Teoria da Intencionalidade aplicada à encontrabilidade. A Intencionalidade é um termo utilizado em letra maiúscula porquanto designa um direcionamento e não um propósito ou objetivo, pois seu conceito fenomênico se aplica à teoria do conhecimento e não à teoria da ação humana (MIRANDA, 2010).

A Intencionalidade, segundo Sokolowski (2004¹³), é a doutrina nuclear da fenomenologia, ela explica que cada ato de consciência que nós realizamos, cada ato de consciência que nós temos como algo intencional, é consciência de ou experiência de algo ou outrem (MIRANDA, 2010, p.155).

A autora considera que a noção de experiência do usuário (ou sujeito informacional) está presente no fenômeno infocomunicacional por meio deste termo, visto que a Intencionalidade do sujeito é carregada de experiências, necessidades e competências (aqui entendemos tanto as informacionais quanto as tecnológicas), entendimento, cognição e satisfação, fornecendo, inclusive, subsídios para a estruturação de sistemas e ambientes informacionais.

Miranda (2010) entende que a Intencionalidade é também importante para o campo da Organização e Representação da Informação na Ciência da Informação ao ser utilizada para o entendimento do significado e da relevância da informação, sempre com enfoque no comportamento informacional dos sujeitos.

Com a Intencionalidade, podemos compreender que o sujeito informacional tem papel importante inclusive no desenvolvimento tecnológico. A colaboração dos sujeitos é fundamental para a evolução da *Web*, pois, aliada às tecnologias da *Web* dos Dados, vêm contribuindo sobremaneira para que novas perspectivas para ela sejam traçadas, delineando uma *Web Pragmática* emergente.

Além disso, com a quantidade de recursos, serviços e ambientes informacionais digitais disponíveis, são os sujeitos que estabelecem seus critérios de avaliação dessas fontes, pois têm suas experiências, conhecimento e habilidades como respaldo, o que caracteriza a *information literacy*.

Para Miranda (2010, p.273), portanto,

A intencionalidade de informação para a *findability* significa *direcionalidade* de informação e se funda na experiência de cada sujeito (*user experience* para controle na produção, organização e partilha de informação); a informação que é produzida é sempre *acerca de* e *dirigida a*, isto é, um sujeito, com a sua experiência, cria informação *acerca de* e *dirigida a* para atingir seus objetivos. E é nesse sentido, da capacidade da experiência do usuário, da consciência Intencional, que se baseia a *web* da inovação, do paradigma atual.

Morville (2005a) alia o conceito de encontrabilidade particularmente às ações de navegação e busca que a impulsionam. Entendemos que para que possa se efetivar, é necessário que o conceito abarque as características relacionadas aos sistemas e sujeitos, portanto, na nossa perspectiva, a Intencionalidade se torna um atributo fundamental para

¹³ SOKOLOWSKI, R. **Introdução à Fenomenologia**. São Paulo: Ed. Loyola, 2004.

que todo o sistema e ambiente informacional seja projetado com enfoque nas experiências e habilidades dos sujeitos informacionais.

4.1.9 Mobilidade, convergência e ubiquidade

Mobilidade, convergência e ubiquidade correspondem ao último atributo e são aspectos que permeiam os ambientes informacionais híbridos. Estão associados ao meio ambiente, externo aos sistemas e ambientes informacionais, mas que os incluem, dinamizando-os e potencializando as possibilidades dos sujeitos em encontrar a informação por meio de diferentes dispositivos e em diferentes contextos e situações.

A perspectiva da sociotecnologia da informação do pesquisador espanhol Fernando Sáez Vacas alicerça este atributo, tendo em vista que estamos a vivenciar um *Nuevo Entorno Tecnossocial* (NET), formado pelo conjunto das redes de telecomunicações, computadores e suas tecnologias (RUD¹⁴), pelas redes sociais (organizações, grupos/comunidades e indivíduos), com enfoque nos nativos digitais (FUMERO, 2007).

Esses elementos tendem a trazer a esse ambiente (entorno) novas perspectivas e ações mediadoras no ciberespaço. Desse modo, percebemos que não há segregação entre o virtual e o real ou entre o ciberespaço e o mundo físico.

4.2 Recomendações de encontrabilidade da informação

Nesta subseção, apresentamos uma lista de recomendações de encontrabilidade da informação a partir do referencial teórico apresentado e dos atributos de encontrabilidade da informação definidos, considerando a necessidade de projetar e avaliar sistemas e ambientes encontráveis. As recomendações são:

1. Elaborar taxonomias navegacionais coerentes com a terminologia utilizada pelos sujeitos;
2. Implementar recursos de classificação social (*folksonomia*) e de navegação por meio das *tags* atribuídas pelos sujeitos.
3. Investir na mediação infocomunicacional dos sujeitos.
4. Ampliar as possibilidades dos sujeitos em encontrar a informação por meio de diferentes ambientes e dispositivos.
5. Investigar os comportamentos, as competências, as experiências e a Intencionalidade dos sujeitos.

¹⁴ Sáez Vacas (2007) caracteriza a Rede Universal Digital (RUD) como um conjunto heterogêneo composto por diferentes redes: a Internet, as redes de telefonia fixa e móvel, as redes wi-fi, as redes *Global Positioning System* (GPS), as redes de energia elétrica entre outras. Para o autor, essas redes estão cada dia mais interoperáveis e formam um imenso tecido de redes, muito complexo e quase invisível.

6. Aplicar recomendações e avaliações de acessibilidade e de usabilidade.
7. Investir na utilização de *affordances* para orientar os sujeitos no espaço (*wayfinding*) no decorrer da navegação.
8. Investir na utilização de *affordances* para a *query* e os resultados de busca.
9. Proporcionar busca pragmática.

Quanto à elaboração das taxonomias navegacionais (Recomendação 1), devem ser observados os seguintes aspectos: categorização, controle terminológico, relacionamento entre os termos e multidimensionalidade (AQUINO; CARLAN; BRASCHER, 2009). A aplicação do *card sorting* com participação de sujeitos que representam o público-alvo pode ser um caminho para a construção dessas taxonomias.

Conforme mencionado, a utilização de *folksonomias* (Recomendação 2), aliada aos vocabulários controlados, contribuem para a encontrabilidade da informação. Todavia, a mediação dos sujeitos precisa ser incentivada (Recomendação 3), mormente quanto à produção e à organização da informação, neste caso, conforme também foi observado por Morville e Callender (2010) na elaboração da anatomia da busca.

As possibilidades atuais de acesso à informação por meio de diferentes dispositivos revelam a necessidade dos ambientes informacionais serem projetados para atuarem tanto em ambientes analógicos quanto digitais, bem como em todos os dispositivos a partir de tecnologias e práticas específicas, porém que mantenham consistência no nível de interação com o sujeito (Recomendação 4). As diversas possibilidades de acesso contribuem para que os sujeitos encontrem a informação mais facilmente.

As Recomendações 5 e 6 estão associadas aos estudos com sujeitos. Em estruturas de organização *top-down*, os estudos junto aos sujeitos podem auxiliar na organização e na representação da informação, bem como em aspectos formais de interface e em outros aspectos do conteúdo informacional, conduzindo o ambiente para a acessibilidade, para a usabilidade (VECHIATO; VIDOTTI, 2009) e também para a encontrabilidade da informação. Em estruturas de organização *bottom-up*, a Intencionalidade dos sujeitos pode ser utilizada para a personalização de serviços, para a contextualização dos resultados de busca etc.

As Recomendações 7, 8 e 9 mantêm relação direta com os mecanismos de busca e com a navegação. As *affordances*, conforme evidenciamos, constituem um atributo significativo a ser aplicado na interface de sistemas e ambientes informacionais e se tornam um caminho profícuo para o desenvolvimento das pesquisas em Arquitetura da Informação Pervasiva.

Momento no tocante à Recomendação 9, a busca pragmática pode contemplar os recursos autocomplete / autosugestão e *mashups*, que possibilitam ao sistema atribuir sentido e contexto à *query*, conforme Andrade e Monteiro (2012). Em paralelo, também se constituem como *affordances*, fornecendo aos sujeitos possibilidades de ações concernentes à sua necessidade de informação, o que pode proporcionar a descoberta e o encontro da informação. Destarte, quanto maiores as possibilidades de relacionamento entre a *query* e a necessidade do sujeito, bem como entre os resultados da pesquisa, maiores serão as possibilidades de encontrar a informação a partir dos mecanismos de busca e outros Sistemas de Recuperação da Informação (SRI).

5 CONSIDERAÇÕES FINAIS

Este texto apresentou aspectos epistemológicos, teóricos e práticos com vistas à incorporação do conceito ‘encontrabilidade da informação’ no escopo da Ciência da Informação, tendo como ponto de partida a concepção técnica de Peter Morville.

Epistemologicamente, a encontrabilidade da informação se insere na Ciência da Informação no paradigma pós-custodial, que privilegia o acesso à informação e é significativamente influenciado pelas Tecnologias de Informação e Comunicação. Teoricamente, o conceito ‘encontrabilidade da informação’ deriva da evolução da *Web* na perspectiva das dimensões da linguagem e do conceito de mediação infocomunicacional.

Para a aplicação da encontrabilidade da informação em ambientes informacionais digitais, foram propostos recomendações e atributos, sendo estes últimos: taxonomias navegacionais; *folksonomias*; *wayfinding*; *affordances*; descoberta de informações; acessibilidade e usabilidade; mediadores; Intencionalidade; e mobilidade, convergência e ubiquidade.

Concluímos que os ambientes informacionais digitais, se projetados com base nas perspectivas conceituais e práticas de um conceito de encontrabilidade ajustado para a Ciência da Informação e alicerçado na mediação infocomunicacional, e considerados os aspectos sociotécnicos que permeiam o projeto desses ambientes e a Intencionalidade dos sujeitos, possibilitam melhorias na recuperação, no acesso e na apropriação da informação.

REFERÊNCIAS

- ANDRADE, I. A. de. **As dimensões semântica e pragmática da web e dos mecanismos de busca no ciberespaço.** 2012. 127 f. Dissertação (Mestrado Profissional em Gestão da Informação) – Universidade Estadual de Londrina, Londrina, 2012. Disponível em: <<http://www.bibliotecadigital.uel.br/document/?code=vtls000181112>>. Acesso em: 15 maio 2013.

- ANDRADE, I. A. de.; MONTEIRO, S. D. Diretrizes para busca de informação no ciberespaço. **Informação@profissões**, Londrina, v. 1, n. 1/2, p. 187-203, jul./dez. 2012. Disponível em: <<http://www.uel.br/revistas/uel/index.php/infoprof/article/viewFile/14595/12263>>. Acesso em: 21 ago. 2013.
- AQUINO, I. J.; CARLAN, E.; BRASCHER, M. B. Princípios classificatórios para a construção de taxonomias. **Pontodeacesso**, Salvador, v. 3, n. 3, p. 196-215, dez. 2009. Disponível em: <<http://www.portalseer.ufba.br/index.php/revistaici/article/view/3626/2744>>. Acesso em: 15 jul. 2013.
- ASSIS, J. de; MOURA, M. A. Folksonomia: a linguagem das tags. **Encontros bibli**: revista eletrônica de biblioteconomia e ciência da informação, v. 18, n. 36, p. 85-106, jan./abr. 2013. Disponível em: <<https://periodicos.ufsc.br/index.php/eb/article/view/1518-2924.2013v18n36p85/24523>>. Acesso em: 10 ago. 2013.
- BATLEY, S. Information architecture for information professionals. Oxford: Chandos Publishing, 2007.
- DELEUZE, G. **A dobrA**: Leibniz e o barroco. Campinas: Papirus, 1991.
- FARIA, M. M. de. Card sorting: noções sobre a técnica para teste e desenvolvimento de categorizações e vocabulários. **Revista digital de biblioteconomia e ciência da informação**, Campinas, v. 7, n. 2, p. 1-9, jan./jun. 2010. Disponível em: <<http://www.sbu.unicamp.br/seer/ojs/index.php/rbc1/article/view/436/297>>. Acesso em: 15 jul. 2013.
- FUMERO, A. M. Contexto sociotécnico. In: FUMERO, A.; ROCA, G. **Redes Web 2.0**. Fundación Orange, 2007. p. 8-64. Disponível em: <http://www.fundacionorange.es/areas/25_publicaciones/WEB_DEF_COMPLETO.pdf>. Acesso em: 12 fev. 2011.
- JORENTE, M. J. V. **Ciência da informação**: mídias e convergência de linguagens na web. São Paulo: Cultura Acadêmica, 2012.
- KOO, L. C. **Web 3.0**: impacto na sociedade de serviços: uma análise da comunicação contemporânea. 2011. 228 f. Tese (Doutorado em Comunicação e Semiótica) – Pontifícia Universidade Católica de São Paulo, São Paulo, 2011. Disponível em: <http://www.sapientia.pucsp.br/tde_busca/arquivo.php?codArquivo=13429>. Acesso em: 10 mar. 2013.
- LANDSHOFF, R. **Findability**: elementos essenciais para as formas de encontro da informação em bibliotecas digitais. 2011. 127 f. Dissertação (Mestrado em Tecnologias da Inteligência e Design Digital) – Pontifícia Universidade Católica de São Paulo, São Paulo, 2011. Disponível em: <http://www.sapientia.pucsp.br//tde_busca/arquivo.php?codArquivo=13775>. Acesso em: 15 maio 2013.
- LYNCH, P. J.; HORTON, S. **Web style guide**: basic design principles for creating web sites. 3. ed. 2009. Disponível em: <<http://webstyleguide.com/wsg3/>>. Acesso em: 24 maio 2013.

MALHEIRO, A.; RIBEIRO, F. Paradigmas, serviços e mediações em Ciência da Informação. Recife: Néctar, 2011.

MARCOS, M. Information architecture & findability: Peter Morville interview. **El profesional de la información**, v. 16, n. 3, p. 268-269, maio/jun. 2007. Disponível em: <<http://www.elprofesionaldelainformacion.com/contenidos/2007/mayo/12.pdf>>. Acesso em 04 jun. 2014.

MARTINS, A. A. L. **Mediação:** reflexões no campo da ciência da informação. 2010. 253 f. Dissertação (Mestrado em Ciência da Informação) – Escola de Ciência da Informação, Universidade Federal de Minas Gerais, Belo Horizonte, 2010. Disponível em: <http://www.tdf-biblio.ufv.br/bitstream/handle/1843/ECID-88MHR9/dissertacao_ana_amelia.pdf?sequence=1>. Acesso em: 20 abr. 2013.

MIRANDA, M. K. F. de O. **O acesso à informação no paradigma pós-custodial:** da aplicação da intencionalidade para findability. 2010. 353 f. Tese (Doutorado em Informação e Comunicação em Plataformas Digitais) – Faculdade de Letras, Universidade do Porto, Porto, 2010. Disponível em: <<http://repositorio-aberto.up.pt/bitstream/10216/50422/2/tesedoutmajormiranda000112543.pdf>>. Acesso em: 15 mar. 2013.

MONTEIRO, S. D. A dobra semiótica e os agenciamentos maquínicos: por uma ontologia das tecnologias de informação e comunicação. In: CERVANTES, B. M. N. (Org.). **Horizontes da organização da informação e do conhecimento**. Londrina: EDUEL, 2012. p. 63-96.

MONTEIRO, S. D. O ciberespaço: o termo, a definição e o conceito. **Datagramazero**, v. 8, n. 3, p. 1-29, jun. 2007. Disponível em: <http://www.dgz.org.br/jun07/Art_03.htm>. Acesso em: 30 jun. 2011.

MONTEIRO, S. D.; ABREU, J. G. de. O pós-moderno e a organização do conhecimento no ciberespaço: agenciamentos maquínicos. **Datagramazero**, v. 10, n. 6, p. 1-9, dez. 2009. Disponível em: <http://www.dgz.org.br/dez09/Art_05.htm>. Acesso em: 30 jun. 2011.

MORVILLE, P. **Ambient findability**. Sebastopol: O'Reilly, 2005a.

MORVILLE, P. Libraries at the crossroads of ubiquitous computing and the internet. **Online**, v. 29, n. 6, nov./dez 2005b. Disponível em: <<http://www.infotoday.com/online/nov05/morville.shtml>>. Acesso em: 04 jun. 2014.

MORVILLE, P.; CALLENDER, J. **Search patterns:** design for discovery. Canadá: O'Reilly, 2010.

MORVILLE, P.; ROSENFELD, L. **Information architecture for the world wide web**. 3. ed. Sebastopol: O'Reilly, 2006.

MORVILLE, P.; SULLENGER, P. Ambient findability: libraries, serials, and the internet of things. **The serials librarian**, v. 58, n. 1-4, p. 33-38, 2010.

OLIVEIRA, F. I. da S. **Affordances:** a relação entre agente e ambiente. 2005. 99 f. Dissertação (Mestrado em Filosofia) – Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília, 2005. Disponível em: <http://www.athena.biblioteca.unesp.br/exlibris/bd/bma/33004110041P1/2005/oliveira_fis_me_ma.pdf>. Acesso em: 02 nov. 2013.

RESMINI, A.; ROSATI, L. **Pervasive information architecture:** designing cross-channel user experiences. Burlington: Elsevier, 2011.

SÁEZ VACAS, F. Contextualización sociotécnica de la web 2.0: vida y sociedad en el nuevo entorno tecnossocial. In: FUMERO, A.; ROCA, G. **Redes Web 2.0.** Fundación Orange, 2007. p. 96-122. Disponível em: <http://www.fundacionorange.es/areas/25_publicaciones/WEB_DEF_COMPLETO.pdf>. Acesso em: 12 fev. 2011.

SANTAELLA, L. A tecnocultura atual e suas tendências futuras. **Signo y pensamiento**, v. 30, p. 30-43, jan./jun. 2012. Disponível em: <<http://revistas.javeriana.edu.co/index.php/signoypensamiento/article/view/2408/1692>>. Acesso em: 22 maio 2013.

SANTOS, P. L. A. da C.; VIDOTTI, S. A. B. G. Perspectivismo e tecnologias de informação e comunicação: acréscimos à Ciência da Informação? **Datagramazero**, v. 10, n. 3, p. 1-10, jun. 2009. Disponível em: <http://dgz.org.br/jun09/Art_02.htm>. Acesso em: 8 jul. 2012.

SILVA, A. M. da. RIBEIRO, F. **Das <<ciências>> documentais à Ciência da Informação:** ensaio epistemológico para um novo modelo curricular. Porto: Ed. Afrontamento, 2002.

SILVA, A. M. da. RIBEIRO, F. **Recursos de informação:** serviços e utilizadores. Lisboa: Universidade Aberta, 2010.

SPAGNOLO, L.; BOLCHINI, D.; PAOLINI, P.; DI BLAS, N. Beyond findability: search-enhanced information architecture for content-intensive Rich Internet Applications. **Journal of information architecture**, v. 2, n. 1, p. 19-36, 2010. Disponível em: <<http://journalofia.org/volume2/issue1/03-spagnolo/>>. Acesso em: 12 jun. 2013.

VECHIATO, F. L. **Encontrabilidade da informação:** contributo para uma conceituação no campo da ciência da informação. 2013. 206 f. Tese (Doutorado em Ciência da Informação) – Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília, 2013. Disponível em: <http://www.marilia.unesp.br/Home/Pos-Graduacao/CienciadaInformacao/Dissertacoes/Tese_de_Doutorado_-Fernando_Luiz_Vechiato.pdf>. Acesso em: 15 jan. 2014.

VECHIATO, F. L.; VIDOTTI, S. A. B. G. Recomendações de usabilidade e de acessibilidade em projetos de ambientes informacionais digitais para idosos. **Tendências da pesquisa brasileira em ciência da informação**, v. 5, n. 1, p. 1-23, 2012. Disponível em: <<http://inseer.ibict.br/ancib/index.php/tpbci/article/view/87/132>>. Acesso em: 06 ago. 2013.

VECHIATO, F. L.; VIDOTTI, S. A. B. G. Subsídios teórico-metodológicos para a construção de ambientes informacionais digitais. In: BORGES, M. M.; CASADO, E. S. (Orgs.). **A ciência da informação criadora de conhecimento.** v. 2. Coimbra: Imprensa da Universidade de Coimbra, 2009. p. 287-299. Disponível em: <http://www.eventos-iuc.com/ocs/public/conferences/1/schedConfs/1/actas_EDIBCIC2009_2.pdf>. Acesso em: 15 jan. 2012.

CONVERSÃO DE REGISTROS EM XML PARA MARC 21: UM MODELO BASEADO EM XSLT¹⁵

*RECORD CONVERSION FROM XML TO MARC 21: A MODEL BASED ON XSLT
STYLESHEETS*

Fabrício Silva Assumpção
Plácida Leopoldina Ventura Amorim da Costa Santos

Resumo: Considerando a existência de sistemas de gerenciamento de bibliotecas que não utilizam os Formatos MARC 21 ou quaisquer outros padrões de metadados internacionalmente aceitos, este trabalho tem por objetivo elaborar um modelo para a conversão de registros de distintos padrões de metadados codificados com a *Extensible Markup Language* (XML) em registros nos Formatos MARC 21, tendo como um dos pontos centrais do modelo a utilização de folhas de estilo de transformação elaboradas com a linguagem *Extensible Stylesheet Language for Transformation* (XSLT). Após uma revisão de literatura, é apresentado o modelo e seus processos: mapeamento, elaboração de regras de conversão, elaboração da folha de estilo XSLT, verificação da folha de estilo, exportação e conversão dos registros. Como resultados alcançados, destaca-se que o modelo para a conversão de registros (1) busca apresentar uma generalização e uma formalização dos processos e dos componentes necessários à conversão de registros utilizando folhas de estilo XSLT e (2) apresenta novas possibilidades para o intercâmbio de dados que favorecem a interoperabilidade entre diferentes aplicações de informática que comunicam dados sobre recursos informacionais.

Palavras-chave: Conversão de registros. Padrões de metadados. Folhas de estilo XSLT. Formatos MARC 21. MARCXML.

Abstract: Considering the existence of integrated library systems that do not use neither MARC 21 Formats nor any other international metadata standard, this paper aims to formulate a model for conversion of records created using different metadata standards and encoded in Extensible Markup Language (XML) into MARC 21 records, by using stylesheets created with Extensible Stylesheet Language for Transformation (XSLT). After a literature review, we present the model and its process: mapping, formulation of the conversion rules, XSLT stylesheet formulation and test, and record exporting and conversion. As conclusions, we highlight that the model for record conversion (1) aims to present a generalization and formalization of the processes and components needed for record conversion using XSLT stylesheets and (2) presents new possibilities for the data interchange and supports interoperability between and among different computer applications that communicate data about information resources.

Keywords: Record conversion. Metadata standards. XSLT stylesheets. MARC 21 Formats. MARCXML.

1 INTRODUÇÃO

Uma vez que suas preocupações podem ser de naturezas diversas (BORKO, 1968, p. 3; LE COADIC, 2004, p. 25), a Ciência da Informação busca aporte em campos com

¹⁵ Pesquisa desenvolvida com financiamento da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES).

especificidades que auxiliam em suas discussões. Assim, a Ciência da Informação mostra-se como um campo interdisciplinar, relacionando-se com campos como a Biblioteconomia e a Ciência da Computação (SARACEVIC, 1996; LE COADIC, 2004). Como apontado por Alves (2010, p. 122), em razão de seu caráter interdisciplinar, a Ciência da Informação

utiliza-se, em parte, dos processos de tratamento descritivo da informação (TDI) desenvolvidos na disciplina de Catalogação na área de Biblioteconomia para solucionar os problemas inerentes a essa ciência: tornar acessível e disponível a variedade crescente de recursos informacionais, bem como promover uma melhora na recuperação desses recursos.

A catalogação, entendida como “um processo de representação documentária que desde a antiguidade atua como instrumento de acesso a informação e ao documento” (PEREIRA; SANTOS, 1998, p. 123), busca “individualizar um item documentário e ao mesmo tempo multidimensionar suas formas de acesso por meio de recursos tecnológicos” (PEREIRA; SANTOS, 1998, p. 122).

No domínio bibliográfico, a partir da década de 1960 surgiram padrões de metadados que, com as estruturas para a codificação, permitiram que os dados resultantes das atividades de catalogação fossem processados e intercambiados entre aplicações de informática. Atualmente, alguns dos principais padrões de metadados internacionalmente utilizados para o intercâmbio de registros no domínio bibliográfico são os Formatos MARC 21.

Os registros criados com esses padrões, de modo geral, ainda são codificados com a norma ISO 2709 (INTERNATIONAL ORGANIZATION FOR STANDARDIZATION, 2008), que define uma estrutura para a codificação semelhante àquela utilizada na década de 1960. Além da ISO 2709, os Formatos MARC 21 dispõem hoje de estruturas para codificação tecnologicamente mais vantajosas e mais condizentes com a atualidade, como é caso da *Extensible Markup Language* (XML).

No Brasil, o suporte a importação e a exportação de registros nos Formatos MARC 21 tem se tornado uma característica buscada nos sistemas de gerenciamento de bibliotecas (CAFÉ; SANTOS; MACEDO, 2001; CÔRTE et al., 1999; PARANHOS, 2004). No entanto, em uma análise exploratória, observa-se a existência de sistemas que não satisfazem esse requisito. Alguns desses sistemas, entre eles o *Personal Home Library* (PHL), utilizam padrões de metadados próprios.

A não utilização de padrões de metadados internacionalmente aceitos, como os Formatos MARC 21, reduz as possibilidades de intercâmbio de registros e pode ocasionar conflitos durante a migração entre sistemas de gerenciamento de bibliotecas. Diante dos

sistemas de gerenciamento de bibliotecas que não utilizam qualquer padrão de metadados internacionalmente aceito, tem-se como problema a questão: como converter os registros exportados por esses sistemas em registros nos Formatos MARC 21?

Observa-se que os sistemas de gerenciamento de bibliotecas que não utilizam padrões de metadados internacionalmente aceitos têm, às vezes, seus padrões de metadados próprios baseados nas tabelas de seus bancos de dados. Nesses casos, as diferenças entre os bancos de dados dos sistemas inviabilizam a criação de uma aplicação de informática que possa ser utilizada na conversão dos registros exportados a partir de diferentes bancos de dados. Assim, para cada banco de dados diferente é necessário criar uma nova aplicação ou modificar uma já existente, o que se mostra dispendioso (ZAFALON, 2012, p. 24, 27 e 28).

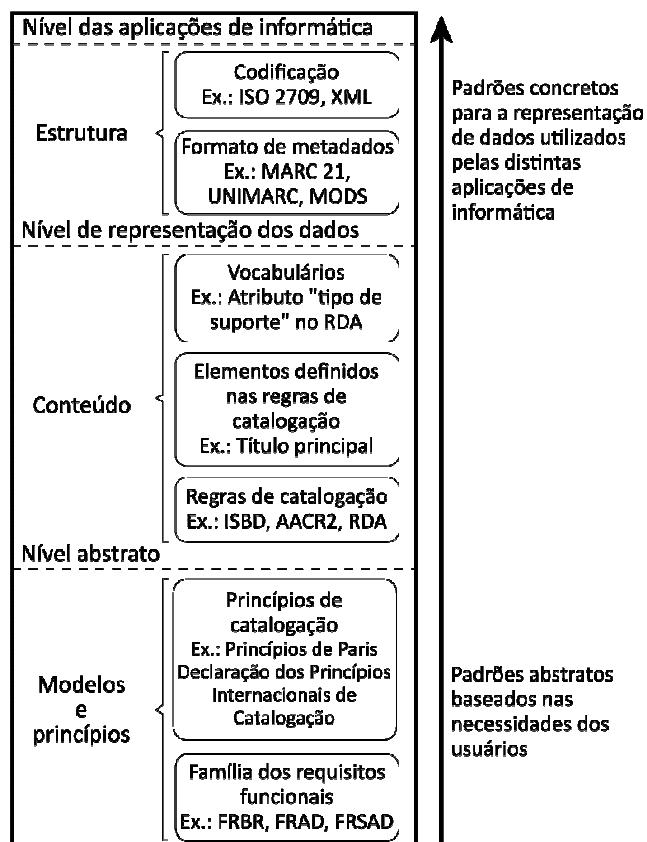
Por outro lado, as tecnologias atualmente disponíveis para a codificação trouxeram novas possibilidades para a conversão de registros entre padrões de metadados. Uma dessas tecnologias é a XML, que trouxe a possibilidade do uso de folhas de estilo elaboradas com a linguagem *Extensible Stylesheet Language for Transformation* (XSLT) na transformação de documentos XML de modo a atender distintos propósitos.

Considerando a provável efemeridade das aplicações de informática específicas para a conversão dos registros de cada sistema e a possibilidade do uso de folhas de estilo XSLT como instrumentos para a conversão, este trabalho tem por objetivo apresentar um modelo para a conversão de registros de distintos padrões de metadados codificados com a XML em registros nos Formatos MARC 21, tendo como um de seus pontos centrais a utilização de folhas de estilo de transformação elaboradas com a linguagem XSLT.

2 PADRÕES DE METADADOS E CODIFICAÇÕES

Os instrumentos de descrição utilizados no domínio bibliográfico podem ser categorizados de diversas formas. Entre as categorizações existentes, é utilizada aqui a categorização desenvolvida por Picco e Ortiz Repiso (2012, p. 149) apresentada na FIGURA 1.

FIGURA 1 – Instrumentos do domínio bibliográfico



Fonte: Picco e Ortiz Repiso (2012, p. 149, tradução nossa).

As autoras consideram três principais níveis: um nível abstrato abrangendo os modelos e os princípios do domínio bibliográfico; um nível de representação dos dados, com os instrumentos destinados ao conteúdo das representações; e um nível relacionado às aplicações de informática, compreendendo os formatos ou padrões de metadados e as codificações.

Uns dos principais padrões de metadados utilizados no domínio bibliográfico são os Formatos MARC 21. Um registro MARC 21, em qualquer um de seus formatos (para dados bibliográficos, de itens, de autoridade, de classificação ou para informação comunitária), envolve três componentes: a estrutura do registro, a designação do conteúdo e o conteúdo (LIBRARY OF CONGRESS, 1996). Seguindo a categorização de Picco e Ortiz Repiso (2012), a **estrutura do registro** é dada pela codificação, a **designação do conteúdo** é provida pelo padrão de metadados e o conteúdo está de acordo com regras de catalogação e vocabulários.

Enquanto padrão de metadados, cada Formato MARC 21 apresenta “um conjunto predeterminado de metadados (atributos codificados ou identificadores de uma entidade) metodologicamente construídos e padronizados” (ALVES, 2010, p. 47-48). Em cada

Formato MARC 21, os metadados estão materializados nas designações de conteúdo, que podem ser de três tipos: etiqueta de campo (composta por três caracteres numéricos), indicador (composto por um caractere) ou código de subcampo (composto por um caractere alfanumérico).

A codificação, que permite que os metadados e os valores a eles atribuídos sejam processados por aplicações de informática, é entendida por Balby (1995, p. 31) como as regras para a estruturação dos dados em meio legível por computador. Para a autora, as codificações são definidas por normas, “conjuntos de regras, externos ao formato de intercâmbio, que orientam a distribuição dos dados no meio físico de armazenamento (fita, disquete etc.)”.

Os Formatos MARC 21 são derivados no Formato MARC, desenvolvido pela *Library of Congress* (LC) na década de 1960. Uma vez que, em sua origem, o MARC tinha entre seus objetivos possibilitar o intercâmbio de registros via fitas magnéticas, foi utilizada com esse padrão uma codificação condizente com o tipo de suporte em questão, a fita magnética.

A codificação de caráter sequencial utilizada com o MARC foi reconhecida primeiramente nos Estados Unidos como um padrão nacional, posteriormente, em 1981, foi reconhecida pela *International Organization for Standardization* (ISO) como a norma ISO 2709: *Documentation – Format for bibliographic information interchange on magnetic tape* (INTERNATIONAL ORGANIZATION FOR STANDARDIZATION, 2008).

Apesar da norma ISO 2709 ter sido atualizada no decorrer dos anos, nota-se que não houve mudanças substanciais na codificação, de modo que os registros nos atuais Formatos MARC 21 são, em sua maior parte, codificados quase que da mesma forma com que eram codificados os registros na década de 1960, seja para propósitos de recuperação, por exemplo, via protocolo Z39.50, de importação e de exportação entre sistemas de gerenciamento de bibliotecas ou de armazenamento em bancos de dados.

O desenvolvimento das tecnologias de informática trouxe, além de outros suportes para o armazenamento e posterior intercâmbio de registros, outros ambientes digitais, como a Web. Com vistas ao uso das tecnologias de informática nos processos de catalogação, buscaram-se alternativas à codificação tradicionalmente utilizada com os registros nos Formatos MARC 21.

Algumas alternativas foram buscadas pela LC, que, na década de 1990, iniciou estudos sobre a utilização da *Standard Generalized Markup Language* (SGML) na codificação de registros. Com o surgimento da XML, a LC voltou-se para essa tecnologia e,

em 2002, lançou o *MARC 21 XML Schema*, mais conhecido como MARCXML: esquema XML que apresentava uma forma de codificação de registros MARC 21 utilizando a XML (LIBRARY OF CONGRESS, 2004, 2014).

A linguagem de marcação MARCXML foi pautada em uma série de considerações: ser um esquema simples e flexível para a codificação de registros MARC com a XML; não causar perdas no conteúdo quando os registros são convertidos a partir de registros ISO 2709; permitir a recriação de registros ISO 2709 a partir de registros MARCXML; facilitar a apresentação, a edição e a conversão dos dados por meio de folhas de estilo; permitir a validação dos registros; e facilitar a criação de ferramentas para a utilização, manipulação e conversão de registros MARC (LIBRARY OF CONGRESS, 2004).

3 CONVERSÃO RETROSPECTIVA E REUTILIZAÇÃO DE METADADOS

Em relação à conversão de registros no domínio bibliográfico, a literatura apresenta estudos e relatos que podem ser classificados em dois principais tipos: a **conversão retrospectiva** e a **reutilização de metadados** (*repurposing metadata*).

A conversão retrospectiva envolve a inserção dos dados de registros analógicos, de modo geral presentes em fichas catalográficas, em registros digitais (ASENSI ARTIGA; RODRÍGUEZ MUÑOZ, 2001; CASTRO, 2003; OLIVEIRA et al., 1998). Segundo Bowman (2007, p. 331), a conversão retrospectiva tem acompanhado a catalogação desde o início da automação de bibliotecas, o que ocorreu a partir da década de 1960 com a criação do Formato MARC e dos padrões dele derivados.

Devido à ampla utilização de sistemas digitais de gerenciamento de bibliotecas e de outras aplicações de informática destinadas ao gerenciamento, ao armazenamento e/ou à disseminação de recursos informacionais, por exemplo, os repositórios e as bibliotecas digitais, a literatura sobre conversão passou a preocupar-se não somente com a conversão retrospectiva (do analógico para o digital), mas também com a conversão dos registros já presentes no ambiente digital. Dada à existência de diversos padrões de metadados, tornou-se necessário converter os registros criados com um padrão em registros de acordo com outros padrões, por exemplo, para possibilitar o intercâmbio ou a migração desses registros entre sistemas que não adotam o mesmo padrão (ASENSI ARTIGA; RODRÍGUEZ MUÑOZ, 2001; RUDIĆ; SURLA, 2009; SCHMIDT; PATEL, 1999). Esse tipo de conversão tem sido chamado de reutilização de metadados e sua necessidade estende-se ainda mais devido a diversidade de padrões internacionais, nacionais e locais (WOODLEY, 2008).

Para Woodley (2008, p. 6), o processo de reutilização de metadados compreende um amplo conjunto de atividades: converter ou transformar registros de um padrão de metadados para outro; migrar de um padrão legado para outro; integrar registros criados de acordo com diferentes padrões; e coletar ou agregar registros criados utilizando um padrão compartilhado pela comunidade ou diversos padrões.

Na literatura são encontrados relatos e estudos sobre a reutilização de metadados tanto no contexto dos catálogos dos sistemas de gerenciamento de bibliotecas, quanto em outros ambientes informacionais, tais como os repositórios institucionais.

Averkamp e Lee (2009) apresentam um *workflow* para a reutilização dos metadados de teses e de dissertações, oriundos da base de dados *ProQuest UMI Dissertation Publishing*, na criação de registros para inserção em um repositório institucional e em um catálogo de biblioteca. Keenan (2010) relata a reutilização de registros no padrão Dublin Core da base de dados *U.S. Congressional Serial Set, 1817-1994* para a criação de registros bibliográficos no Formato MARC 21, visando à inserção desses em um catálogo de biblioteca. Essa base de dados oferece aos seus assinantes os registros em Dublin Core sem qualquer custo adicional, ao passo que vende seus registros no Formato MARC 21. A reutilização dos metadados nesse caso resultou em uma significativa redução dos custos, se comparada à compra dos registros já no Formato MARC 21.¹⁶

Na literatura brasileira são encontrados estudos e relatos sobre a conversão retrospectiva (CASTRO, 2003; DIAS, 1999; PEREZ; LIMA, 2002) e sobre a reutilização de metadados (BOICA; OLIVEIRA, 2008; MURAKAMI, 2012; RAPOSO; OLIVEIRA; SHINOTSUKA, 1985).¹⁷ Este estudo, partindo da necessidade de converter registros presentes em ambientes digitais, é considerado do tipo reutilização de metadados e seu principal resultado, um modelo para a conversão de registros, é apresentado na seção seguinte.

4 MODELO PARA A CONVERSÃO DE REGISTROS

Como abordado, a necessidade de converter registros digitais no domínio bibliográfico relaciona-se à utilização dos dados em uma aplicação de informática diferente daquela da qual os dados proveem, sendo necessário, portanto, considerar na conversão os

¹⁶ Segundo Keenan (2010), o custo estimado para a compra e a inserção dos registros seria de US\$ 25.669,71, enquanto que o custo da reutilização foi de US\$ 1.129,05, incluindo o tempo gasto pela equipe da biblioteca no planejamento, na pesquisa e no desenvolvimento de scripts.

¹⁷ Os estudos sobre a reutilização de metadados, principalmente os nacionais, não empregam essa denominação. Em vez disso, utilizam apenas os termos conversão ou migração.

padrões de metadados e as codificações aceitas pela aplicação de informática que receberá os registros.

Em razão da ampla utilização da ISO 2709 para o intercâmbio de dados no domínio bibliográfico, algumas aplicações de informática, principalmente do tipo sistema de gerenciamento de bibliotecas, possibilitam a importação de registros MARC 21 apenas se codificados de acordo com tal norma. Outras aplicações, no entanto, permitem a importação de registros MARC 21 apenas se codificados com a XML e seguindo o MARCXML.

Considerando essas duas situações, o modelo para a conversão de registros foi elaborado diante da seguinte necessidade: registros de acordo com um padrão de metadados qualquer codificados com a XML precisam ser convertidos em registros em um dos Formatos MARC 21, codificados com a XML (MARCXML) ou com a ISO 2709.

Uma vez que os registros de origem estarão codificados com a XML e os registros de destino poderão também estar codificados com essa tecnologia, foi incluída no modelo uma folha de estilo criada com a *Extensible Stylesheet Language for Transformation* (XSLT).

A linguagem XSLT pode ser utilizada na criação de folhas de estilo (documentos com regras que deverão ser executadas por uma aplicação de informática específica) para a transformação de um documento XML que está de acordo com uma linguagem de marcação em um documento XML de acordo com outra linguagem de marcação ou mesmo em um documento que não seja XML (RAY, 2001; TIDWELL, 2008; W3C, 2007).

Eito Brun (2008, p. 95) destaca a importância da XSLT para as situações em que a XML é utilizada para transferir e intercambiar dados estruturados com diferentes esquemas, ou seja, com diferentes linguagens de marcação. Tais situações requerem uma tecnologia que permita transformar os documentos baseados em um esquema em outro diferente. Para essa transformação uma “folha de estilo XSLT estabelecerá as equivalências entre os elementos e os atributos do esquema em que estão os documentos XML e os elementos e os atributos do esquema ao qual se deseja converter” (EITO BRUN, 2008, p. 95, tradução nossa).

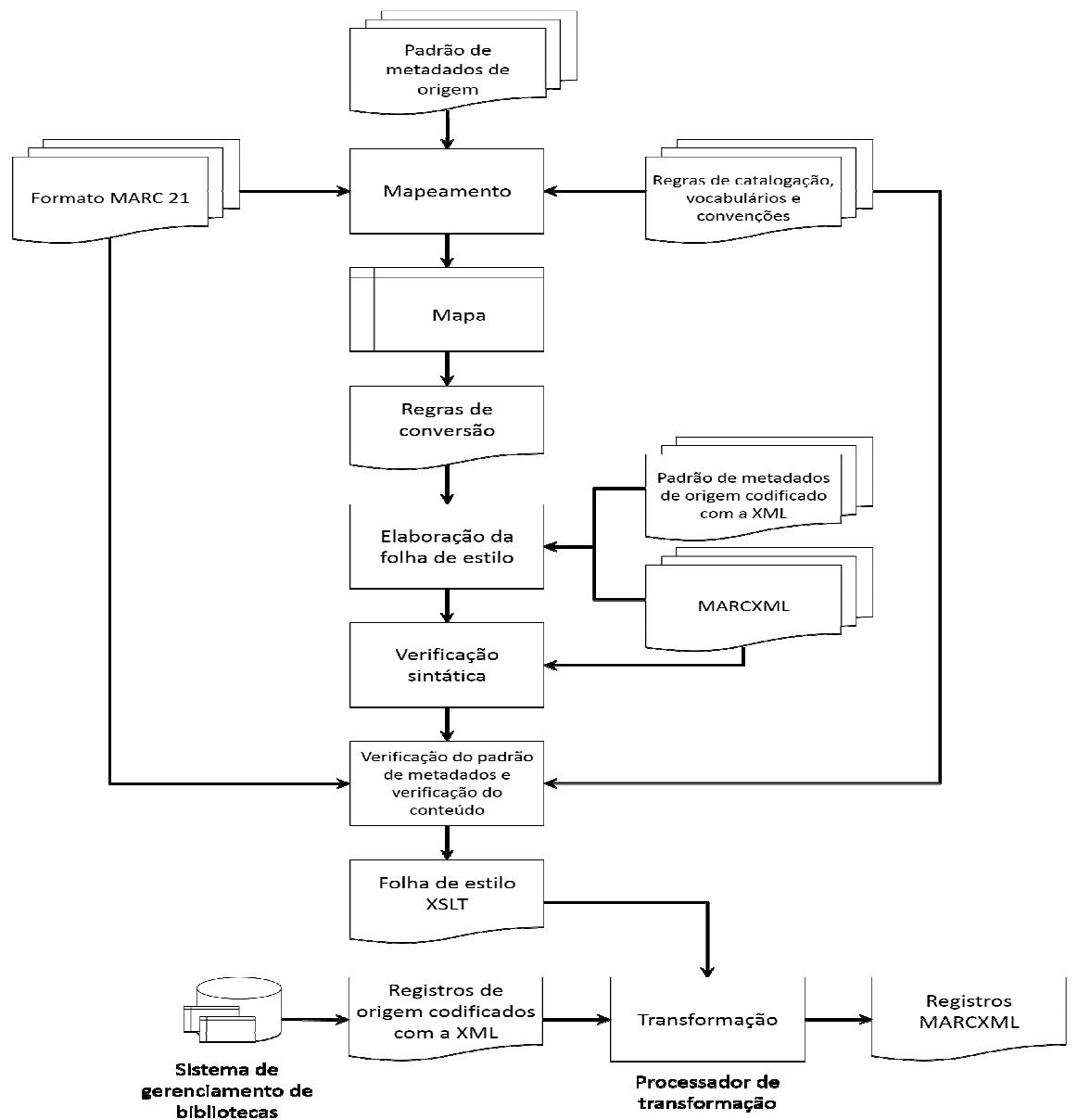
A utilização de folhas de estilo XSLT na conversão de registros é relatada em alguns estudos internacionais que consistem, principalmente, em relatos demonstrando os procedimentos e os instrumentos utilizados (KEITH, 2004; KEENAN, 2010; KURTH; RUDDY; RUPP, 2004; RUDIĆ; SURLA, 2009). Sobre a literatura nacional, concorda-se com os apontamentos realizados por Zafalon (2012, p. 22) de que,

Na literatura nacional, diagnósticos e situações de conversão de bases de dados são relatados no sentido de apresentar as experiências e os

procedimentos metodológicos utilizados em diversas instituições. Estas publicações, de certa forma, tendem a deixar de explicitar as concepções de ordem teórico-metodológica envolvidas nesse processo.

Apesar de, de modo geral, apresentarem sequências de procedimentos e conjuntos de instrumentos mais ou menos semelhantes, nota-se nos estudos sobre a conversão de registros pouca ou nenhuma sistematização de um modelo ou de uma estrutura teórico-metodológica que compreenda os procedimentos e os instrumentos que comumente relacionam-se à conversão. Assim, para suprir essa lacuna e auxiliar na conversão de registros para os Formatos MARC 21, é apresentado na FIGURA 2 o modelo para a conversão de registros resultado deste estudo.

FIGURA 2 – Modelo para a conversão de registros



Fonte: Elaborada pelo autor

O modelo apresentado na FIGURA 2 é sintetizado da seguinte forma:

- a documentação do padrão de metadados de origem, a documentação do Formato MARC 21 de destino, as regras de catalogação, os vocabulários e as convenções são utilizados no mapeamento, que tem como resultado um mapa indicando as correspondências existentes entre os metadados dos padrões de origem e de destino;
- as correspondências indicadas no mapa são redigidas como regras de conversão;
- as regras de conversão, a documentação da codificação do padrão de metadados de origem com a XML e a documentação do MARCXML são utilizadas na elaboração da folha de estilo XSLT;
- durante a elaboração da folha de estilo XSLT são realizadas a verificação sintática, a verificação do padrão de metadados e a verificação do conteúdo;
- um editor XML realiza a verificação sintática com base no esquema do MARCXML, assegurando que os registros obtidos a partir da conversão utilizando a folha de estilo estão em acordo com o MARCXML;
- um agente humano realiza a verificação do padrão de metadados, com base na documentação do Formato MARC 21, e a verificação do conteúdo, com base nas regras de catalogação, nos vocabulários e nas convenções, assegurando, assim, que os registros obtidos a partir da conversão estão de acordo com o padrão de metadados de destino e com os instrumentos de descrição;
- a folha de estilo XSLT já finalizada é inserida no processador de transformação;
- os registros de origem são exportados a partir do sistema de gerenciamento de bibliotecas e inseridos no processador de transformação;
- o processador transforma os registros de origem em registros no Formato MARC 21 codificados de acordo com o MARCXML.

Para uma maior elucidação do modelo apresentado na Figura 2 seus processos e seus componentes são descritos nas seções seguintes.

4.1 Mapeamento e elaboração das regras de conversão

O processo de estabelecer os relacionamentos entre metadados semanticamente equivalentes em diferentes padrões é denominado na literatura como mapeamento (*mapping* ou *crosswalking*). O resultado do mapeamento é chamado de mapa (*map* ou *crosswalk*) e consiste em uma representação visual dos relacionamentos, das equivalências e das lacunas entre os padrões de metadados mapeados (KURTH; RUDDY; RUPP, 2004, p. 154; ST. PIERRE; LAPLANT, 1998; WOODLEY, 2008, p. 3).

St. Pierre e LaPlant (1998) destacam que o mapeamento inclui um mapa semântico e especificações para a conversão que indicam as transformações requeridas para converter o conteúdo do registro do padrão de origem em um registro de acordo com o padrão de destino.

Rudić e Surla (2009, p. 952) apontam que uma descrição de conversão pode conter uma tabela e regras de conversão. Essa tabela de conversão é o mapa, e as regras de conversão são as “especificações para a conversão” (ST. PIERRE; LAPLANT, 1998).

Para Machovec (2002, p. 2), o mapeamento entre quaisquer padrões não será perfeito e será mais ou menos bem sucedido dependendo dos padrões mapeados. Um dos fatores que influenciam o mapeamento e o sucesso da conversão é o nível de granularidade dos padrões de metadados mapeados (MACHOVEC, 2002, p. 1), portanto, esse é um importante aspecto a ser levado em conta.

Segundo Alves, Simionato e Santos (2012, p. 3), o termo granularidade tem origem na Ciência da Computação, porém passa ser utilizado também no universo bibliográfico para referir-se aos níveis de detalhe em que um recurso informacional pode ser descrito. Para as autoras, a granularidade é dividida em dois níveis: a granularidade fina (*fine-granularity*), quando a descrição apresenta um alto nível de detalhamento, e a granularidade grossa (*coarse-granularity*), quando a descrição possui um baixo nível de detalhamento.

Woodley (2008, p. 7) destaca que os metadados no padrão de origem podem não estar bem definidos ou podem conter uma mistura de dados que, no padrão de destino, podem estar em diferentes campos. Nesses casos, a identificação de um dado dentro do conteúdo de um metadado pode não ser possível, pode requerer a manipulação dos dados diversas vezes, ser difícil, demorado e repleto de erros (MACHOVEC, 2002, p. 2; WOODLEY, 2008, p. 7).

Considerando os diferentes níveis de granularidade entre os padrões, as correspondências do tipo *um-para-muitos* e *muitos-para-um* são frequentes. Woodley (2008, p. 7) destaca a falta de correspondências perfeitas entre os metadados, o que requer o estabelecimento de correspondências entre metadados com significados próximos.

Algum metadado do padrão de origem pode não possuir correspondente no padrão de destino, ocasionando no descarte de seu conteúdo durante a conversão. Situações inversas também são possíveis: algum metadado do padrão de destino pode não ter um correspondente no padrão de origem. Nesses casos, se o metadado não é obrigatório, pode ser deixado em branco, caso seja obrigatório, é necessário prover o conteúdo do mesmo, por

exemplo, por meio de um valor padrão (*default*) (MACHOVEC, 2002; ST. PIERRE; LAPLANT, 1998).

Outro ponto a ser observado, segundo St. Pierre e LaPlant (1998), são as propriedades dos metadados do padrão de origem e do padrão de destino, por exemplo, obrigatoriedade, repetitividade e tipo de dado.

Apesar dos padrões de metadados existirem de forma independente dos instrumentos de descrição (regras de catalogação, vocabulários e convenções) e vice-versa, observa-se que os padrões de metadados, em razão do conjunto de metadados de que dispõem e do modo com que tais metadados estão arranjados, condicionam a utilização dos instrumentos de descrição, e, em decorrência da oferta de mais ou de menos possibilidades que aquelas previstas pelos instrumentos de descrição, influenciam o modo com que os dados são registrados. Assim, entendendo que o modo com que um dado é registrado em um padrão pode diferir do modo como o mesmo dado é registrado em outro, é necessário considerar no mapeamento os instrumentos de descrição utilizados na criação dos registros a serem convertidos.

Durante a elaboração do mapa e após a sua finalização, devem ser redigidas regras de conversão de modo a facilitar a criação da folha de estilo XSLT pelo profissional da Ciência da Computação. Essas regras são redigidas em linguagem natural indicando os relacionamentos fornecidos no mapa e, opcionalmente, provendo orientações sobre os instrumentos de descrição utilizados na criação dos registros de origem.

A importância dessas regras está no fato de que o profissional da Ciência da Computação pode não possuir o conhecimento suficiente dos padrões de metadados e dos instrumentos de descrição para interpretar o mapa.

4.2 Elaboração da folha de estilo XSLT

Além do mapa e das regras de conversão, para a elaboração da folha de estilo é necessário (1) escolher um editor XML, um processador de transformação e a versão da XSLT, e (2) conhecer as especificações das linguagens de marcação dos padrões de metadados de origem e de destino, além do conhecimento de XSLT. Essas especificações definem o modo com que os metadados compreendidos pelos padrões e os valores a eles atribuídos são expressos na XML.

A especificação da linguagem de marcação utilizada com o padrão de metadados de origem nem sempre está disponível em uma *Document Type Definition* (DTD) ou em esquema XML, ou mesmo acessível ao profissional usuário do sistema de gerenciamento de bibliotecas. Nesses casos, uma possibilidade é apreender a estrutura definida pela

linguagem de marcação a partir dos documentos XML que estão de acordo com ela, ou seja, a partir da análise dos registros exportados em XML.

A elaboração da folha de estilo compreende principalmente dois passos:

- a inclusão da declaração XML, a inclusão do elemento do documento dentro do qual será inserido todo o conteúdo da folha de estilo, e a inclusão das declarações dos *namespaces*;
- a elaboração das regras de transformação com base no mapa e nas regras de conversão.

As regras de transformação selecionam e manipulam os elementos e os conteúdos do documento de entrada fazendo com que sejam organizados e/ou transformados para se adequarem aos elementos do documento de saída; algumas das ações realizadas pelas regras de transformação são:

- incluir o conteúdo de um elemento em um elemento de saída;
- reunir o conteúdo de dois ou mais elementos para formar o conteúdo de um elemento de saída;
- separar o conteúdo de um elemento e distribuí-lo entre dois ou mais elementos de saída;
- editar o conteúdo de um elemento (substituir, acrescentar ou remover caracteres) para formar o conteúdo de um elemento de saída;
- criar um elemento de saída com um conteúdo que não pôde ser obtido a partir do documento de entrada;
- incluir, reunir, separar, editar e criar elementos e conteúdos com base em condições.

4.3 Verificação da folha de estilo

Durante a elaboração da folha de estilo, é necessário verificar os registros resultantes da conversão para checar se estão de acordo com o resultado esperado. A verificação busca assegurar que os registros convertidos com a folha de estilo estão de acordo com o MARCXML, com o padrão de metadados (um dos Formatos MARC 21) e com os instrumentos de descrição. Assim, a verificação inclui três etapas que proveem um *feedback* indicando se há ou não a necessidade de modificações no mapa, nas regras de conversão e/ou na folha de estilo:

- **verificação sintática:** realizada por editores XML, é a análise do registro MARCXML para verificar se o mesmo está de acordo com a linguagem de marcação MARCXML;

- **verificação do padrão de metadados:** é a análise das designações do conteúdo (etiquetas, indicadores e códigos de subcampo) presentes nos registros convertidos para verificar se estão de acordo com o padrão de metadados;
- **verificação do conteúdo:** é a confrontação dos dados presentes nos registros convertidos com as regras de catalogação, os vocabulários e as convenções utilizadas pela instituição; tem por objetivo checar a adequação dos registros aos instrumentos de descrição, mostrando se o resultado da conversão atende ou não aos requisitos de qualidade definidos nesses instrumentos.

4.4 Exportação e conversão dos registros

Entre os processos do modelo, a exportação dos registros é o mais dependente do sistema de gerenciamento de bibliotecas. Para a utilização do modelo proposto é necessário que os registros exportados estejam codificados com a XML, não importando qual seja a linguagem de marcação ou o padrão de metadados utilizado com essa codificação. Assim, se os registros não estão codificados com a XML e não há qualquer possibilidade de assim codificá-los, o modelo para a conversão de registros não poderá ser aplicado tal como foi elaborado.

Entende-se que a exportação pode ocorrer de dois modos. O primeiro é por meio de uma interface, seja ela gráfica ou não, provida pelo sistema de gerenciamento de bibliotecas. Um segundo modo ocorre quando o sistema não oferece uma interface de exportação, mas é possível acessar sua base de dados por meio do sistema de gerenciamento de banco de dados e, então, exportar os registros codificados com a XML.

Com a folha de estilo XSLT elaborada e os registros exportados, o próximo processo é converter os registros do padrão de metadados de origem codificados com a XML em registros no Formato MARC 21 de acordo com o MARCXML. Os componentes necessários a essa conversão são: a folha de estilo XSLT, um processador de transformação compatível com a versão da XSLT utilizada na folha de estilo e o documento XML contendo os registros exportados.

De modo geral, a folha de estilo XSLT e o documento XML contendo os registros exportados são inseridos no processador que, a partir de algum comando, inicia a transformação. Ao final da transformação, o processador gera um documento XML contendo os registros convertidos de acordo com o MARCXML.

A conversão dos registros MARCXML para ISO 2709 é um processo opcional, realizado somente quando é necessário obter registros nos Formatos MARC 21 codificados

com a ISO 2709. Para essa segunda conversão pode ser utilizada a aplicação disponibilizada pela LC¹⁸ ou a que acompanha a suíte de aplicativos gratuita MarcEdit¹⁹, entre outras.

Para a validação do modelo elaborado, realizou-se sua aplicação na conversão de registros bibliográficos exportados pelo sistema de gerenciamento de bibliotecas *Personal Home Library* (PHL). Os procedimentos utilizados na aplicação do modelo e os resultados obtidos são detalhados em Assumpção (2013, p. 81-101).

5 CONSIDERAÇÕES FINAIS

A existência de sistemas de gerenciamento de bibliotecas que não possibilitam a exportação dos dados nos Formatos MARC 21 e nem em qualquer padrão de metadados internacionalmente utilizado, fomenta a necessidade de estudos sobre a reutilização de metadados. Essa necessidade está pautada na consideração de que as instituições usuárias desses sistemas, por não disporem de seus registros de acordo com padrões de metadados internacionalmente aceitos, poderão estar impossibilitadas de participarem de programas de catalogação cooperativa e estarão sujeitas a problemas e a retrabalhos durante a migração entre sistemas de gerenciamento de bibliotecas.

Diante dessa consideração, o objetivo deste estudo foi elaborar um modelo para a conversão de registros de distintos padrões de metadados codificados com a XML em registros nos Formatos MARC 21, tendo como um dos pontos centrais do modelo a utilização de folhas de estilo de transformação elaboradas com a linguagem XSLT.

Sobre o modelo elaborado e sua aplicação na conversão de registros bibliográficos do sistema PHL, são traçadas as seguintes considerações:

- o modelo busca apresentar uma generalização e uma formalização dos processos e dos componentes necessários à conversão de registros em XML utilizando folhas de estilo XSLT, processos e componentes estes até então dispersos na literatura e abordados no contexto de realidades individuais;
- nos ambientes digitais, a conversão de registros não deve ser uma tarefa realizada apenas por profissionais da Ciência da Informação ou apenas da Ciência da Computação. A Ciência da Informação possui o conhecimento dos padrões de metadados e dos instrumentos de descrição, portanto, tem a competência necessária para traçar mapas estabelecendo correspondências entre os metadados de diferentes

¹⁸ Disponível em: <<http://www.loc.gov/standards/marcxml/>>. Acesso em: 6 jul. 2014.

¹⁹ Disponível em: <<http://marcedit.reeset.net>>. Acesso em: 6 jul. 2014.

padrões, ao passo que à Ciência da Computação cabe a transposição dos mapas para aplicações de informática que, manipulando os registros do padrão de metadados de origem, possam convertê-los para o padrão de metadados de destino. Nesse sentido, o modelo proposto é uma tentativa de delineamento dos papéis exercidos pelos profissionais dessas áreas durante as atividades de conversão de registros;

- entre os fatores que interferem no tempo demandado na elaboração de uma folha de estilo XSLT estão a complexidade dos padrões de origem e de destino, incluindo o número de campos/subcampos, a especificidade e o nível de granularidade dos padrões, e o quanto mínima é a perda de dados tolerada;
- o grau de complexidade das instruções da folha de estilo relaciona-se diretamente à perda de dados na conversão. Quanto maior a manipulação (junção, separação, substituição, etc.) dos dados do documento XML de origem, melhor esses dados poderão ser alocados no padrão de metadados de destino;
- em razão das diferenças entre os padrões de origem e de destino, após a conversão, pode ser necessário completar ou modificar os registros convertidos, o que deve ser levado em conta pelas instituições em seus planos para conversão; e
- ao considerar o tempo demandado na elaboração da folha de estilo, deve ser levada em conta a quantidade de registros que serão convertidos, de modo a melhor ponderar sobre a relação custo-benefício da conversão utilizando o modelo.

A proposição inicial desta pesquisa era a elaboração de um modelo para a conversão que favorecesse o intercâmbio de registros e minimizasse os conflitos durante a migração de registros de sistemas que não utilizam padrões de metadados internacionalmente aceitos.

Os resultados obtidos com a utilização do modelo são registros em um dos Formatos MARC 21 e codificados com a XML seguindo o MARCXML. Tais registros podem ser codificados com a ISO 2709 de modo a atender a uma necessidade específica, por exemplo, a importação em um sistema de gerenciamento de bibliotecas. No entanto, uma vez que os registros convertidos estão de acordo com o MARCXML, uma ampla gama de possibilidades passa a existir, principalmente devido às possibilidades trazidas pela XML (a utilização de folhas de estilo XSLT na transformação de documentos é uma dessas possibilidades) e ao papel que os Formatos MARC 21 podem desempenhar como padrões capazes de intermediar a troca de registros entre diferentes padrões de metadados no domínio bibliográfico.

Ou seja, além de prover uma solução para o problema de como converter os registros exportados em registros nos Formatos MARC 21, esta pesquisa apresentou novas

possibilidades para o intercâmbio de dados que favorecem a interoperabilidade entre diferentes aplicações de informática que comunicam dados sobre recursos informacionais. Nesse sentido, observa-se que, apesar de estar voltado ao contexto dos Formatos MARC 21, o modelo elaborado pode ser aplicado na conversão de registros para outros padrões de metadados, sendo essa possibilidade uma consideração a respeito da aplicação do modelo em outras comunidades além da comunidade de bibliotecas.

Por fim, cabe ressaltar que este estudo não defende que o profissional da Ciência da Informação seja um especialista na construção de folhas de estilo, mas sim que tal profissional tenha conhecimento das tecnologias desenvolvidas na Ciência da Computação que possam ser úteis às suas atividades, favorecendo, assim, o uso estratégico dessas tecnologias em prol do desenvolvimento e do gerenciamento de ambientes informacionais digitais.

REFERÊNCIAS

- ALVES, R. C. V. **Metadados como elementos do processo de catalogação**. 2010. 132f. Tese (Doutorado em Ciência da Informação) – Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília, 2010. Disponível em: <<http://base.repositorio.unesp.br/handle/unesp/103361>>. Acesso em: 7 jul. 2014.
- ALVES, R. C. V.; SIMIONATO, A. C.; SANTOS, P. L. V. A. da C. Aspectos de granularidade na representação da informação no universo bibliográfico. In: ENCONTRO NACIONAL DE CATALOGADORES, 1., 2012, Rio de Janeiro. **Trabalhos...** Rio de Janeiro: Fundação Biblioteca Nacional, 2012. Disponível em: <<http://pt.scribd.com/doc/109274547/Aspectos-de-granularidade-na-representacao-da-informacao-no-universo-bibliografico>>. Acesso em: 6 jul. 2014.
- ASENSI ARTIGA, V.; RODRÍGUEZ MUÑOZ, J. V. El proceso de catalogación automatizada. In: PINTO MOLINA, M. (Org.). **Catalogación de documentos: teoría y práctica**. 2. ed. rev. e atual. Madrid: Síntesis, 2001. p. 105-143.
- ASSUMPÇÃO, F. S. **Conversão de registros em XML para MARC 21**: um modelo baseado em folhas de estilo XSLT. 135 f. Dissertação (Mestrado em Ciência da Informação) – Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília, 2013. Disponível em: <<http://base.repositorio.unesp.br/handle/unesp/93658>>. Acesso em: 14 maio 2014.
- AVERKAMP, S.; LEE, J. Repurposing ProQuest Metadata for Batch Ingesting ETDs into an Institutional Repository. **Code4Lib Journal**, n. 7, 2009. Disponível em: <<http://journal.code4lib.org/articles/1647>>. Acesso em: 6 jul. 2014.
- BALBY, C. N. Formatos de intercâmbio de registros bibliográficos: conceitos básicos. **Cadernos da F.F.C.**, v. 4, n. 1, p. 29-35, 1995.
- BOICA, A. L.; OLIVEIRA, L. H. M. de. Conversão de metadados do padrão Dublin Core para o RDF. **Global Science and Technology**, v. 1, n. 2, p. 8-13, dez./mar. 2008.

Disponível em: <<http://rioverde.ifgoiano.edu.br/periodicos/index.php/gst/article/view/13>>. Acesso em: 6 jul. 2014.

BORKO, H. Information Science: what is it? **American Documentation**, v. 19, n. 1, p. 3-5, 1968.

BOWMAN, J. H. Retrospective conversion: The early years. **Library History**, v. 23, p. 331-340, Dec. 2007. <<http://dx.doi.org/10.1179/174581607x254811>>. Acesso em: 06 jul. 2014.

CAFÉ, L.; SANTOS, C. dos; MACEDO, F. Proposta de um método para escolha de software de automação de bibliotecas. **Ciência da Informação**, v. 30, n. 2, p. 70-79, maio/ago. 2001. Disponível em: <<http://dx.doi.org/10.1590/S0100-19652001000200009>>. Acesso em: 6 jul. 2014.

CASTRO, F. F. de. Conversão retrospectiva de registros bibliográficos. In: CONGRESSO BRASILEIRO DE CIÊNCIAS DA COMUNICAÇÃO, 24., 2003, Belo Horizonte. **Anais...** São Paulo: Intercom, 2003. Disponível em: <<http://www.portcom.intercom.org.br/navegacaoDetalhe.php?option=trabalho&id=42503>>. Acesso em: 6 jul. 2014.

CÔRTE, A. R. et al. Automação de bibliotecas e centros de documentação: o processo de avaliação e seleção de softwares. **Ciência da Informação**, v. 28, n. 3, p. 241-256, set./dez. 1999. Disponível em: <<http://dx.doi.org/10.1590/S0100-19651999000300002>>. Acesso em: 06 jul. 2014.

DIAS, M. do R. I. **Catalogação e qualidade**: breve estudo. Marília: UNESP/CGB, 1999.

EITO BRUN, R. **Lenguajes de marcas para la gestión de recursos digitales**: aproximación técnica, especificaciones y referencia. Gijón: Trea, 2008.

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION. **ISO 2709:2008**: Abstract. Geneva, 2008. Disponível em: <http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=41319>. Acesso em: 6 jul. 2014.

KEENAN, T. M. Why Purchase When You Can Repurpose? Using Crosswalks to Enhance User Access. **Code4Lib Journal**, n. 11, 2010. Disponível em: <<http://journal.code4lib.org/articles/3604>>. Acesso em: 6 jul. 2014.

KEITH, C. Using XSLT to manipulate MARC metadata. **Library Hi Tech**, v. 22, n. 2, p. 122-130, 2004. Disponível em: <<http://dx.doi.org/10.1108/07378830410524549>>. Acesso em: 6 jul. 2014.

KURTH, M.; RUDDY, D.; RUPP, N. Repurposing MARC metadata: using digital project experience to develop a metadata management design. **Library Hi Tech**, v. 22, n. 2, p. 153-165, 2004. Disponível em: <<http://dx.doi.org/10.1108/07378830410524585>>. Acesso em: 6 jul. 2014.

LE COADIC, Y-F. **A ciência da informação**. 2. ed. Brasília: Briquet de Lemos, 2004.

LIBRARY OF CONGRESS. **MARC 21 XML Schema**: official web site. Washington, D.C., 2014. Disponível em: <<http://www.loc.gov/standards/marcxml/>>. Acesso em: 6 jul. 2014.

LIBRARY OF CONGRESS. **MARC XML Design Considerations**. Washington, D.C., 2004. Disponível em: <<http://www.loc.gov/standards/marcxml/marcxml-design.html>>. Acesso em: 6 jul. 2014.

LIBRARY OF CONGRESS. **The MARC 21 Formats**: Background and Principles. Revised November 1996. Washington, D.C.: Library of Congress, 1996. Disponível em: <<http://www.loc.gov/marc/96principl.html>>. Acesso em: 6 jul. 2014.

MACHOVEC, G. S. Challenges and issues with metadata crosswalks. **Information Intelligence Online Libraries and Microcomputers**, v. 20, n. 4, p. 1-3, Apr. 2002.

MURAKAMI, T. Migração de registros de uma Tabela para o MARC21. In: BIBLIOTECÁRIOS sem fronteiras. [S.l.: s.n.], 2012. Disponível em: <<http://bsf.org.br/2012/01/16/migracao-de-registros-de-uma-tabela-para-o-marc21>>. Acesso em: 6 jul. 2014.

OLIVEIRA, N. M. et al. Compact Disc Cataloging - CatCD: análise de um instrumento para conversão retrospectiva no Sistema de Bibliotecas da UNICAMP. **Perspectivas em Ciência da Informação**, v. 3, n. 1, p. 41-46, jan./jun. 1998. Disponível em: <<http://portaldeperiodicos.eci.ufmg.br/index.php/pci/article/view/608>>. Acesso em: 6 jul. 2014.

PARANHOS, W. M. M. da R. Fragmentos metodológicos para projetos e execução de gestão informatizada de coleções de documentos e serviços em bibliotecas. **Encontros BIBLI**, n. esp., p. 14-32, 2º sem. 2004. Disponível em: <<https://periodicos.ufsc.br/index.php/eb/article/view/1518-2924.2004v9nesp2p14>> http://www.loc.gov/marc/96principl.html. Acesso em: 6 jul. 2014.

PEREIRA, A. M.; SANTOS, P. L. V. A. da C. O uso estratégico das tecnologias em catalogação. **Cadernos da F.F.C.**, v. 7, n. 1/2, p. 121-131, 1998.

PEREZ, D. R.; LIMA, P. O projeto de conversão retrospectiva de registros bibliográficos: uma experiência do sistema de bibliotecas da Pontifícia Universidade Católica do Rio de Janeiro. In: SEMINÁRIO NACIONAL DE BIBLIOTECAS UNIVERSITÁRIAS, 12., 2002, Recife. **Anais...** Recife: UFPE, 2002. Disponível em: <<http://alfarrabiosroger.files.wordpress.com/2009/12/37-a.pdf>>. Acesso em: 6 jul. 2014.

PICCO, P.; ORTIZ REISO, V. RDA, el nuevo código de catalogación: cambios y desafíos para su aplicación. **Revista Española de Documentación Científica**, v. 35, n. 1, p. 145-173, enero-marzo 2012. Disponível em: <<http://dx.doi.org/10.3989/redc.2012.1.848>>. Acesso em: 6 jul. 2014.

RAPOSO, M. de F. P.; OLIVEIRA, V. L. S. de; SHINOTSUKA, F. H. Mudança para o formato CALCO: uma experiência. **Revista de Biblioteconomia de Brasília**, v. 13, n. 1, p. 21-26, jan./jun. 1985.

RAY, E. T. **Aprendendo XML**. Rio de Janeiro: Campus, 2001.

RUDIĆ, G.; SURLA, D. Conversion of bibliographic records to MARC 21 format. **The Electronic Library**, v. 27, n. 6, p. 950-967, 2009. Disponível em: <<http://dx.doi.org/10.1108/02640470911004057>>. Acesso em: 6 jul. 2014.

SARACEVIC, T. Ciência da informação: origem, evolução e relações. **Perspectivas em Ciência da Informação**, v. 1, n. 1, p. 41-62, jan./jun. 1996. Disponível em:

<<http://portaldeperiodicos.eci.ufmg.br/index.php/pci/article/view/235>>. Acesso em: 6 jul. 2014.

SCHMIDT, N.; PATEL, A. MARC record conversion: a generalized approach. **Computer Standards & Interfaces**, v. 21, p. 287-297, 1999. Disponível em: <[http://dx.doi.org/10.1016/S0920-5489\(99\)00007-0](http://dx.doi.org/10.1016/S0920-5489(99)00007-0)>. Acesso em: 6 jul. 2014.

ST. PIERRE, M.; LAPLANT, W. P. **Issues in Crosswalking Content Metadata Standards**. Baltimore: National Information Standards Organization, 1998. Disponível em: <http://www.niso.org/publications/white_papers/crosswalk>. Acesso em: 6 jul. 2014.

TIDWELL, D. **XSLT**. 2nd ed. Beijing: O'Reilly, 2008.

W3C. **XSL Transformations (XSLT) Version 2.0**: W3C Recommendation 23 January 2007. Cambridge, 2007. Disponível em: <<http://www.w3.org/TR/xslt20>>. Acesso em: 6 jul. 2014.

WOODLEY, M. S. Crosswalks, Metadata Harvesting, Federated Searching, Metasearching: Using Metadata to Connect Users and Information. In: BACA, M. (Org.). **Introduction to Metadata**. 2nd ed. Los Angeles: Getty Research Institute, 2008. Disponível em: <http://www.getty.edu/research/publications/electronic_publications/intrometadata/path.pdf>. Acesso em: 6 jul. 2014.

ZAFALON, Z. R. **Scan for MARC**: princípios sintáticos e semânticos de registros bibliográficos aplicados à conversão de dados analógicos para o Formato MARC21 bibliográfico. 2012. 169 f. Tese (Doutorado em Ciência da Informação) – Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília, 2012. Disponível em: <<http://base.repositorio.unesp.br/handle/unesp/103386>>. Acesso em: 6 jul. 2014.

APLICAÇÃO DE DADOS INTERLIGADOS ABERTOS APOIADA POR ONTOLOGIA

LINKED OPEN DATA APPLICATION SUPPORTED BY ONTOLOGY

Linair Maria Campos
Maria Luiza de Almeida Campos

Resumo: Aplicações de dados interligados abertos têm se constituído em poderoso instrumento para promover a difusão do conhecimento, na medida em que colocam em contexto informações de naturezas diversas e afins, fornecendo ao receptor um significado estabelecido justamente pelas interações entre essas informações. Nesse cenário, ontologias atuam como um mecanismo conceitual que dará suporte à consistência dessa rede de ligações entre conceitos, e, consequentemente, às aplicações que se proponham a explorá-la. O objetivo desse trabalho é apresentar uma aplicação de dados interligados abertos como prova de conceito, apoiada pelo uso de uma ontologia, com o foco em servir como base para explicitar informações relacionadas a teorias científicas de uma área do conhecimento. A fundamentação teórico-metodológica parte de um levantamento na literatura sobre ontologias relacionadas ao conhecimento científico como apoio para aplicações de dados interligados abertos. Com base nessa fundamentação, tem-se como resultado um exemplo de aplicação real de dados interligados abertos que permite apontar as conexões existentes entre um conjunto de teorias presentes em artigos científicos, as quais versam sobre um mesmo conceito, alvo de um objeto de pesquisa. Esse objeto de pesquisa é referenciado em instâncias na ontologia proposta, a qual pode ser disponibilizada como uma base de dados interligados aberta. Esse exemplo de aplicação permitiu evidenciar possíveis interligações não antecipadas entre as teorias presentes nas instâncias da ontologia e a Dbpedia, bem como responder a uma série de questões com base no modelo da ontologia. Como conclusão, são apontadas evidências da importância do uso de ontologias como uma forma de apoiar aplicações de dados interligados abertos.

Palavras-chave: Ontologias. Dados interligados abertos. Teorias científicas. Aplicativo.

Abstract: Open linked data applications have constituted a powerful tool to promote the diffusion of knowledge, as it put into context related information of various natures, giving the recipient an established meaning represented by the interactions between these information. In this scenario, ontologies serve as a conceptual mechanism that will support the consistency of this network of connections between concepts, and consequently, the applications that propose to exploit it. The aim of this paper is to present an application of linked open data as a proof of concept, supported by the use of an ontology with a focus on serving as a basis for explicit information related to scientific theories of an area of knowledge. The theoretical and methodological basis consists of a survey of the literature on ontologies related to scientific knowledge as support for open linked data applications. Based on this reasoning, we have as a result an example of an implementation of open linked data, which point out the differences between a set of theories found in scientific articles, which deal with the same concept, the target of a search object connections. This research object is referenced in the proposed ontology instances, which can be made available as a basis for open linked data. This sample application has highlighted possible unanticipated interconnections between ontology instances and DBpedia, as well as answers a series of questions based on the ontology model. In conclusion, it is pointed out evidence of the importance of using ontologies as a way to support the open linked data applications.

Keywords: Ontologies. Linked open data. Scientific Theory. Application.

1 INTRODUÇÃO

Artigos científicos costumam seguir uma metodologia de pesquisa, que implica em adotar teorias, métodos e técnicas de pesquisa. O conjunto de teorias usado pelo autor reflete um recorte possível, dentre visões afins e contrárias a seu ponto de vista. Este ponto de vista implica um conjunto de pressupostos sobre o mundo que é alvo de seu estudo, ou seja, seu paradigma de pesquisa, o qual fornece um arcabouço conceitual e filosófico para guiá-lo, norteando inclusive a escolha de métodos a serem adotados (PONTEROTTO, 2005).

As noções de paradigma e teoria às vezes se confundem, mas um paradigma possui um escopo mais amplo que uma teoria (MERTENS, 2010). Enquanto um paradigma fornece uma visão geral de mundo, uma teoria representa um conjunto de conceitos relacionados, definições e proposições para explicar e predizer um determinado fenômeno (COHEN *et al.*, 2007, p.12). Um paradigma, então, poderia nortear a formulação de uma teoria, ou, em outras palavras, uma teoria poderia estar alinhada com um paradigma.

Em uma teoria os conceitos ocupam um lugar central, pois são elementos básicos para a elaboração de proposições, que, por sua vez, são elementos básicos das teorias (GRENON; SMITH, 2011). Dessa forma, uma teoria possui um conjunto básico de conceitos que lhes confere um escopo (FRIEDMAN, 2003), sendo que a maneira pela qual esse mundo é percebido depende desse conjunto de conceitos (COHEN *et al.*, 2007, p.14). Dessa forma, dado um determinado conceito que é alvo de um objeto de pesquisa, pode-se ter atrelado a ele uma série de teorias afins e antagônicas entre si.

Entretanto, os detalhes de um objeto de pesquisa, ou seja, seus conceitos, teorias, proposições, métodos, dentre outros, nem sempre são explicitados e por vezes são até adotados de forma não consciente pelo autor. Tal fato pode ocasionar uma maior dificuldade na compreensão do contexto em que se insere a problemática do objeto de pesquisa, e especificamente, na busca por teorias afins e em contraponto com a posição do autor. Cabe destacar ainda que o critério adotado para a seleção das teorias pelo autor é naturalmente limitado pela possibilidade de se manipular um volume cada vez maior de publicações disponíveis e das limitações dos próprios mecanismos existentes para a sua recuperação.

Idealmente, na ausência de mecanismos que automatizem a descoberta de teorias relacionadas a objetos de pesquisa (dada a complexidade dessa tarefa), estas poderiam ser descritas em bases de dados temáticas, e disponibilizadas através de metadados acessíveis de forma livre e padronizada. Para isso, faz-se necessária a existência de vocabulários que

contenham os conceitos a serem utilizados para a descrição de objetos de pesquisa e suas teorias e pressupostos. Nesse sentido, tendo-se em mente a descoberta de conhecimento relacionado, como, por exemplo, outras teorias de autores com pressupostos afins com os adotados por um pesquisador, aplicações de dados interligados abertos podem ser úteis. Embora o foco do presente artigo não esteja voltado para a elaboração dos vocabulários necessários para a descrição das teorias, é possível ilustrar a utilidade de aplicações de dados interligados abertos nesse cenário através do uso de ontologias de mais alto nível, incentivando-se, como trabalhos futuros, a criação de tais vocabulários de forma mais detalhada.

Nesse sentido, o uso de uma aplicação de dados interligados abertos de teorias presentes em artigos científicos apoiada por ontologia poderia ajudar a entender o contexto de tais teorias e a sua possível relação com outras, permitindo ainda evidenciar possibilidades de aplicações de saberes de natureza interdisciplinar. O objetivo desse trabalho é apresentar uma implementação de prova de conceito de uma aplicação de dados interligados abertos apoiada por uma ontologia com o foco em explicitar informações relacionadas a teorias que embasam artigos científicos, exemplificadas em uma área do conhecimento. A proposta é dita prova de conceito pelo fato de não se constituir em uma aplicação com interface customizada para uso geral. Embora essa customização não seja uma tarefa complexa, ainda não se encontra concluída. Nesse sentido, a ferramenta apresentada é um trabalho em aperfeiçoamento fruto de um esforço colaborativo conduzido no âmbito de um Grupo de Pesquisa em Ciência da Informação.

Dados abertos são instrumentos para o avanço do conhecimento científico, na medida em que se constituem em fontes de dados úteis de livre acesso que têm sido publicadas por diferentes organizações de interesse público (LIU *et al.*, 2011), muitas delas ligadas à comunidade científica. Quando esses dados são interligados em um contexto, podemos perceber de forma mais precisa o seu significado e a partir daí, obter conhecimento.

É importante destacar que o significado da informação é obtido não através da especificação do seu significado em si, o que não é possível de forma inequívoca, mas especificando como a informação interage com outra informação (JAIN *et al.*, 2011), formando um contexto em que a semântica pode ser percebida de forma indireta.

Corroborando com essa afirmativa, situa-se Kitcher (1989) que, no contexto de uma discussão sobre como se dá o entendimento do conhecimento, comenta que “compreender o fenômeno não é simplesmente uma questão de reduzir as “incompreensibilidades

fundamentais”, mas de perceber conexões, padrões comuns, naquilo que inicialmente parecia ser situações diferentes”. Para que essa rede de informações com significado se estabeleça é necessário que pessoas façam um esforço extra na codificação de informações em representações passíveis de processamento automático. Com esse esforço computadores terão condições de processarem, interpretarem e concatenarem dados. Nesse cenário se situa a importância de se planejar o modelo dessas informações, ou seja, quais os conceitos, as suas naturezas, características, e de que forma se relacionam uns com os outros. Esse modelo ajuda a delinear a precisão das possibilidades de busca e entendimento da informação, tanto para as máquinas quanto para as pessoas. A máquina cumpre o seu papel de manipular de forma eficiente e sistemática grandes quantidades de informação, muitas vezes a partir de ferramentas de software já existentes, que podem ser usadas de modo combinado, ainda que customizadas, ou adaptadas para demandas específicas. As pessoas, aqui entendidas como os profissionais da informação, cumprem o papel de fornecer os princípios que o software deve seguir para que possa tratar a informação de forma inteligente, por exemplo, os modelos, que não podem ser definidos à revelia da validação de pessoas e seus pontos de vista, sejam eles sociais, políticos, epistemológicos e/ou pragmáticos.

Nesse cenário, o foco do presente artigo se apresenta no âmbito dos estudos em dados abertos interligados enfatizando a questão da interligação semântica de dados de artigos científicos da área de Ciência da Informação, com foco no estudo de Ontologias.

O restante do artigo está estruturado como se segue: na seção 2 é apresentado o panorama de dados abertos interligados e sua aplicabilidade e ainda ferramentas que podem ser usadas no contexto da aplicação proposta; na seção 3 é feita uma discussão de trabalhos relacionados a abordagens de dados interligados abertos voltada para artigos científicos na temática de ontologias; na seção 4 é apresentada a aplicação de dados interligados abertos desenvolvida e os resultados obtidos e, por fim, na seção 5, conclusões e trabalhos futuros.

2 DADOS INTERLIGADOS ABERTOS: CONCEITO E FERRAMENTAS PARA INTERLIGAÇÃO DE CONCEITOS

Bizer *et al.* (2009) definem dados interligados abertos²⁰ como “um conjunto de boas práticas para estruturar e publicar dados estruturados na web”.

As iniciativas de dados interligados abertos, doravante referenciadas como LOD, têm tomado proveito do avanço de padrões web e são baseadas nos seguintes princípios: (1)

²⁰ *Linked Open Data* (LOD) em inglês.

uso de URIs (*uniform resource identifier*) como nomes para entidades; (2) uso de URIs via HTTP (*hypertext transfer protocol*), de modo que se possam buscar informações por esses nomes na web; (3) informações úteis associadas às URIs, usando padrões tais como RDF (*Resource Description Framework*) e SPARQL (*Protocol and RDF Query Language*) (W3C, 2008); (4) inclusão de associações com outras URIs, de modo que se possam descobrir mais entidades (BIZER *et al.*, 2009).

URIs via HTTP na prática se constituem em um mecanismo para atribuir a cada entidade (concreta, abstrata, ou ainda um conceito qualquer) na web um identificador único, através do qual o recurso pode ser referenciado, ligado a outros recursos, ou se pode recuperar uma descrição do recurso que a URI representa.

RDF é um formato padrão para representação de dados na web. Esse formato permite que se representem fatos através de triplas na forma de sujeito, predicado e objeto, que, por sua vez, representam entidades concretas ou abstratas do mundo real. Cabe destacar que conteúdos RDF podem estar disponíveis através de documentos (páginas) RDF, mas também podem estar disponíveis através de bancos de dados de triplas RDF. Seja como for, a linguagem SPARQL permite buscas nesse conteúdo, distribuído em diferentes locais na web, de forma transparente, como se fosse uma única fonte de dados. Além disso, SPARQL também inclui um protocolo para criação de serviços de fornecimento de dados na web (SPARQL endpoints), os quais são acessíveis de forma usual através da web, e que aceitam pesquisas, sendo os resultados fornecidos em formatos padronizados tais como XML e RDF, o que facilita a sua interligação com outros dados na web (D'AQUIN, 2012).

Essa interligação, que pode ser automática ou semi-automática, é feita ligando-se URIs e pode ser de duas formas: através do casamento de conceitos definidos em ontologias ou através de casamento de suas instâncias, ou seja, quando se tenta determinar se dois indivíduos se referem a uma mesma entidade de um domínio (SINGH *et al.*, 2013). Ainda, de acordo com Singh *et al.* (2013), muitas ferramentas têm sido propostas para ambas as formas de interligação, porém a tarefa tem se mostrado árdua, e muitas das propostas são voltadas para domínios específicos, ou para um tipo específico de dados. Entretanto, a ferramenta SILK (Silk Link Discovery Framework) (ISELE *et al.*, 2010) parece se destacar nesse cenário, sendo uma das mais usadas e eficientes para interligar instâncias de dados em RDF.

SILK é uma ferramenta de código aberto (*open source*) para descoberta semi-automática de interligações entre entidades de diferentes fontes de dados RDF, de acordo com heurísticas pré-definidas que podem ser configuradas pelo usuário.

A ferramenta SILK possui as seguintes características principais: (i) suporta a geração de links do tipo *owl:sameAs* (para indicar equivalência de instâncias), assim como outros tipos de links RDF; (ii) fornece uma linguagem declarativa flexível para especificar as condições para que sejam propostas ligações; (iii) pode ser empregada em ambientes distribuídos sem que seja necessário replicar os conjuntos de dados localmente; (iv) pode ser usada em situações onde termos de diferentes vocabulários são misturados e não há um esquema consistente por sobre esses dados; (v) implementa vários métodos para aumentar o desempenho e a carga na rede (VOLZ *et al.*, 2009).

Uma forma alternativa ao uso de ferramentas prontas para determinação automática de links é a implementação, em um software específico, de algoritmos baseados em padrões (*pattern-based algorithms*), os quais exploram padrões de nomeação de URIs para propor interligações. Por exemplo, dadas duas fontes de dados RDF alvo, se o ISBN de um livro fizer parte de duas URIs diferentes dessas fontes, o algoritmo poderá propor que se tratam de conceitos equivalentes, gerando a relação *owl:sameAs* para conectar as duas URIs (YU, 2011). A partir daí, por exemplo, queries SPARQL podem ser utilizadas para apresentar os resultados de conceitos considerados semelhantes, ou seja, que estejam ligados pela relação *owl:sameAs*. O desafio que se coloca a partir daí é identificar quais os conceitos que podem ser interligados e quais as relações possíveis de se estabelecer entre eles. Para isso é preciso conhecer os elementos básicos do domínio.

3 TRABALHOS RELACIONADOS A ONTOLOGIAS E DADOS LIGADOS ABERTOS

Na busca por trabalhos relacionados, foram consideradas ontologias e aplicações de dados interligados abertos que pudessem ser aplicadas à descrição de pesquisas científicas. Optou-se por um escopo de busca mais amplo, em vez de buscar apenas por trabalhos com foco em descrição de teorias, por entender que as teorias se situam no escopo das pesquisas científicas, levando a conceitos relacionados, como, por exemplo, autores, áreas de pesquisa e linha filosófica adotada. Cabe destacar que essa ontologia que aqui se coloca é a do artefato de software, e não a ontologia da filosofia, e seu papel é especificar e formalizar as entidades, relações e axiomas que correspondem ao ponto de vista ontológico adotado pelo pesquisador.

3.1 Ontologias aplicáveis ao conhecimento científico

Ontologias, em seu uso na Ciência da Informação, fornecem um modelo para representar os pressupostos epistemológicos e ontológicos, relevantes para o entendimento

de pesquisas e seu tratamento computacional através das iniciativas de dados interligados abertos, mas sua elaboração é um processo custoso. Por essa razão, é importante utilizar em alguns casos ontologias já existentes, devido à redução de esforços e ao favorecimento da padronização:

Uma parte fundamental no processo de publicação inclui a determinação e criação de vocabulários e ontologias que fornecem o modelo de dados subjacente aos conjuntos de dados interligados, objetivando uma visão mais integrada dos dados e a maximização da interoperabilidade semântica entre conjuntos de dados e entre produtores-consumidores. Outra parte importante na definição dos modelos de dados é a maximização do reúso e da extensão de vocabulários e ontologias existentes. Este é um elemento fundamental na redução do esforço envolvido no consumo e integração de conjuntos de dados interligados (CORDEIRO *et al.*, 2011, p.84).

Na literatura, foram encontradas algumas iniciativas de descrição de vocabulários com foco em aspectos específicos, tais como Bonomi *et al.* (2006) para descrição de dados de publicações científicas da arqueologia, Biolchini *et al.* (2006) para revisão sistemática de artigos de engenharia de software, Malheiros e Marcondes (2011) para publicação científica na área biomédica, Bibliographic Ontology (BIBO, 2014) para descrever citações e referências bibliográficas na web semântica, Colibrary (GENDARMI; LANUBILI, 2009) para descrever dados bibliográficos e de revisões de usuários. Essas iniciativas são voltadas, de modo geral, para os dados bibliográficos da publicação, como por exemplo, seu autor, veículo de publicação, ou então para capturar os elementos da temática da publicação, ou ainda a estrutura de afirmações científicas feitas nos artigos, neste último caso com o foco em descoberta de conhecimento.

De fato, até onde pudemos perceber, não foi encontrada uma proposta de ontologia voltada especificamente para publicações científicas que contemplasse detalhes sobre as teorias utilizadas por eles. Entretanto, a ontologia Phylo (GRENON; SMITH, 2011), ontologia da Filosofia, criada por Pierre Grenon e Barry Smith, este último filósofo e ontologista que atua na comunidade biomédica, foi considerada aplicável aos propósitos do presente trabalho, mediante adaptação. Traçando um modelo em alto nível da Filosofia, onde se insere a Epistemologia, Phylo permite uma visão geral sobre aspectos epistemológicos e ontológicos da pesquisa científica, incluindo as teorias e suas relações:

Os conceitos filosóficos são o mais próximo que chegaremos a unidades básicas de atividades filosóficas. Proposições filosóficas são em primeira aproximação feita de conceitos. E as teorias são feitas de proposições (GRENON; SMITH, 2011, p.17).

A ontologia proposta por Smith propõe, de forma preliminar, quatro categorias ou classes principais, a saber: *Philosophical Object* (Objeto Filosófico), *Philosophical Field*

(Área da Filosofia), *Person (philoshoper)* (Pessoa (filósofo)), e *Group of Persons (group of philosophers)* (Grupo de pessoas (grupo de filósofos)). Para a classe *Philosophical Object* (Objeto Filosófico), Smith propõe as seguintes subclasses:

- *concept* (conceito: por exemplo, o conceito de forma)
- *proposition* (proposição: por exemplo, que forma existe)
- *theory* (teoria: por exemplo, teoria das formas, de Platão)
- *argument* (argumento: por exemplo, argumento do Terceiro Homem, de Platão)
- *method* (método: por exemplo, o método dialético).

A ontologia Phylo propõe ainda um conjunto de relações binárias, como, por exemplo: *memberOf* (entre *person* e *group*), *subgroupOf* (entre *group* e *group*), *workedOn* (entre *person* e *concept* e entre *group* e *concept*), *subconceptOf* (entre *concept* e *concept*), *activeIn* (entre *person* e *field* e entre *group* e *field*), *inField* (entre *concept* e *field* e entre *philosophical entity* e *field*), *subfieldOf* (entre *field* e *field*).

Cabe destacar que a ontologia Phylo embora conte com a associação de um conceito a uma pessoa ou grupo de pessoas, ela não explicita que esse mesmo conceito pode ter definições diferentes, de acordo com o viés determinado pelo grupo de pessoas que trabalha no conceito. Por exemplo, podem-se citar pelo menos três definições diferentes para o conceito de tropo²¹ (ROJEK, 2008, p.367), cada uma defendida por um grupo de pessoas diferente. Dessa forma, como esse importante aspecto não fica explicitado, pode dar a entender que grupos de pessoas distintos trabalham com a mesma noção, o que é equivocado. Na verdade, elas podem trabalhar o mesmo conceito, porém com diferentes pontos de vista. Isso é determinado pela definição. Se pudermos dizer que as definições são as mesmas, então podemos chegar à conclusão que grupos distintos convergem em relação a um conceito. Do contrário, fica explicitada a divergência.

Embora a ontologia Phylo possua a relação *subConceptOf* que relaciona *concept* a *concept*, Smith deixa claro que essa relação é apenas para fins de especialização de um conceito, não estando essa especialização vinculada a uma teoria: “Nós usamos *subconceptOf* como a relação entre dois conceitos quando o primeiro é uma especialização do segundo. Além disso entendemos que a especialização é definitória e não sujeita a debate filosófico, assim, em particular, não é dependente de uma teoria” (GRENON; SMITH, 2011, p.17).

Por outro lado, a ontologia Phylo possui a entidade *proposition*, através da qual se poderia diferenciar os conceitos usados nas teorias. Entretanto, para que fosse possível

²¹ São elas: Teoria de Tropos sem Universais, Teoria de Tropos com Universais Determináveis e Teoria de Tropos com Universais Concretos.

chegar à conclusão que as teorias convergem, teríamos de associar um conjunto de proposições (com seus respectivos argumentos) às respectivas correntes filosóficas. Se tomarmos a definição de conceito de Dahlberg (1978a), onde este é uma unidade de conhecimento formada pelo referente, o termo e proposições sobre o referente, é possível considerar *proposition* como sendo parte de *concept* e, assim, associá-las indiretamente a *group of persons* (representantes das correntes filosóficas). Nesse caso, teríamos a noção de um conceito com um viés, este determinado pelo grupo de pessoas.

3.2 Aplicações de dados interligados abertos relacionadas ao conhecimento científico

Apesar de o foco do presente trabalho não ser implementar uma aplicação de LOD para *descrição* de recursos, é importante disseminar o uso e compreender os aspectos envolvidos nesse tipo de iniciativa, que se constitui em uma realidade cada vez mais proeminente na web dos dias atuais:

Criar maneiras de facilitar o uso das informações e elaborar um ambiente no qual as informações estejam integradas deve ser objetivo dos profissionais da informação, que certamente têm um espaço na instrumentalização dessas tecnologias (SANTOS NETO *et al.*, 2013, p.86).

Dentre as aplicações de LOD pesquisadas relatadas a seguir, observou-se de fato que o diferencial entre elas é ontologia que apoia a aplicação em si. Entretanto, algumas aplicações apresentam aspectos úteis para uma aplicação de LOD, os quais podem ser utilizados em trabalhos futuros.

Börner *et al.* (2012) apresentam uma solução denominada VIVO, a qual tem o foco na informação sobre pesquisa e pesquisadores e fornece, além de uma ontologia, ferramentas para gerenciar a ontologia e uma plataforma para usar a ontologia e criar e gerenciar os dados interligados abertos a ela associados. Bonomi *et al.* (2006), por sua vez, apresentam um protótipo de uma biblioteca eletrônica (*e-Library*), a qual permite descrições e buscas semânticas nas coleções descritas, cujo vocabulário é voltado especificamente para publicações científicas da área de Arqueologia.

Kramer *et al.* (2012) propõem o uso de RDF com o foco em descrever e interligar dados trabalhados pelas pesquisas de ciências sociais na web. Contempla, dentre outros, os seguintes aspectos: interligação de publicações, pessoas e organizações a dados de pesquisa, interligação a dados geográficos, a estudos relacionados, e interligação de dados a licenças. Os autores utilizam-se do recurso de identificadores persistentes largamente usados na publicação acadêmica, como por exemplo DOIs (*Digital Object Identifiers*), para identificar unicamente os dados de pesquisa, assim como nas citações da publicação. Dessa forma, é

possível identificar, por exemplo, publicações que utilizem as mesmas fontes de dados, ou ainda, publicações que utilizem as mesmas bases teóricas. Embora essa proposta se aproxime do objetivo do presente trabalho, é importante destacar que existe uma diferença primordial, posto que Kramer *et al.* explicitam o aporte teórico na forma de links para as publicações científicas e não para as teorias ali utilizadas nem as linhas epistemológicas adotadas. Nesse sentido, funcionam mais como uma rede de citações bibliográficas, embora contextualizadas.

Todas essas aplicações evidenciam a utilidade e versatilidade de aplicações de LOD no que tange a representar de forma articulada dados do conhecimento científico. O desafio que se coloca para essas aplicações parece centrar-se não na infraestrutura tecnológica, mas sim na concepção do modelo da ontologia do domínio, uma vez que os resultados por elas apresentados são todos, em sua essência, dependentes dessa ontologia e nela apoiados.

4 APLICAÇÃO DE LOD APOIADA POR ONTOLOGIAS: MÉTODOS E RESULTADOS

Nas seções anteriores foi ilustrado o papel das ferramentas de software e dos aspectos epistemológicos e ontológicos voltados para a representação do domínio escolhido. Apontou-se também, embora não esteja no escopo do presente trabalho, a importância dos aspectos metodológicos, voltados para a compreensão do problema. Partindo desses conhecimentos, espera-se poder demonstrar a utilidade da aplicação de LOD para interligar teorias presentes em artigos científicos da temática de ontologias, e, através de um exercício simples dessa aplicação, voltar a lembrar a importância do papel do profissional da informação no uso de ferramentas de software e na modelagem da informação, conforme proposto por Campos *et al.* (2012).

4.1 Ontologia do conhecimento científico para uso na aplicação LOD

Conforme a recomendação de publicação de dados interligados abertos, procedeu-se à descrição de um conjunto de dados relacionados a teorias presentes em publicações científicas da temática de ontologias. Para isso, utilizou-se a ontologia do conhecimento científico que é apresentada a seguir. Essa ontologia foi inspirada na ontologia Phylo, e, tal como esta, é uma proposta preliminar, ainda em aperfeiçoamento.

Das entidades propostas por Smith na ontologia Phylo, para os propósitos do presente trabalho, foram reutilizadas as seguintes: conceito, teoria, área, pessoa e grupo de pessoas. Cabe observar que a ontologia Phylo é específica da Filosofia, enquanto que na presente proposta, o foco foi adaptado para a descrição do conhecimento contido em artigos

de pesquisa acadêmica, objetivando contemplar aspectos básicos desse conhecimento. A adaptação se aplica, pois está ancorada na descrição dos aspectos epistemológicos dos artigos, sendo que a Epistemologia é um ramo da Filosofia. Assim sendo, à semelhança de Smith, a FIGURA 1 ilustra exemplos de uso dos conceitos reutilizados.

Cabe lembrar que se faz necessário o desenvolvimento de vocabulários que descrevam as instâncias dos conceitos de alto nível descritos pela ontologia proposta, e que poderiam ser descritas em bases de dados temáticas, conforme mencionado na introdução do presente artigo. Porém esse não é o foco desse trabalho, e, nesse sentido, serão utilizados exemplos, de modo a viabilizar a utilidade da aplicação LOD proposta.

Ainda, com base na ontologia Phylo, a ontologia para o conhecimento científico também possui a preocupação para definir suas relações e minimizar ambiguidade na definição de conceitos. Corroborando com essa preocupação, na Ciência da Informação, Dahlberg (1978a, 1978b) destaca a importância das definições, na medida em que tornam explícito o conteúdo dos conceitos, fornecendo os elementos que vão forjar as relações entre eles. Dahlberg, através de sua Teoria do Conceito, propõe ainda que as definições revelam um conjunto de características comuns que são úteis para construir qualquer sistema classificatório (DAHLBERG, 1983).

Considerando a importância da definição dos conceitos da ontologia, explicamos a seguir de que modo são entendidas as entidades presentes na ontologia proposta, a saber: *research object* (objeto de pesquisa), *concept* (conceito), *biased concept* (conceito com viés), *theory* (teoria), *field* (área), *person* (pessoa) e *group of persons* (grupo de pessoas), *proposition* (proposição), *basic proposition* e *differential proposition*. Por motivos de simplificação, não será considerada a entidade *argument*.

Por *research object* entende-se a classe genérica, que agrupa os conceitos de *concept*, *biased concept*, *bias*, *proposition*, *basic proposition*, *differential proposition* e *theory*.

Por *field* entende-se a área de pesquisa na qual se insere um pesquisador ou seu objeto de pesquisa. Uma área de pesquisa pode estar subordinada a outra área de pesquisa. Por *person* entende-se o pesquisador que atua ou atuou em uma área de pesquisa, e o qual trabalha ou trabalhou em um objeto de pesquisa. Por *group of persons* entende-se a posição filosófica de um grupo de pessoas a qual pertence um pesquisador.

Por *concept* entendemos um objeto básico alvo de uma temática de pesquisa.

Por *biased concept* entendemos um *concept* porém com um viés que o difere de outros *biased concept*. Esse viés se manifesta através de um conjunto específico de

proposições, a saber, *differential proposition*. Os *biased concept* são assim especializações de um *concept* na medida em que todos concordam com um conjunto básico de proposições que está definida em *concept (basic proposition)*, embora cada um deles possua seu próprio conjunto de proposições que os difere entre si. *Biased concept* foi criada para reforçar a noção do viés de um grupo de pessoas trabalhando em um conceito. Considera-se que essa entidade expressa um ponto de vista consensual de um grupo de pessoas sobre um determinado conceito, sendo definida através de um conjunto de proposições específico. Naturalmente, considera-se que o ponto de vista expressa uma diferença, mas também algum consenso sobre alguma coisa. Por exemplo, têm-se duas visões distintas para um mesmo conceito básico de tropo feito por grupos diferentes de pessoas: tropos são abstratos e tropos são particulares. Essas proposições podem nos levar a conclusões importantes, dependendo do escopo de instâncias de teorias de tropos considerado. Tomando-se o ponto de vista de Rojek (2008), por exemplo, todas as teorias de tropos têm em comum a proposição de que “tropos são particulares” (esse seria o conceito básico de tropo). Em contrapartida, algumas teorias de tropos consideram que tropos são abstratos, enquanto que outras consideram que tropos são concretos. Dessa forma, esse conceito consensual, baseado no qual derivam conceitos afins, porém com diferentes vieses, é expresso na ontologia como *concept*. Os diferentes pontos de vista são os *biasedConcept*. Seria possível representar ambos como *concept*, mas nesse caso não ficaria explícita essa diferença importante de que existe um conjunto básico de proposições consensuais, *basicProposition*, que dá origem a um *concept* que é um conceito base para uma divergência de posições de pesquisa, divergência essa representada pelo diferente conjunto de proposições específicas, *differentialProposition*, que está associado a cada *biasedConcept*.

Um *concept* é definido ou trabalhado por um *groupOfPersons*, mas a tendência é que esse grupo seja de um escopo mais amplo do que o grupo de pessoas que trabalha no *biasedConcept*. Por essa razão, entende-se que os *biasedConcept*, são, de fato, trabalhados por um *GroupOfPersons* que é um *subgrupo* de um determinado *GroupOfPersons*. Entretanto, para fins de simplificação, esses detalhes não serão explicitados no mapa conceitual proposto para a ontologia do conhecimento científico (FIGURA 1).

Por *proposition* entendemos uma afirmativa sobre um referente no mundo. Uma proposição é expressa através de conceitos, embora sejam, por sua vez, parte de um conceito, como propõe Dahlberg. *Proposition* pode ser *differentialProposition* ou *basicProposition*.

Por *theory* entendemos um conjunto de afirmativas sobre o mundo, incluindo o que Giere (2004) descreve como teoria, lei e princípio.

Quanto às relações, as principais que foram adotadas, reutilizadas da Phylo, são: *memberOf* (membro de), *workedOn* (trabalhou em), *subgroupOf* (subgrupo de), *subconceptOf* (sub conceito de), *inField* (na área), *subfieldOf* (sub área de), *activeIn* (atua em).

As relações *subgroupOf*, *subfieldOf* e *subconceptOf* são usadas para denotar especialização para as entidades *Group*, *Field* e *Concept*, respectivamente, no sentido original adotado por Smith. Além dessas, foram utilizadas as relações *IsA* (gênero-espécie), para associar um *biasedConcept* a um *Concept* implicando um viés teórico distinto, *partOf* (e sua inversa *hasPart*), para associar um *BiasedConcept* a uma *Theory*, e para associar um *concept* a uma *basicProposition*. A relação *partOf* também é usada para associar uma *proposition* a uma *Theory*. Cabendo observar que nem toda *proposition* que compõe uma *Theory* pode estar associada a um *BiasedConcept* que faz parte da *Theory*. Por exemplo, uma *Theory* pode apresentar a *proposition* “formas existem”, sem que essa proposição específica esteja associada a um *BiasedConcept*.

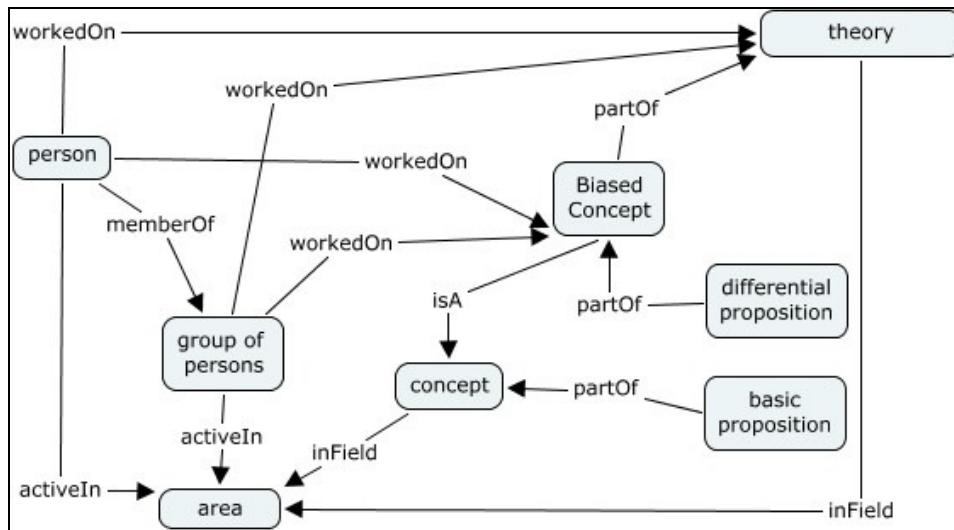
A relação *inField* entre um conceito objeto de pesquisa e uma área de pesquisa ocorre quando é o caso do conceito ser estudado pela área, como, por exemplo, no caso do conceito de categorias, o qual é estudado pela Ciência da Informação. Cabe destacar que o fato de um objeto de pesquisa ser estudado por uma área não impede que ele seja estudado também por outra área, embora provavelmente através de teorias diversas.

A relação *workedOn* é a relação entre uma *person* ou *groupOfPersons* e uma *theory*. Por exemplo, entre D.C. Williams e a Teoria de Tropos. É usada também para relacionar *person* ou *groupOfPersons* a *concept* ou a *BiasedConcept*.

Os conceitos da ontologia proposta são ilustrados através de um mapa conceitual, na FIGURA 1, sendo apoiado em uma adaptação da ontologia Philo. Nem todas as relações explicadas anteriormente estão ilustradas na FIGURA 1, por motivos de simplificar a visualização do modelo. Por exemplo, o conceito de *Proposition* foi omitido. Dessa forma, suas relações com *theory* (*partOf*) e com *differential proposition* e *basic proposition* (*IsA*) foram omitidas.

Um trecho do código OWL de um conjunto de instâncias desse modelo, usado no experimento descrito na seção 4 é ilustrado na FIGURA 2.

FIGURA 1: mapa conceitual simplificado da ontologia para o conhecimento científico



A ontologia proposta permite responder a várias perguntas, tais como: (i) de quais teorias o conceito x faz parte; (ii) de quais teorias o conceito x faz parte e são trabalhadas por um grupo de pessoas y; (iii) a que área pertence a pessoa x que trabalha na Teoria y; (iv) que pessoas trabalham na teorias x e em que áreas; (v) de que teorias fazem parte os conceitos x com a proposição y; (vi) qual o grupo de pessoas que trabalha no conceito x.

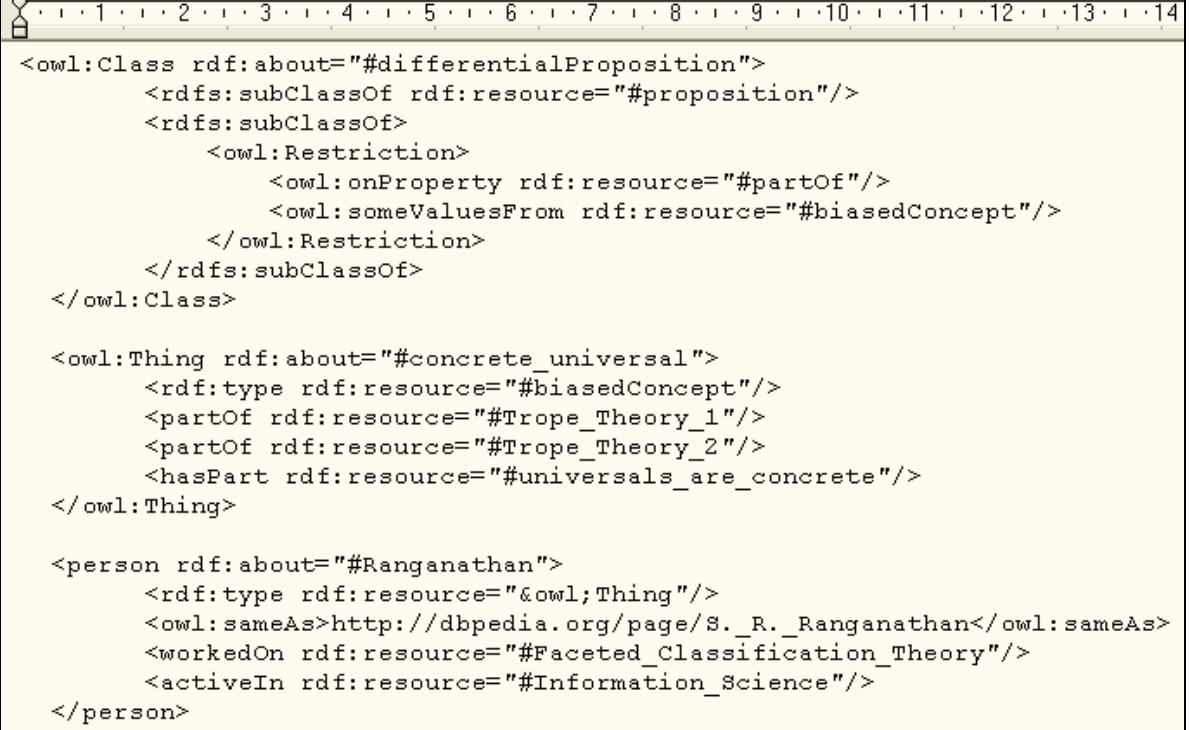
O conjunto de instâncias da ontologia do conhecimento utilizado como exemplo ao ser explorado responde a algumas questões específicas, conforme ilustrado na FIGURA 2. Essas respostas são fornecidas com base em Rojek (2008) e Dahlberg (1978a).

Algumas das questões respondidas são: (i) o *concept* “universal” faz parte das *theory* de Tropos1, Tropos2, Tropos3 e Teoria do Conceito; (ii) o *biasedConcept* de “universal concreto”, trabalhado pelo *groupOfPersons* “filósofos da pan-unidade”, faz parte das *theory* de Tropo1 e Tropo2; (iii) a *person* Ingetraut Dahlberg trabalha na Teoria do Conceito e atua na área da Ciência da Informação; (iv) as *person* Aristoteles, Ingarden e Husserl trabalham na *theory* Teoria de Tropos na *area* da Filosofia; (v) o *biasedConcept* “non-abstract universal”, com a *basicProposition* “universals are not abstract”, faz parte das *theory* Teorias de Tropos 1, 2, 3 e 4; o *biasedConcept* “concrete universal” com a *differentialProposition* “universals are concrete” faz parte das *theory* Teorias de Tropos 1 e 2; (vi) o *groupOfPersons* “realismo moderado” trabalha no *biasedConcept* “abstract universal”.

Uma semântica mais rica de descrição, como proposto, pode ser o ponto de partida para ajudar a construir perguntas mais objetivas e bem construídas. A modelagem das questões relevantes para o domínio, incluindo seus relacionamentos e a descrição e

interligação semântica dos seus dados, permitem resultados mais precisos. Por um lado, por auxiliar o interessado a formular a sua pergunta de forma objetiva e, por outro lado, por permitir que esses aspectos possam ser considerados como critério pelas ferramentas de software para trazer informação relevante relacionada.

FIGURA 2: Trecho da ontologia OntoSC



```


    <rdfs:subClassOf rdf:resource="#proposition"/>
    <rdfs:subClassOf>
        <owl:Restriction>
            <owl:onProperty rdf:resource="#partOf"/>
            <owl:someValuesFrom rdf:resource="#biasedConcept"/>
        </owl:Restriction>
    </rdfs:subClassOf>
</owl:Class>

<owl:Thing rdf:about="#concrete_universal">
    <rdf:type rdf:resource="#biasedConcept"/>
    <partOf rdf:resource="#Trope_Theory_1"/>
    <partOf rdf:resource="#Trope_Theory_2"/>
    <hasPart rdf:resource="#universals_are_concrete"/>
</owl:Thing>

<person rdf:about="#Ranganathan">
    <rdf:type rdf:resource="&owl;Thing"/>
    <owl:sameAs>http://dbpedia.org/page/S.\_R.\_Ranganathan</owl:sameAs>
    <workedOn rdf:resource="#Faceted_Classification_Theory"/>
    <activeIn rdf:resource="#Information_Science"/>
</person>

```

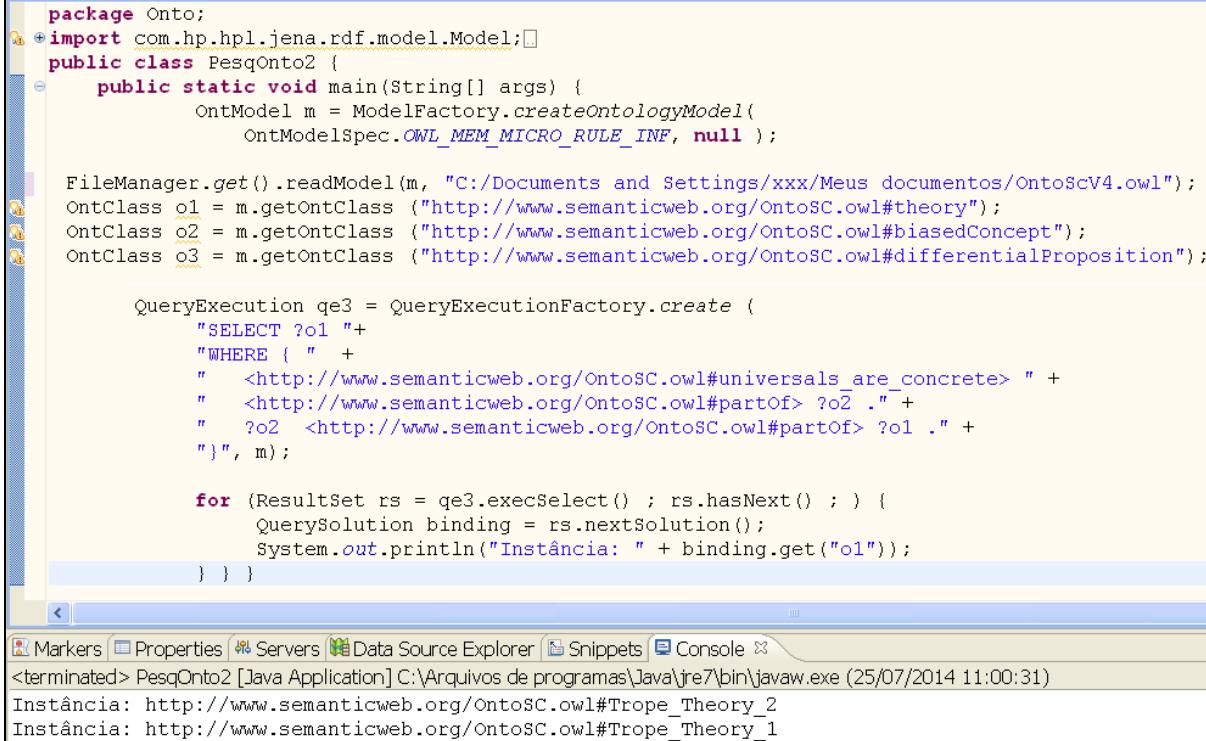
4.2 Aplicação LOD e resultados obtidos

A aplicação de LOD proposta consiste em uma prova de conceito de uma ferramenta de software, mas envolve mais do que simplesmente uma codificação em uma linguagem de programação. Para que essa aplicação fosse concebida, foi necessário todo um estudo, conforme foi explanado nas seções anteriores, que permitiu o embasamento necessário para a compreensão dos problemas e dos requisitos a serem implementados na solução computacional. Além disso, envolveu um esforço da definição do modelo de uma ontologia, da seleção de um conjunto de dados adequado e da descrição desse conjunto como instâncias da ontologia, e ainda, da codificação da aplicação. Por questões de espaço, será apresentada uma breve explanação das funcionalidades da aplicação bem como do conjunto de dados utilizado.

A aplicação foi implementada em Java com uso da API Jena (2014), sendo a ontologia desenvolvida por meio da ferramenta Protégé OWL (2014). Fornece funcionalidades para que sejam feitas consultas, incluindo inferências, na base de instâncias

OWL cujo trecho é apresentado na FIGURA 2. Foi testada com sucesso para responder às perguntas apresentadas na seção 4.1, sendo que o resultado de um desses testes é ilustrado na FIGURA 3, o qual visa responder à pergunta “quais as teorias que têm como parte a proposição diferencial *universals_are_concrete*”.

FIGURA 3: Trecho do código Java que faz busca na ontologia OntoSC, com resultado



```

package Onto;
import com.hp.hpl.jena.rdf.model.Model;
public class PesqOnto2 {
    public static void main(String[] args) {
        OntModel m = ModelFactory.createOntologyModel(
            OntModelSpec.OWL_MEM_MICRO_RULE_INF, null );
        FileManager.get().readModel(m, "C:/Documents and Settings/xxx/Meus documentos/OntoScV4.owl");
        OntClass o1 = m.getOntClass ("http://www.semanticweb.org/OntoSC.owl#theory");
        OntClass o2 = m.getOntClass ("http://www.semanticweb.org/OntoSC.owl#biasedConcept");
        OntClass o3 = m.getOntClass ("http://www.semanticweb.org/OntoSC.owl#differentialProposition");

        QueryExecution qe3 = QueryExecutionFactory.create (
            "SELECT ?o1 " +
            "WHERE { " +
            "    <http://www.semanticweb.org/OntoSC.owl#universals_are_concrete> " +
            "    <http://www.semanticweb.org/OntoSC.owl#partOf> ?o2 ." +
            "    ?o2 <http://www.semanticweb.org/OntoSC.owl#partOf> ?o1 ." +
            "}" , m);

        for (ResultSet rs = qe3.execSelect() ; rs.hasNext() ; ) {
            QuerySolution binding = rs.nextSolution();
            System.out.println("Instância: " + binding.get("o1"));
        } } }

```

The screenshot shows an IDE interface with the Java code above. Below the code, the Eclipse IDE's bottom bar is visible with tabs for Markers, Properties, Servers, Data Source Explorer, Snippets, and Console. The Console tab shows the output of the application's execution:

```

<terminated> PesqOnto2 [Java Application] C:\Arquivos de programas\Java\jre7\bin\javaw.exe (25/07/2014 11:00:31)
Instância: http://www.semanticweb.org/OntoSC.owl#Trope_Theory_2
Instância: http://www.semanticweb.org/OntoSC.owl#Trope_Theory_1

```

A interligação de dados se dá através dos elementos *theory*, *person* e *field*, sendo feita com a Dbpedia²². A relação para indicar a interligação com a Dbpedia é owl:sameAs. Para a descoberta dos possíveis recursos alvo da Dbpedia (no caso do exemplo, Ranganathan) foi feita uma busca manual, que, por ocasião de uma versão futura, poderá ser substituída pela busca automatizada da ferramenta SILK. A busca na DBpedia também foi implementada e apresentada como resultado complementar. Um extrato do código com a query SPARQL utilizada é ilustrada na FIGURA 4. Essa query tem como finalidade exibir todas as informações encontradas na Dbpedia associada ao recurso indicado na relação owl:sameAs. Dessa forma, foi possível complementar as informações contidas na ontologia OntoSC relacionadas a *Person* e de forma organizada, através de ligações nomeadas, trazendo uma semântica adicional que é um atrativo para o aprendizado. No caso de Ranganathan, por exemplo, foi possível obter a partir da Dbpedia informações sobre as

²² www.dbpedia.org

susas obras notáveis (*notableworks*), a saber, *Prolegomena of Library Classification*, *Ramanujan: the Man and the Mathematician*, *The Five Laws of Library Science*, dentre outras informações de interesse, conforme ilustrado na FIGURA 4.

FIGURA 4: Trecho do código Java com a query SPARQL para busca na Dbpedia com parte do resultado



```

String service = "http://dbpedia.org/sparql";
String query = "PREFIX dbo:<http://dbpedia.org/ontology/> " +
    "SELECT ?PropertyValue ?PropertyName " +
    "WHERE { " +
    "<http://dbpedia.org/resource/S._R._Ranganathan> " +
    "?PropertyName ?PropertyValue." +
    " }";

```

Markers Properties Servers Data Source Explorer Snippets Console <terminated> PesqDbpedia [Java Application] C:\Arquivos de programas\Java\jre7\bin\javaw.exe (28/0
http://dbpedia.org/property/notableworks
Prolegomena to Library Classification@en
http://dbpedia.org/property/notableworks
Ramanujan: the Man and the Mathematician@en
http://dbpedia.org/property/notableworks|
The Five Laws of Library Science@en

5 CONCLUSÃO

Através de LOD, a web passa a disponibilizar um volume enorme de dados, não só documentos, que ao serem modelados em triplas RDF, permite que a web seja encarada como um grande banco de dados passível de inspeção por ferramentas de software já disponíveis na atualidade. Nesse cenário, ontologias são importante instrumento para explicitar de forma menos ambígua as informações que se disponibilizar de forma livre na web.

Para mostrar essa realidade, apresentou-se um exemplo de aplicação de dados interligados abertos apoiada por uma ontologia para descrever teorias científicas ligadas a áreas do conhecimento. Essa aplicação foi concebida a partir de uma revisão teórica na área de ontologias relacionadas ao conhecimento científico e em dados interligados abertos, tendo como objetivo servir como prova de conceito da utilidade e da viabilidade de se implementar uma solução tecnológica apoiada pela ontologia. A partir do modelo da ontologia proposta, foi criado um conjunto de instâncias de teste o qual serviu para evidenciar, como resultado, a utilidade do modelo para responder a um conjunto pré-determinado de perguntas. Trabalhos futuros podem abordar a questão de se criar um conjunto real de instâncias e de tornar a implementação publicamente disponível através da adaptação de sua interface de busca para a web. Como conclusão tem-se que o uso de

ontologias é útil para ajudar a conceber um modelo mais preciso com foco em responder perguntas, cujas respostas transcendem os dados da base de instâncias local, estendendo o conhecimento para uma nuvem de possibilidades ampliadas, concretizando o objetivo pretendido das aplicações de dados interligados abertos.

REFERÊNCIAS

- BIBO, **The Bibliographic Ontology**, 2014. Disponível em: <<http://bibliontology.com/>>. Acesso em: 27 jul. 2014.
- BIZER, C., HEATH, T., BERNES-LEE, T. Linked Data. The story so far. **International Journal on Semantic Web and Information Systems**, v. 5., n. 3., p. 1-22, 2009.
- BONOMI, A. *et al.* A framework for ontological description of archaeological scientific publications. In: ITALIAN SEMANTIC WEB WORKSHOP, 3., Pisa. **Proceedings...**, Pisa: CEUR, 2006.
- BÖRNER *et al.* VIVO: A Semantic Approach to Scholarly Networking and Discovery. **Synthesis Lectures on the Semantic Web: Theory and Technology**, v. 1., n. 7., p.1-178, 2012.
- CAMPOS, L.M. et al. Dados abertos interligados e o espaço do profissional de informação: Uma aplicação no domínio da enfermagem. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 13., Rio de Janeiro. **Anais eletrônicos...** Rio de Janeiro: ICICT, 2012.
- COHEN, L.; MANION, L.; MORRISON, K. **Research Methods in Education**. New York: Routledge, 2007.
- CORDEIRO, K.F. *et al.*. An approach for managing and semantically enriching the publication of Linked Open Governmental Data. In: WORKSHOP IN APPLIED COMPUTING FOR ELECTRONIC GOVERNMENT, 3., Florianópolis. **Proceedings...** Florianópolis: WCGE, 2011.
- DAHLBERG, I. A referent-oriented, analytical concept theory of Interconcept. **International Classification**, v. 5, n. 3, p. 122-151, 1978a.
- DAHLBERG, I. Conceptual compatibility of ordering systems. **International Classification**, v. 10, n. 2, p. 5-8, 1983.
- DAHLBERG, I. Teoria do conceito. **Ciência da Informação**, v. 7, n. 2, p. 101-07, 1978b.
- D'AQUIN, M.: Putting Linked Data to Use in a Large Higher-Education Organisation. In: INTERACTING WITH LINKED DATA WORKSHOP, 1., Heraklion. **Proceedings...** Heraklion: ILD, 2012: p. 9-21.
- FRIEDMAN, K. Theory construction in design research: Criteria: Approaches, and methods. **Design Studies**, v. 24, n. 6, p. 507-522, 2003.
- GENDARMI, D.; LANUBILI, F. 2009. Colibrary: Linking Communities of Readers on the Social Semantic Web. In: SCRIPTING AND DEVELOPMENT FOR THE SEMANTIC WEB WORKSHOP, 6., Heraklion, Proceedings..., Heraklion: CEUR-WS.org, 2009, p.1-2.

- GIERE, R. How Models Are Used to Represent Reality. **Philosophy of Science**, v. 71, Supplement, p. 742-752, 2004.
- GRENON, P; SMITH, B. Foundations of an Ontology of Philosophy. **Synthese**, v. 182, n. 2, p. 185-204, 2011.
- ISELE, R.; JENTZSCH, A.; BIZER, C. Silk Server - Adding missing Links while consuming Linked Data. In: INTERNATIONAL WORKSHOP ON CONSUMING LINKED DATA, 1., Shanghai, **Proceedings...** Shanghai: CEUR-WS.org, 2010, p.1-12.
- JAIN, P. *et al.* **How To Make Linked Data More than Data**, 2010. Disponível em: <<http://corescholar.libraries.wright.edu/cgi/viewcontent.cgi?article=1094&context=knoesis>>. Acesso em: 28 jul. 2014.
- JENA. **A Semantic Web Framework for Java**, 2014. Disponível em: <<http://jena.sourceforge.net/>>. Acesso em 28/07/2014.
- KITCHER, P. Explanatory Unification and the Causal Structure of the World. In: KITCHER, P.; SALMON, W. **Scientific Explanation**. Minneapolis: University of Minnesota Press, 1989.
- KRAMER, S. *et. al.* **Using RDF to Describe and Link Social Science Data to Related Resources on the Web**, 2012. Disponível em: <<http://pub.uni-bielefeld.de/publication/2524325>>. Acesso em: 25 ago. 2014.
- LIU, Q; *et al.* Linking Australian Government Data for Sustainability Science: A Case Study. In: WOOD, D. **Linking Government Data**, New York: Springer, 2011.
- MALHEIROS, L.R.; MARCONDES, C.H.. Identificación de los rasgos de descubiertas científicas en artículos biomedicos. **Revista EDICIC**, v. 1, n. 4, p. 30-48, 2011.
- MERTENS, D. A. Research and evaluation in education and psychology. Thousand Oaks: Sage, 2010.
- PONTEROTTO, J.G. Qualitative Research in Counseling Psychology: A Primer on Research. Paradigms and Philosophy of Science. **Journal of Counseling Psychology**, v. 52, n. 2, p. 126-136, 2005.
- PROTÉGÉ OWL, **Editor de Ontologias em OWL**, 2014. Disponível em: <<http://protege.stanford.edu/>>. Acesso em: 28 jun. 2014.
- ROJEK, P. Three Trope Theories. **Axiomathes**, v. 18, n. 1., p. 359-377, 2008.
- SANTOS NETO, A.L. *et al.* Tecnologias de dados abertos para interligar bibliotecas, arquivos e museus: um caso machadiano. **Transinformação**, v. 25, n. 1, p. 81-87, 2013.
- SINGH, R, *et al.* A graphical user interface for SILK data link discovery framework. In: International Symposium on Open Collaboration, 9., New York. **Proceedings...** New York: ACM, 2013, p.241-242.
- VOLZ, J. *et al.* Silk – A Link Discovery Framework for the Web of Data. Workshop on Linked Data on the Web, 2., Madrid. **Proceedings...** Madrid: LDOW, 2009, p. 1-6.
- W3C, 2008. **SPARQL Query Language for RDF**, 2008. Disponível em: <<http://www.w3.org/TR/rdf-sparql-query/>>. Acesso em: 10 jul. 2014.

YU, L. A developer's guide to the Semantic Web. New York: Springer, 2011.

USUÁRIOS EM PRÁTICAS PROFISSIONAIS DE BIBLIOTECÁRIOS E ANALISTAS DE TI

USERS IN PROFESSIONAL PRACTICES OF LIBRARIANS AND IT PROFESSIONALS

Eliane Cristina de Freitas Rocha
Adriana Bogliolo Sirihal Duarte

Resumo: Este artigo tem como objetivo apresentar comparações das práticas de bibliotecários e analistas de Tecnologia da Informação em aspectos concernentes à usabilidade, acessibilidade, construção e uso de sistemas de informação digitais, resultantes de pesquisa de doutorado, realizada junto a 17 profissionais bibliotecários e 16 analistas de tecnologia da informação. Entre os entrevistados com perfil de analistas de requisitos/analistas de sistemas e bibliotecários com experiência em construção de sistemas digitais, verificou-se a aplicação de técnicas de pesquisa comuns para realização de estudos de usuário: entrevistas e análise de domínio, no projeto de sistemas informatizados. Já as técnicas qualitativas (observação, entrevista) e métodos técnicos da engenharia de software e usabilidade (prototipação) são aplicados por analistas de tecnologia da informação com perfil de designers de interação ou especialistas em usabilidade.

Palavras-chave: Bibliotecário. Analista de sistemas. Estudo de usuário. Prática profissional.

Abstract: This paper aims to present comparisons of librarians and information technology analysts' professional practices concerning usability, accessibility, building and use of digital information systems, as final results of doctoral dissertation research held among 17 librarians and 16 technology information analysts. It has been found, among interviewers with system's requirements/systems' analysts profiles and librarians experienced in building digital systems, the use of research techniques in common to carry on user studies: interviews and domain analysts, in projecting information systems. Otherwise, qualitative techniques (observation, interviews) and software engineering and usability techniques (prototyping) were applied by Information Technology Analysts in interaction designer or usability specialist roles.

Keywords: Librarian. System analyst. User studies. Professional practice.

1 INTRODUÇÃO

Este artigo tem como principal objetivo apresentar resultados e reflexões originadas em trabalho de doutorado (ROCHA, 2013) que visou responder à seguinte questão: “De que maneira têm sido abordados os usuários da informação nas práticas profissionais de bibliotecários e analistas de sistemas?”.

Se, por um lado, o trabalho procurava abordar uma importante lacuna observada na literatura da Ciência da Informação (CI) - carência de trabalhos acadêmicos brasileiros que refletam sobre as práticas efetivas de trabalho de bibliotecários em relação aos seus usuários - por outro, explorava pontos de intersecção das práticas profissionais relativas aos projetos de sistemas de informação, usabilidade e acessibilidade. Tais pontos são explorados em

revisão bibliográfica e teórica, com breve exposição apresentada na próxima seção deste trabalho (seção 2).

Em sequência, foi realizado um estudo de campo com 17 profissionais bibliotecários e 16 analistas de tecnologia da informação, por meio de entrevistas semiestruturadas, que permitiram levantar as técnicas de coleta de dados de pesquisa utilizadas para abordagem de usuários da informação (observação, diários de campo, questionários, entrevistas, entre outras) no projeto de sistemas e em aspectos relacionados à usabilidade, acessibilidade e suporte ao usuário. Os resultados do trabalho de campo são apresentados na seção 3. Por fim, são tecidas considerações finais em relação a aproximações e diferenças das práticas profissionais na seção 4.

2 ABORDAGENS DOS USUÁRIOS NA CI E NA COMPUTAÇÃO

Com os propósitos de conhecer os usuários de unidades de informação e contribuir com o seu planejamento, González Teruel (2005) aponta os estudos de usuários na CI como estudo das necessidades da informação, estudo de demanda de informação, estudo de usos da informação, estudos de satisfação com os serviços informacionais, e estudos de impacto; por outro lado, os estudos de usuários, na visão de Figueiredo (1994) também envolvem o estudo das comunidades de usuários.

Tanto Choo (2003) quanto González Teruel (2005) reconhecem que tais estudos podem ser orientados ao sistema (abordagem tradicional, de orientação metodológica quantitativa) – “buscam caracterizar como a informação flui no ambiente externo do indivíduo pelos sistemas sociais” (CHOO, 2003, p. 68), que privilegia o emprego de técnicas quantitativas – ou orientados ao usuário (abordagem alternativa, tradicionalmente associada a métodos qualitativos) – buscam revelar “as preferências, necessidades cognitivas e psicológicas do indivíduo, como elas afetam a busca e os padrões de comunicação da informação” (CHOO, 2003, p. 68-69).

Se, no âmbito da CI, o campo de estudos de usuários se volta para o avanço tanto de teorias quanto de técnicas para compreensão dos usuários da informação, a área de Computação, no seu subcampo denominado engenharia de software, procurou desenvolver estudos com a visão do desenvolvedor de software que tornam os sistemas interativos mais “eficientes, robustos, livres de erros, e de fácil manutenção” (BARBOSA; SILVA, 2010, p. 8). Foi somente a partir da quarta geração de computadores (a geração dos minicomputadores do final da década de 1970) em diante, que a figura do usuário de sistemas com suas necessidades passa a ser considerada no processo de software, embora as abordagens da engenharia de software surgidas a partir daquele momento fossem centradas

usualmente na qualidade de construção do software a partir de uma visão de “dentro para fora”, com ênfase “na representação de dados, algoritmos que processam esses dados, arquitetura do sistema e tudo o mais que permite um sistema interativo funcionar” (BARBOSA; SILVA, 2010, p. 8).

A concepção de sistemas interativos construídos de uma perspectiva de “fora para dentro”, com ênfase primeira no contexto de uso (necessidades e interesses dos usuários, domínio da atividade, tarefas realizadas, impactos no ambiente, etc) é o foco dos estudos da área de Interação Humano-Computador (IHC) – surgida em âmbito mundial a partir da década de 1990 – e, “sob alguns aspectos, também [da] área de Engenharia de Requisitos” (BARBOSA; SILVA, 2010, p. 9). As duas abordagens – de fora para dentro e de dentro para fora – talvez se assemelhem à distinção existente no campo da CI dos estudos de usuários da informação centrados no sistema, por um lado, e centrados no usuário, por outro.

A Engenharia de Requisitos refere-se aos processos de desenvolvimento (extração, discussão e análise de requisitos) e gestão de requisitos (documentação, validação e manutenção de requisitos), conforme Attarha e Modiri (2011). Na análise de requisitos, a “equipe [...] de desenvolvimento [...] trabalha] com o cliente e os usuários finais do sistema para descobrir [...] informações sobre o domínio da aplicação, que serviços o sistema deve oferecer, o desempenho exigido do sistema, as restrições de hardware e assim por diante” (SOMMERVILLE, 2003, p. 104). As principais técnicas de extração de requisitos são entrevistas, cenários, realização de *brainstorming*, observação e análise, grupos de foco, reuso de requisitos, prototipação, *Soft System Methodology* (SSM), JAD (*Joint Application Development*) e QFD (*Quality Function Deployment*) (ATTARHA; MODIRI, 2011, p. 182).

Embora os usuários sejam essenciais no processo de desenvolvimento de sistemas computadorizados, em especial nas fases de elicitação e definição de requisitos, implantação e operação de sistemas (NASCIMENTO, 2003), o estudo sistemático acerca dos papéis dos usuários se desenvolveu em campo de estudo à parte da engenharia de software – a IHC – donde foi gestada a engenharia de usabilidade. Barbosa e Silva (2010, p.10-11), apresentam cinco tópicos de interesse de estudo da IHC: natureza da interação humano-computador e suas consequências na vida das pessoas; contexto de uso (cultura, sociedade, organizações, ambiente); características humanas (físicas, perceptivas, emocionais, entre outras) que interferem no uso de sistemas; arquitetura de sistemas e

interfaces (como os dispositivos de interação e os estilos de interação) e o processo de desenvolvimento de sistema interativo visando alta qualidade de uso.

O campo de estudos é bastante amplo, e, para os objetivos deste artigo, chama-se atenção do leitor de duas importantes orientações de pesquisa em IHC: por um lado, o desenvolvimento teórico que passa pela elaboração de modelos e leis da interação humano-computador advindos da influência da psicologia cognitiva (em suas variadas expressões desde o cognitivismo até a cognição situada e distribuída) e da psicologia experimental, bem como da semiótica²³; e, por outro lado, o desenvolvimento de técnicas para projeto de interação orientados por abordagens do *design* enquanto disciplina pelas engenharias de requisitos e de usabilidade – esta última com a proposição de técnicas de Design Centrado no Usuário (DCU).

Segundo Fallman (2011), é possível distinguir três ondas de desenvolvimento dos estudos da IHC: a primeira, da década de 1980, com base nos modelos da engenharia cognitiva de processamento humano, associados à perspectiva do computador como ferramenta (fase em que se desenvolveu o conceito de usabilidade e técnicas de DCU, como enunciadas por NIELSEN, 1993, entre outros autores); a segunda, dos anos 1990, com base teórica do design participativo, cognição distribuída e teoria da atividade contou com o emprego de técnicas da etnografia, etnometodologia e fenomenologia, entre outros; a terceira, a partir do final dos anos 1990, conta com a progressiva incorporação de circuitos digitais em produtos diversificados (computação embarcada), a ascensão das tecnologias móveis (como celulares, *smartphones* e *tablets*, entre outros dispositivos), o que faz com que o projeto da interação das pessoas com tantos dispositivos (e não apenas computadores) seja cada vez mais complexo. Na terceira onda, o estudo da interação das pessoas com o computador se estende para estudos ampliados de design da interação: “o design de produtos interativos que fornecem suporte às atividades cotidianas das pessoas, seja no lar ou no trabalho” (PREECE; ROGERS; SHARP, 2005, p.28), preocupação que também se liga à avaliação de comportamento informacional de usuários em ambientes de

23 Barbosa e Silva (2010) apresentam algumas abordagens teóricas em IHC como as inspiradas na psicologia experimental (como a lei de Hick-Hyman - que estima o tempo requerido para opções de escolha em interfaces; ou a lei de Fitts – que relaciona o tempo médio para se alcançar um alvo na interface); na psicologia cognitiva aplicada (como o modelo de processador humano de Card, Moran e Newell; como a aplicação dos princípios da Gestalt à elaboração de interfaces); da engenharia cognitiva (com as teorias da atividade de inspiração cognitivista até a aplicação de teorias da atividade e Leontiev e Engeström); da cognição distribuída e das abordagens etnometodológicas; e na semiótica (tanto a aplicação da semiótica quanto da proposição da engenharia semiótica).

entretenimento, apresentada por Case (2007). Questões como o design de experiência e a ampliação da multidisciplinaridade nos projetos de software, com a interação de profissionais do design e da comunicação, se estabelecem.

Ao longo de seu desenvolvimento, a disciplina de IHC problematizou diversas qualidades de uso dos sistemas interativos como usabilidade, acessibilidade, comunicabilidade, aplicabilidade, usabilidade universal, experiência de uso, entre outros. De tais qualidades, a usabilidade é definida pela norma ISO 9241 (NBR 9241), citada por Barbosa e Silva (2010, p.29), como “a capacidade que apresenta um sistema interativo de ser operado, de maneira eficaz, eficiente e agradável, em determinado contexto de operação, para a realização das tarefas de seus usuários”. Já a acessibilidade de sistemas computacionais é definida por Barbosa e Silva (2010) como importante qualidade de uso dos sistemas associada à flexibilidade da interface para se adaptar aos usuários com diferentes necessidades (cegos, surdos, com problemas motores, disléxicos, entre outros). A qualidade da experiência de uso é um conceito em discussão, mas endereça frequentemente aspectos subjetivos da relação dos usuários com os produtos (PREECE; ROGERS; SHARP, 2005). A *findability*²⁴ é, por sua vez, a qualidade da construção de sistemas associada à arquitetura de ambientes digitais e não é inserida inteiramente nos estudos de IHC.

O projeto de sistemas que tenha boas qualidades de uso (como boa usabilidade, acessibilidade, entre outras) está associado ao desenvolvimento de processos de construção de produtos interativos centrados nos usuários, com a proposição de técnicas de DCU e processos de engenharia de usabilidade e suas derivações (design de interação, design de informação, design de experiência)²⁵. Silva (2012) comenta que não há consenso sobre o processo de design de interação na área de IHC, sendo este composto por etapas iterativas: a) identificação de necessidades (por técnicas como entrevistas, observação e investigação contextual) e análise de requisitos de usuário e análise de tarefas; b) design e redesign (projeto da interação e interfaces, com uso de protótipos) e c) avaliação (com o emprego de

²⁴ Propriedade que reflete uma boa arquitetura de informação de um ambiente informacional relativa à facilidade com que o usuário encontra o que precisa em suas estratégias de navegação e busca (por *browsing*, *searching* ou busca direta) (MORVILLE; ROSENFELD, 2006).

²⁵ Há emergência de conceitos e práticas relativas ao projeto/design de produtos interativos como: 1) Design de interação: projeto de interações das pessoas com produtos interativos, o que inclui o design de interfaces; 2) Design da informação: projeto de organização dos conteúdos em sua forma e linguagem, levando em conta aspectos da arquitetura da informação; 3) Design de experiência: projeto de um produto interativo que leva em conta metas de alto nível que os usuários têm, independentemente do produto (exemplo: projetar a experiência de um carro de luxo que instigue sentimentos ou atitudes leva a incluir atributos e características diferentes de um carro popular).

testes com usuários – teste de usabilidade e comunicabilidade; ou testes sem a presença dos usuários – avaliação heurística, teste por *check-list*, percurso cognitivo). Por outro lado, na perspectiva da área do design, também a condução de projetos interação passa por etapas semelhantes, conforme a metodologia do *design thinking* (VIANNA *et al.*, 2012) – método que busca a inovação de produtos e serviços em aspectos humanos, que envolve quatro etapas: a) Imersão (compreensão empática do usuário/cliente/consumidor por meio de observação e métodos etnográficos com vistas a identificar um *insight* para a solução de um problema de design); b) análise e síntese (reflexões a partir do *insight*, com uso de técnicas como o mapa de empatia - um diagrama dividido em seis áreas cujo centro é a caracterização do cliente estudado pelas observações de seu comportamento); c) Ideação (gerar ideias inovadoras, com participação de equipes multidisciplinares); d) Prototipação.

Barbosa e Silva (2010) mencionam que algumas técnicas auxiliam no design de um novo produto ou de um novo projeto de interação em substituição de um já existente: entrevistar usuários ou fazer pesquisas de campo para aprender sobre o produto antigo; consultar processos e normas para entender as restrições de uso; fazer análise de *logs* de uso e fazer uma análise competitiva. As técnicas para conhecer o perfil do usuário mais citadas são os questionários, entrevistas, grupos focais e workshops, observação natural e estudo de documentação (PREECE; ROGERS; SHARP, 2005, p.235). Barbosa e Silva (2010) acrescentam à lista o *brainstorming*; a investigação contextual, na qual o investigador (*designer*) exerce o papel de aprendiz do trabalho do usuário (modelo mestre-aprendiz); e a classificação por cartões (técnica do *card sorting*) para organização da informação. Para estarem mais próximos do universo do usuário, os projetistas devem registrar os perfis dos usuários de acordo com características relevantes à construção do sistema – como familiaridade com computadores, conhecimento do domínio da aplicação, idade, entre outros aspectos. Barbosa e Silva (2010) apresentam a técnica de criação de personas – personagens fictícios representativos dos perfis típicos de usuários do sistema – para melhorarem a definição do sistema.

Conhecido o usuário, é importante saber como eles realizam suas tarefas, através da análise de suas tarefas que envolve, normalmente, sua representação como uma sequência de sentenças curtas (conhecidas como listas de atividades) e a classificação das coisas envolvidas na realização das atividades (objetos e agentes). Há modelos para formalização das análises de tarefas como o GOMS (*Goals, Operators, Methods and Selection Rules*), HTA (*Hierarchical Task Analysis*), CTT (*ConcurTaskTrees* – Árvores de Tarefas Concorrentes), entre outros (BARBOSA; SILVA, 2010). As técnicas de coleta de dados

para realização de análise de tarefas envolvem observação e entrevista. Os estudos com inspiração etnográfica têm sido abordados na literatura como possibilidades de melhor compreensão das tarefas dos usuários (PREECE; ROGERS; SHARP, 2005).

Após conhecer o perfil dos usuários e suas tarefas, são traçadas metas de usabilidade - as mais conhecidas são facilidade de uso, facilidade de aprendizado, boa memorabilidade, segurança no uso, eficiência no uso, eficácia no uso, boa utilidade - e metas baseadas na experiência - vistas como subjetivas e associadas à atratividade, motivação, satisfação, entre outros aspectos – para os sistemas e aplicativos (PREECE; ROGERS; SHARP, 2005).

Com relação ao conjunto de técnicas propostas tanto para o projeto e avaliação de interações centradas no usuário da área de IHC, percebe-se que algumas delas são intensamente utilizadas pelo profissional designer ou analista, como a técnica da prototipação²⁶ (BALDUS; MACIEL; SOUZA, 2012). Baldus, Maciel e Souza (2012) comentam que é um grande desafio compreender o que o usuário precisa e que há relativa pouca utilização das técnicas de DCU (como análise dos usuários e suas tarefas, bem como avaliações interativas das propostas de interface) durante o desenvolvimento de software. Por outro lado, há dificuldades inerentes ao processo de análise de requisitos de sistemas computacionais – de falta de compreensão do usuário sobre os sistemas e dos analistas do domínio do usuário referenciada em Schwiderski (2011). Chama atenção, também, a pouca atenção conferida ao uso dos sistemas em si mesmos na literatura de IHC ou da engenharia de software. Sendo o uso dos sistemas colocado na etapa final do ciclo de vida de desenvolvimento de software, os processos de operação e treinamento de usuários recebem pouca atenção pela engenharia de software e Schwiderski (2011) aponta, em seu estudo de campo, carência de treinamentos dos usuários.

Para os propósitos deste artigo, serão apresentadas as abordagens dos profissionais bibliotecários e analistas de Tecnologia da Informação (TI) em relação a alguns dos aspectos que aproximam a área de CI com a computação: como os profissionais bibliotecários e analistas de TI abordam usuários para projetar ambientes informacionais;

²⁶ A prototipação é uma técnica utilizada tanto para eliciar e validar requisitos no âmbito da engenharia de requisitos, quanto para projetar interfaces e interações, no âmbito da IHC. Consiste em se desenvolver protótipos de alta ou baixa fidelidade do produto, como produtos interativos desenvolvidos em ferramentas como Axure ou Balsamic ou desenhos em papel, respectivamente. Tais protótipos são apresentados aos potenciais usuários para melhorar a comunicação usuário-analista e melhorar a definição de requisitos e/ou interfaces/interações do sistema.

como lidam com a usabilidade e a acessibilidade e realizam atividades de atendimento e treinamento de usuários.

3 PRÁTICAS DE TRABALHO DE BIBLIOTECÁRIOS E ANALISTAS DE TI RELATIVAS AOS USUÁRIOS

Foram realizadas entrevistas semiestruturadas com 17 bibliotecários e 16 analistas de TI, no período de junho a setembro de 2012. No quadro 1, são apresentadas os bibliotecários participantes (B1 a B17), atuantes em 13 diferentes instituições, em diversos tipos de biblioteca. No quadro 2, por sua vez, são visualizados os analistas de tecnologia da informação (A1 a A16), atuantes em fábricas de software (desenvolvedoras de software e prestadoras de serviço de TI, numeradas de A a F), centros de inovação (Centros A e B - empresas atuantes em projetos e produtos inovadores na área de TI) ou em empresas usuárias de tecnologia da informação (Instituição A; Banco A; Banco B; Empresa E; Empresa F).

QUADRO 1 – Participantes da pesquisa bibliotecários

Empresa		Idade	Gênero	Função
Universidade A (Privada)	B1	48	M	Diretor de biblioteca
	B2	53	F	Bibliotecária de referência
	B3	47	F	Coordenadora do setor de referência
Universidade B (Pública)	B4	36	F	Bibliotecária
	B5	32	F	Bibliotecária
Faculdade A (Privada)	B6	31	M	Diretor do sistema de bibliotecas
	B7	26	F	Bibliotecária e arquivista
Faculdade B (Privada)	B8	34	F	Bibliotecária
Empresa A (Autarquia)	B9	49	F	Bibliotecária
Empresa B (Privada, sem fins lucrativos)	B10	45	F	Bibliotecária
Escola A (Privada – 1º e 2º graus)	B11	29	F	Bibliotecária
Escola B (Pública – 1º grau)	B12	47	F	Bibliotecária
Biblioteca pública A	B13	38	M	Bibliotecário
Biblioteca pública B	B14	60	F	Bibliotecária
Empresa C (Biblioteca digital de instituição pública)	B15	27	M	Analista de pesquisa
Empresa D (Biblioteca digital de uma universidade)	B16	29	F	Designer instrucional
Fundação A (Acervo Biblioteca Nacional)	B17	50	F	Bibliotecária

Fonte: Elaboração própria a partir de dados da pesquisa, coletados entre junho/2012 a setembro/2012.

QUADRO 2 – Participantes da pesquisa analistas de TI

Empresa		Idade	Gênero	Cargo
Fábrica A	A1	43	M	Engenheiro de software
Fábrica B	A2	33	M	Consultor (Analista de requisitos)
Fábrica C	A3	40	M	Analista de requisitos
Fábrica D	A4	33	M	Gerente da equipe de desenvolvimento (Requisitos)
	A5	29	M	Líder de projeto (Engenheiro de usabilidade e requisitos)
Fábrica E	A6	28	M	Analista de sistemas (Consultor de implantação)
Fábrica F	A7	32	M	Analista de suporte
Fábrica G	A8	30	M	Sócio (consultor - designer de interação)
Centro A	A9	28	M	Designer de interação
	A10	31	F	Designer de interação
	A11	28	F	Designer de interface
Centro B	A12	34	F	Pesquisadora (Engenharia de usabilidade)
Banco A	A13	44	F	Coordenadora de equipe de serviços de suporte e infra-estrutura
Instituição A (empresa pública)	A14	37	F	Técnico judiciário (desenvolvimento de sistemas)
Empresa E (área de máquinas)	A15	39	F	Analista de sistemas
Empresa F (área de engenharia)	A16	33	F	Gerente de projeto sênior

Fonte: Elaboração própria a partir de dados da pesquisa, coletados entre junho/2012 a setembro/2012.

3.1 Práticas de abordagens de usuários na concepção de produtos interativos e sistemas de recuperação da informação

A participação do usuário no processo de desenvolvimento dos acervos de bibliotecas e construção de sistemas computacionais foi levantada nas entrevistas junto aos bibliotecários. Com relação à construção de sistemas de informação digitais, seis bibliotecários entrevistados (B5, B6, B7, B8, B15, B16) participavam ou participaram de projetos de sistemas de informação computadorizados, como sistemas de gestão eletrônica de documentos, repositórios digitais e bibliotecas digitais. Como tais projetos normalmente requerem processos que idealmente envolvem usuário – como o de elicitação de requisitos, análise de sistemas e projeto de interfaces da perspectiva computacional – algumas perguntas foram feitas aos bibliotecários entrevistados para se entender a participação dele no processo de desenho de sistemas e a possível aplicação de técnicas de pesquisas de usuário ou DCU em tais processos. Não se nota, nas falas dos bibliotecários com experiências em bibliotecas digitais e repositórios de dados (B6, B7, B15, B16) incorporação de todas as atividades de processo de DCU – como levantamento e modelagem de usuários e suas tarefas; estabelecimento de requisitos de interface e prototipação. Enquanto gestores dos repositórios, os bibliotecários das bibliotecas digitais (Empresa C e Empresa D) se envolvem mais com o ciclo documental - detalhando

processos de indexação e elaboração de resumos (B15, B16), padrões de interoperabilidade de repositórios (B16) – e aplicam conhecimentos relativos ao processamento técnico das bibliotecas e aspectos de gestão da informação, questões que não são contempladas na área da engenharia de software ou da IHC. O processo de representação do material é também um desafio, enfrentado de duas formas: ou com base em um perfil inferido de usuário e nas funcionalidades do sistema ou com a participação de um perfil de usuário (produtor de informação) na indicação de palavras-chave para recuperação – aspectos que evocam os estudos da área de arquitetura da informação.

Em contraste com as experiências dos repositórios e bibliotecas digitais, duas bibliotecárias (B5 e B8) apresentam experiências de análise de requisitos de sistemas de GED (Sistemas de Gerenciamento Eletrônico de Documentos), configurando ferramentas como Alfresco, Oracle ou Sharepoint. Elas relataram participar de processos de definição do escopo e do orçamento de customização e implantação dos sistemas, papel que exerciam em conjunto com a equipe de analistas de TI. A pesquisa de necessidades, nestes casos, era realizada junto ao cliente e/ou usuário, por meio de entrevistas de pessoas-chave na organização - técnica indicada pela engenharia de software e por participantes analistas de sistemas da pesquisa (A4, A5) - e há indícios da aplicação da técnica do mestre-aprendiz (técnica da investigação contextual de levantamento de dados junto a usuários em processos de DCU), com o usuário no papel de mestre e o analista de aprendiz dos seus processos de trabalho.

De maneira análoga à formação das coleções das bibliotecas pelo seu papel funcional, também os sistemas computadorizados podem ser construídos a partir do entendimento do negócio que será informatizado, a partir de uma análise de domínio, entendido o domínio como campo de aplicação do software. Há coleta de documentos relacionados às atividades dos usuários e seu entendimento para propósitos de organização da informação ou de concepção de sistemas. O estudo da documentação do usuário e de sistemas em uso ou similares também é apontado como referência pelos bibliotecários B5, B8 e B15: “porque lá no GED tinha vários sistemas, tinha *Document Image*, [...], *Document Manager*, [...], então a gente tinha que conhecer um pouco [de] [...] qual que era a possibilidade de cada sistema pra poder oferecer pro cliente, né” (B5). O conhecimento das possibilidades dos sistemas também aparece entre os analistas de sistemas A2, A6 e A16: “Normalmente [...], levanta-se a necessidade de um determinado cliente e constroi-se um sistema pra atender aquele cliente ou caso a gente tem o modelo próximo, a gente parte daquele modelo e customiza” (A6).

Além de conhecer dos sistemas similares (técnicas apontada por BARBOSA; SILVA , 2010) para criação de sistemas, também as impressões pessoais e experiências dos analistas contam na construção de sistemas, tanto para bibliotecários no projeto das bibliotecas digitais, quanto para analistas de TI: “Tem um pouco de demanda que vem dos próprios usuários, mas [...] a maioria eu diria que vem internamente, é da cabeça das pessoas. [...] Alguns vêm de usuários ou de relatos de problemas que eles estão tendo ou de validação assim de ideias com *focus groups*” (A1).

Se há semelhanças em aplicação de técnicas para levantar requisitos de sistemas – entrevistas com usuários-chave, análise de documentos/domínio da aplicação, pesquisa de sistemas similares, e até impressões pessoais – notam-se também diferenças entre analistas e bibliotecários na modelagem de sistemas. O emprego de técnicas de análise de requisitos (como o uso de pontos de função²⁷ pelos analistas de TI) e dos protótipos é indicado na fala dos analistas, e não na fala dos bibliotecários. Outro aspecto importante a se notar é que os analistas de TI (como A2, A3, A8) relatam a importância de formalizar/documentar os requisitos levantados em entrevistas – seja com atas de reunião, ou com compartilhamento de planilhas e documentos com a proposta do sistema junto aos clientes ou usuários (A3, A8).

Os clientes – contratantes do sistema – também têm suas próprias impressões que são levadas em conta para a construção do sistema: “às vezes o cliente já faz um levantamento prévio, ele faz um levantamento prévio de necessidades, aí o nosso trabalho é refinar isso a partir daí” (A2). Os interesses do cliente e dos usuários devem ser conjugados para a criação do produto, na visão de A8: “quando a gente trabalha requisitos, [...] a gente avalia o negócio né, dentro da nossa metodologia [...] [que vem do] *design thinking*, [...] essa parte de *problem framing*, de identificar os *stakeholders*, de identificar quais são os ganhos e as perdas de cada [um deles]”.

3.2 A abordagem da usabilidade

A temática da usabilidade não está incorporada no fazer dos profissionais bibliotecários entrevistados, nem nos processos de construção de sistemas de recuperação da informação, nem nos processos de avaliação da biblioteca e seus sistemas. Os bibliotecários que têm funções de projeto de sistemas não cuidam de aspectos relativos aos

²⁷ Técnica para medir o tamanho do software a partir de suas funcionalidades levantadas junto aos usuários. É muito útil para se medir produtividade em projetos de desenvolvimento e manutenção de produtos de software.

projetos de interação e interfaces e é possível dizer que todos os entrevistados bibliotecários não têm pleno entendimento do significado da usabilidade conforme a área de IHC – qualidade de uso dos sistemas associado à sua facilidade de operação na realização de tarefas dos usuários.

Dois bibliotecários (na biblioteca universitária e na biblioteca pública) citam, de passagem, a palavra usabilidade como ligada à utilização do acervo (como se confundissem usabilidade com uso). A facilidade de uso, acessibilidade e *findability* são atributos associados à usabilidade nas bibliotecas digitais (B15 e B16), embora não sejam realizadas pesquisas de usabilidade nos repositórios digitais ou nas bibliotecas digitais pelos bibliotecários. Ao ser perguntada sobre o processo de construção da biblioteca digital, B16 comentou sobre o processo de tratamento documental e aplicação de padrões de interoperabilidade de repositórios digitais:

Entrevistadora: tem alguma metodologia pra construção do repositório?

B16: tem, tem o padrão Oasis, né, Open Archives [...] esse processo é assim – tem a entrada, né, o processo da entrada de conteúdo, que vem pelos depositantes e que eu entro com uma mediação, tem todo tratamento da informação ali dentro do repositório, tem a divisão em comunidades, coleções, né, e tem aqui o usuário final que vai acessar isso do ponto de vista de recuperação da informação.

Se, por um lado, a usabilidade de sistemas de informação não é objeto de trabalho do bibliotecário, nota-se que a usabilidade não é incorporada em todos os processos de trabalho entre os analistas entrevistados com perfil de analistas de requisitos, de negócio e de suporte, sendo que alguns deles não souberem responder às questões específicas sobre a usabilidade no trabalho. Há outros entrevistados com o perfil de analistas de sistemas que vêem a usabilidade como o projeto de interfaces de boa qualidade, sem explicitar sua incorporação nas etapas de concepção dos produtos (A2, A3, A6).

Por outro lado, o ponto de vista dos entrevistados engenheiros de usabilidade e designers de interação e experiência (A4, A5, A10, A11, A12), é o da usabilidade em processo (incorporação das técnicas de DCU nos processos de criação de produtos interativos e de software). A participação da equipe de usabilidade, na visão deles, não pode se restringir a avaliar interfaces ao final do desenvolvimento, dada a limitação o uso de técnicas de inspeção, como avaliação heurística, para garantia da usabilidade, pois ela não reflete maturidade de processos de DCU (A9). Há queixas de falta de valorização da pesquisa de usuários tanto por parte de desenvolvedores de sistemas quanto dos clientes contratantes dos serviços:

Contrataram a gente pra poder fazer teste de usabilidade no sistema, que era horrível, muito ruim, e desenhar os protótipos de um sistema que

funcionaria melhor.[...] Fiz um protótipo muito bacana, [...] [fiz] protótipos funcionais, teste com protótipos, [...], só que são clientes muito mais velhos no mercado, eles estão lá há 20 anos, eles ganham muito dinheiro e quando a pessoa ganha muito dinheiro, ela acha que está fazendo tudo certo né, então quando mostrei pra eles que estavam fazendo umas coisas muito erradas, eles foram muitos reticentes e não aplicaram do jeito que eu queria (A8).

Os entrevistados designers de interação e analistas de sistemas com perfil de engenheiros de usabilidade aplicam técnicas da engenharia de usabilidade – como análise de perfil de usuário, com combinação de técnica quantitativa e qualitativa (A8 e A12) para pesquisar usuários e suas tarefas.

A metodologia, o método que a gente utiliza depende de cada projeto, porque cada projeto lhe impõe restrições e um contexto diferente. Então a gente pode ir a campo pra fazer entrevista, a gente pode ir a campo pra observar, a gente pode ir na casa do usuário, pode fazer uma ciberetnografia [...] [que], faz tempo que eu não faço isso, mas a gente já chegou por exemplo a levantar impressões de usuários sobre determinada [...] marca de celular, utilizando na época o Orkut (A9).

Há predomínio de técnicas qualitativas para se conhecer os usuários e suas tarefas, mas não há sistematização de análise de tarefas, conforme informa A9. São feitas análises de conteúdo das entrevistas categorizando os dados coletados (A9), embora o tempo seja exíguo para fazer uma análise aprofundada de tarefas (A10). Com ou sem aprofundamento, são geradas soluções de interações e interfaces por meio da criação de protótipos a serem validados pelos usuários²⁸.

No que diz respeito às técnicas de avaliação dos projetos de interação e de interfaces (como avaliações heurísticas, inspeção por *check-list* e outras), nota-se que as técnicas de inspeção são mais voltadas à validação de interfaces já finalizadas e mais facilmente integradas ao desenvolvimento, aparentemente. Já os testes de usabilidade que envolvem a observação das interações do usuário com o sistema são técnicas preferidas e privilegiadas para avaliação de usabilidade, ocorrendo ou não em ambiente de controlado (como em laboratório), nas fábricas A, D e G, centro A e centro B. Nota-se referência aos formalismos para condução dos testes de usabilidade conforme apontados na literatura (PREECE; ROGERS; SHARP, 2005), contendo termo de consentimento, questionário pré-teste, entrevista pós-teste, roteiro de tarefas em algumas falas (A5, A8, A12) e, em todas elas, a observação dos usuários em tarefas específicas é relatada. A análise dos resultados do teste

²⁸ No Centro A, o processo de criação de soluções segue um modelo de processo de design inspirado em uma empresa de design americana (IDEO) que envolve conjugar o interesse das pessoas/usuários, negócios e tecnologia. A geração de ideias segue a etapa de compreender os usuários e conjugar os seus interesses com os do cliente e as possibilidades da tecnologia.

de usabilidade pode ser mais ou menos formalizada, e ela é levada à equipe de desenvolvimento para que alterações sejam realizadas nas interfaces ou nas interações dos sistemas.

3.3 A abordagem da acessibilidade

Seis dos participantes bibliotecários não teceram considerações sobre a acessibilidade. Dos participantes que apresentaram suas impressões e as práticas relativas à acessibilidade, nota-se que ela é vista como problemas de adaptação do ambiente físico para pessoas com deficiência (sete bibliotecários da biblioteca universitária, bibliotecas escolares, especializadas e pública B1, B4, B5, B11, B10, B12, B13) ou do ambiente online (um bibliotecário de biblioteca universitária e dois da biblioteca digital B4, B15, B16) para acesso a obras tanto no meio físico quanto online (biblioteca universitária, biblioteca especializada e biblioteca pública - B1, B5, B9, B13); oferta de serviços para pessoas com deficiência (visão de uma bibliotecária da biblioteca pública - B14). Para os participantes das universidades e do setor pesquisado na Biblioteca Nacional, a acessibilidade é uma questão que não é de responsabilidade exclusiva da biblioteca, pois há órgãos na instituição que cuidam especificamente da acessibilidade (B1, B4). Já nas outras bibliotecas (escolar, especializada – B10) a questão da acessibilidade é tratada quando há demanda ou é intratável, por grandes dificuldades de adaptação do ambiente físico (B11, B12) e problemas administrativos/burocráticos/financeiros (B1, B11, B12), os últimos também observados na biblioteca pública (B13). A visão de todos eles, exceto de um participante de biblioteca pública (B14), é compatível com a biblioteca adaptada, na distinção de biblioteca adaptada e acessível²⁹ de Paula e Carvalho (2009), abordando problemas do espaço físico, como se nota na falta dos entrevistados B1 e B13: “acessibilidade eu acho que sobretudo inclusão social (...) é um grande desafio nosso [...]. É uma tendência pros sites serem cada vez mais democráticos, digamos assim” (B15).

De maneira semelhante à atenção destinada à acessibilidade nas bibliotecas, a maior parte dos analistas entrevistados tratam da acessibilidade sob demanda (A3, A4, A5, A9, A10, A11, A12), ou deram respostas vagas quando perguntados sobre a acessibilidade, mostrando não conhecer tão bem o conceito (A2, A3, A13). Há também participantes que disseram não realizar pesquisas sobre as qualidades de uso, de maneira geral, do software e

²⁹ “Biblioteca adaptada é aquela que segue as regras do desenho acessível, com rampas, banheiros adaptados, sinaleiras Braille, entre outras. A biblioteca acessível é a que disponibiliza a informação em qualquer suporte e provê acesso a todas as pessoas que dela necessitam, ou seja, segue os princípios do desenho universal” (PAULA; CARVALHO, 2009, p. 72).

não responderam à pergunta (A6, A7, A15). Outros disseram ou não realizar pesquisas de acessibilidade ou incorporá-la ao processo (A14) de desenvolvimento. Apenas três deles (A1, A12 e A16) relataram a incorporação dos aspectos da acessibilidade em projetos de produto ou pesquisas, e outros dois aproximaram o conceito da acessibilidade aos projetos com boa portabilidade (A16) ou design universal (A8):

Pesquisadora: E acessibilidade, tem alguma preocupação lá?

A16: Lá tem de portabilidade. A gente passa um validadorzinho [de portabilidade] tipo do W3C [desenvolvido internamente na empresa que] vê se o código está conforme as regras do W3C. Ou seja, ele já está meio caminho andado pra ser aberto em qualquer navegador. Portanto, já tá a um caminho para a acessibilidade. [...] E tem uma documentação lá pra se preocupar com a acessibilidade, porque que tem que se preocupar, mas também não é nada assim... uma super regra não.

Entre os participantes, apenas dois dizem tratar da acessibilidade por meio do estabelecimento de diretrizes que precisam ser seguidas no processo de desenvolvimento de produtos (A1, A16) – diretrizes fundamentadas nas regras de acessibilidade da W3C (A16) ou embutidas em guias de estilo para criação de interfaces de produtos (A1). A proximidade da portabilidade ao conceito de acessibilidade é compatível com uma visão da acessibilidade análoga à da biblioteca adaptada (PAULA; CARVALHO, 2009), embora tal visão seja criticada por uma das participantes (A12).

A realização de teste de acessibilidade junto aos usuários aconteceu sob demanda apenas por uma entrevistada (A12) e, à exceção de um trabalho de pesquisa de uma das participantes analistas de TI que tratou do design de uma interface inclusiva com princípios do design universal, e de uma bibliotecária que falou da adaptação dos serviços da biblioteca, a acessibilidade é tratada como uma questão técnica – tanto de adaptação física da biblioteca quanto do cumprimento das normas de acessibilidade no ambiente digital pelos analistas de TI. Não é uma questão avaliada (não há relatos de teste de acessibilidade pelos participantes). Os bibliotecários atuantes nas bibliotecas digitais creditam a responsabilidade pela acessibilidade à equipe de TI e nota-se, nas entrevistas dos analistas de TI pesquisados, que a acessibilidade é tratada sob demanda (como um requisito do sistema) também à semelhança de se tratar a acessibilidade sob demanda na biblioteca.

3.4 Atendimento e treinamento de usuários

O atendimento aos usuários nas bibliotecas pesquisadas se dá por canais de contato direto entre bibliotecário e usuários – e-mail, telefone, pessoalmente – sendo outros canais indiretos menos citados. Tal característica confirma a relação de proximidade do usuário em relação ao bibliotecário, e aponta para a informalidade dos contatos: o serviço de

atendimento é realizado principalmente sob demanda do usuário, como aponta B10: "E muitas das vezes o contato vem da necessidade deles (...). Então, seu contato é basicamente esse, tudo informal". Com relação ao atendimento a usuários, percebe-se que os bibliotecários não relatam grandes problemas, ou relatam dificuldades relativas à inclusão digital ou informacional (dúvidas para entender o resultado da consulta à base, numeração das estantes, normas bibliográficas).

Por outro lado, o contato direto dos analistas de TI com os usuários se dá no processo de criação/construção da concepção de produtos e sistemas, bem como na sua operação, por meio do suporte técnico. No entanto, nem todas as empresas, fábricas ou centros participantes da pesquisa contam com equipes de suporte e atendimento aos usuários. Em algumas fábricas pesquisadas (fábricas A, D e G, Centro A), os entrevistados dizem que não há devida atenção para atendimento e suporte ao usuário, em parte porque os produtos devem ter boa usabilidade para dispensar essa ajuda. Em outros casos (fábricas B, C e Centro A) o suporte e ajuda são fornecidos por contratação, dependendo da natureza do produto e do cliente, enquanto em outros (fábricas F e E, Empresas E e F, Banco A, Instituição A, Centro B) há equipes específicas voltadas para atendimento e suporte a usuários.

Ao contrário dos trabalhadores bibliotecários, que atendem o usuário predominantemente por meio de contatos por telefone, e-mail e pessoalmente, nota-se que predomina o atendimento aos usuários de sistemas em operação por mecanismos formais – como aqueles que prevêem o uso de sistemas de controle de chamados, sistemas de *call center* e *help desk*. Nota-se a perspectiva da eficiência do serviço como balizadora das ações de atendimento ao usuário (número de horas para atendimento, controles da efetividade dos chamados, estruturação dos atendimentos em catálogos de serviço), com perspectiva claramente funcionalista. Os espaços para contato são mais associados à solução de defeitos, mas não voltados para a participação do usuário na construção dos sistemas, para a troca de experiências entre os usuários, ou para fazer o usuário conhecer melhor os serviços ou o sistema. Nota-se que não há iniciativas similares às dos bibliotecários de construção de *blog*, *twitter* ou *facebook* para os usuários dos sistemas. O usuário é tratado de maneira individualizada pelo analista de TI, como um gerador de problemas a serem solucionados, e aqui não se vê claras iniciativas para instruí-lo, diferentemente da perspectiva dos bibliotecários.

Com relação à instrução e treinamento de usuários, sete dos 13 entrevistados bibliotecários disseram não realizar treinamento (casos na biblioteca digital, biblioteca

especializada, biblioteca escolar, biblioteca pública e Biblioteca Nacional) pelas dificuldades de treinamento intrínsecas à instituição de natureza pública em que trabalhavam, que apresentam diversidade grande de usuários (B13, B14, B17) ou pouca disponibilidade de recursos para treinamento, ou por o realizarem sob demanda (caso das bibliotecas especializadas). Dois bibliotecários na biblioteca pública vêem a orientação do usuário, realizada sob demanda do usuário/visitante da biblioteca, como uma forma de treinamento.

Os participantes bibliotecários que realizavam treinamento de usuários, quando envolvidos na instrução para uso de portais digitais (como da CAPES, ou intranet de empresas, ou repositórios ou bibliotecas digitais), orientavam o usuário diretamente no uso dos referidos sistemas, por meio de exercício-prática e uso do *datashow (slides)*. Nas bibliotecas universitárias, verificou-se, também, as práticas de visita guiada e atendimentos específicos e individualizados para orientação de elaboração de trabalhos acadêmicos e normas bibliográficas.

Se nem todos os bibliotecários se envolvem com treinamento de usuários, o treinamento não é atividade que está obrigatoriamente incluída no trabalho dos analistas de TI. Se, no caso das bibliotecas públicas, o treinamento dificilmente acontecia pela própria dispersão do público, também os usuários tratados como público (usuários de recursos *web*, por exemplo) não recebem treinamento, como relatam os participantes designers de interação (A8, A9, A10, A11, A12) e um analista de sistemas/engenheiro de software (A1). Neste caso, o cuidado com manuais de usuário e tutoriais se faz presente.

O treinamento pode ser dado sob contratação, no caso das fábricas B e C. Mas, um dos participantes diz que “dificilmente, hoje, o cliente contrata o treinamento de usuários” (A2), talvez por não visto como necessário pelos mesmos motivos apontados para não realização de treinamento pelos bibliotecários B7, B5 e B8 em sistemas de documentos, relacionados ao fato do usuário já conhecer bem os documentos e dispensaria o treinamento e manuais: “tem contrato que o usuário não quer nem manual, que o cliente não quer nem manual e acontece, que às vezes é uma coisa pra eles muito do negócio, muito óvia”(A3). Porém, uma das participantes – que tem perfil de analista de negócio e acompanha o uso dos sistemas e chamados de suporte (A15) – relata que o treinamento é um dos fatores determinantes de sucesso ou falha na implantação e operação dos sistemas, assim como também Schwiderski (2011), na revisão de literatura.

Já no caso das fábricas D, E e F, o treinamento faz parte do processo padrão dos sistemas e consiste da apresentação do sistema e realização de exercício-prática,

essencialmente, à semelhança da abordagem de treinamento e atendimento de usuários de bibliotecas. Quando o número de usuários é grande, nem todos são treinados (A4, A15). Nestes casos, há treinamento de alguns deles, os quais recebem a incumbência de disseminar o que aprenderam do sistema a outros usuários (são multiplicadores ou usuários-chave). O treinamento pode ser destinado a equipes que não estão associadas ao desenvolvimento do sistema, como equipes de RH (A14, A15). Tal aspecto contrasta com a realização do treinamento dos recursos da biblioteca estar sempre a cargo dos bibliotecários, e não fragmentada no ambiente de trabalho para outras funções ou setores.

3.5 Uso dos sistemas

Entre os profissionais analistas de TI, a análise do uso das funcionalidades dos sistemas não é realizada pela maioria dos entrevistados, sendo que deles, apenas cinco (A1, A2, A8, A12, A15) disseram ter sido realizada algum tipo de análise de uso de sistemas, dos quais dois fazem análise de *logs* para subsidiar mudanças nos mesmos (A1, A8), uma disse ter feito análise de uso em uma pesquisa no trabalho (A12), enquanto outra disse ter realizado uma pesquisa informal de quem efetivamente usa um determinado sistema com objetivos administrativos (A15):

Sabe qual a pesquisa que eu tô fazendo atualmente? Eu tô brincando, né, não tem nada haver com pesquisa, mas que a gente tá vendo é que [a empresa E] contratou x licenças do SAP. Aí, a gente tá fazendo uma auditoria agora e viu que tem muito mais. [...] O gerente [...] falou assim “usa o seu conhecimento dos usuários, que você conhece todo mundo”, que tem muitos anos que eu trabalho lá, “você conhece todas as filiais, [...], procura esses coordenadores, e vê o que pode melhorar” (A15).

Como se nota na fala acima, três pesquisados (A4, A15, A16) têm uma visão normativa do uso dos sistemas pelos usuários. Há crítica acerca da quantidade de funcionalidades que não são utilizadas (como os livros que não consultados nas bibliotecas), o que é visto como problema cuja causa percebida está associada aos processos de construção de software (A4, A16): “eu acho que as pessoas estão começando a ficar mais atentas com isso [ao que é usado nos sistemas] [para] evitar desperdício” (A16).

Nota-se, ainda, na fala de A15 acima, falta de sistematização da condução de pesquisa de uso, a exemplo do que ocorre com a falta de apoio teórico para o trabalho dos bibliotecários. Os outros dois analistas que apontaram realizar pesquisa de uso, o fazem para subsidiar mudanças nos produtos ou serviços (fábricas A e G). Há referência às análises métricas por dois pesquisados (A1 e A8) – como análise de número de visualizações de páginas de um *website* – que também dizem respeito às técnicas de pesquisa similares à pesquisa bibliométrica em bibliotecas, alinhada ao paradigma

tradicional de estudos de usuários: as funcionalidades e recursos utilizados no sistema são medidos quantitativamente e analisados para a tomada de decisões sobre mudanças nas funcionalidades e/ou interface dos aplicativos. Uma entrevistada, embora não faça análise métrica de dados indiretos coletados de usuários (estatísticas de *downloads*) ou diretos (mensagens de *feedback* sobre o produto com avaliações negativas ou positivas reportadas pelos usuários) diz que elas podem subsidiar mudanças, mas que nem sempre o acesso aos dados é possível, pois está nas mãos do cliente (A11).

4 CONSIDERAÇÕES FINAIS

Os processos de trabalho relativos aos usuários – atendimento e treinamento; pesquisa de necessidades e requisitos para sistemas; estudos de uso; abordagem da usabilidade e da acessibilidade – realizados por analistas de TI e bibliotecários não apontam realizações de pesquisas de usuários realizadas de maneira sistemática, à exceção de processos de construção de sistemas e aplicativos realizados por designers de interação e especialistas em usabilidade, que aplicam técnicas de DCU.

Os levantamentos de requisitos de sistemas realizados junto a usuários, embora empreguem técnicas qualitativas – como entrevistas e estudo de documentação dos sistemas existentes ou do domínio com alguma sustentação na engenharia de requisitos – são fundamentadas essencialmente, na vivência prática dos entrevistados (B5, B8, B7, A3).

Embora não seja a regra geral, usabilidade e acessibilidade são temas emergentes ainda não instituídos na prática de todos os profissionais entrevistados. A usabilidade de sistemas computadorizados é presente na prática e na fala dos analistas de TI com atribuições de design de interação e usabilidade, mas não é instituída em todos os processos de software entre os entrevistados com outros perfis (analistas de sistemas, de negócio, de suporte), nem na prática de trabalho dos bibliotecários. A acessibilidade – em sua compreensão como acesso a recursos dos sistemas físicos ou digitais – é mais presente na fala e na prática dos bibliotecários, mas tanto para estes profissionais quanto para os analistas de TI, é uma questão tratada essencialmente sob demanda. Já a incorporação da temática da arquitetura da informação é presente na fala dos bibliotecários e não dos analistas de TI.

De maneira geral, há, também, pouco conhecimento do uso dos sistemas ou das bibliotecas pelos seus usuários, tanto por analistas quanto bibliotecários, como forma de subsidiar mudanças nos sistemas, predominando sua importância para fins administrativos. Bibliotecários e analistas de TI empregam poucos métodos sistemáticos para conhecer seus usuários no cotidiano profissional.

REFERÊNCIAS

- ATTARHA, Mina; MODIRI, Nasser. Focusing on the Importance and the Role of Requirement Engineering. In: INTERNATIONAL CONFERENCE ON INTERACTION SCIENCES (ICIS), 4., 2011, Busan. **Proceedings...** Busan: IEEE, 2011. p. 181-184.
- BALDUS, M. F. O; MACIEL, C.; SOUZA, P.C. Um diagnóstico do uso da modelagem da interação em métodos ágeis no mercado de software. In: SIMPÓSIO DE FATORES HUMANOS EM SISTEMAS COMPUTACIONAIS (IHC 2012), 10., 2012, Cuiabá. **Anais...** Porto Alegre: Sociedade Brasileira de Computação – SBC, 2012. p. 17-20.
- BARBOSA, S. D. J.; SILVA, B. S. **Interação humano-computador.** Rio de Janeiro: Elsevier, 2010.
- CASE, Donald O. **Looking for Information:** A Survey of Research on Information Seeking, Needs and Behavior. 2 ed. Elsevier Ltd., 2007.
- CHOO, C. W. **A organização do conhecimento:** como as organizações usam a informação para criar significado, construir conhecimento e tomar decisões. São Paulo: Editora Senac, 2003. p. 63-120.
- FIGUEIREDO, N. M. Estudos de usuários. In: _____. **Estudos de uso e usuários da informação.** Brasília: IBICT, 1994. p. 7-19; 65-85.
- FALLMAN, Daniel. The new good: exploring the potential of philosophy of technology to contribute to human-computer interaction. In: CONFERENCE ON HUMAN FACTORS IN COMPUTING SYSTEMS (CHI'11). **Proceedings...** Vancouver, BC, Canada: ACM, 2011. p. 1055-1060.
- GONZÁLEZ TERUEL, Aurora. **Los estudios de necesidades y usos de la información:** fundamentos y perspectivas actuales. Ediciones Trea S. L., 2005.
- NASCIMENTO, Luciano Prado Reis. **O usuário e o desenvolvimento de sistemas.** Florianópolis: Visual Books, 2003.
- NIELSEN, J. **Usability Engineering.** Cambridge, MA: Academic Press, 1993.
- MORVILLE, Peter; ROSENFELD, Louis. **Information Architecture for the World Wide Web.** 3 ed. O'Reilly Media Inc, 2006.
- PAULA, S. N.; CARVALHO, J. O. F. Acessibilidade à informação: proposta de uma disciplina para cursos de graduação na área de biblioteconomia. **Ciência da Informação,** Brasília, v. 38, n. 3, p. 64-79, set./dez. 2009 .
- PREECE, J.; ROGERES, Y.; SHARP, H. **Design de interação:** além da interação humano-computador. Porto Alegre: Bookman, 2005.
- ROCHA, Eliane Cristina de Freitas. Usuário da informação, um velho (?) (des)conhecido: usuários da informação em diferentes profissões da informação. 2013. 364 f. Tese (Doutorado em Ciência da Informação) – Universidade Federal de Minas Gerais, Belo Horizonte, 2013.
- SILVA, T. S. **A framework for integrating interaction design and agile methods.** 2012. 110 f. Tese (Doutorado em Ciência da Computação) – Faculdade de Informática, PUCRS, Porto Alegre, 2012.

SCHWIDERSKI, Antônio Carlos. **Desenvolvimento de sistemas de informação do Hospital Colônia Adauto Botelho:** fatores que interferem nos fazeres de levantamento de requisitos. 2011. 94 f. Dissertação (mestrado profissional) – Universidade Estadual de Londrina, Programa de Pós-Graduação em Gestão da Informação, Londrina, 2011.

SOMMERVILLE, Ian. **Engenharia de software.** São Paulo: Addison Wesley, 2003.

VIANNA Maurício et al. **Design Thinking - Inovação em Negócios.** MJV Press, 2012.

WEB SEMÂNTICA: INTRODUÇÃO A RECUPERAÇÃO DE DADOS USANDO SPARQL

SEMANTIC WEB: INTRODUCTION TO DATA RECOVERY USING SPARQL

José Eduardo Santarém Segundo

Resumo: A Web Semântica faz parte do cotidiano de nossas vidas, a publicação de grandes volume de dados através do *Linked Data* é um dos principais produtos materializados como resultado da proposta inicial de Tim Berners-Lee, publicada na *Scientific American*, em 2001. Diante do exposto é necessário concentrar esforços que vão além dos modelos para estruturar e efetivar a publicação de dados em formato semântico, torna-se necessário estudar atentamente modelos de recuperação que tornem usuais estes dados. O objetivo deste trabalho é apresentar o protocolo *SPARQL* e suas possibilidades, como um dos principais aliados no processo de recuperação da informação em ambientes semânticos. A metodologia utilizada nesta pesquisa conta com uma pesquisa bibliográfica e documental seguida de um procedimento exploratório em torno do uso do protocolo *SPARQL*. Como resultado verificou-se a disponibilidade de ferramentas que tornam possível o efetivo uso do protocolo *SPARQL* para recuperação da informação em ambientes semânticos.

Palavras-chave: Web Semântica. Recuperação da Informação. *SPARQL*. *Linked Data*.

Abstract: The Semantic Web is part of our everyday lives, the publishing of large volumes of data through Linked Data is a leading product materialized as a result of the initial proposal by Tim Berners-Lee, published in Scientific American in 2001. Given the above it is necessary to focus efforts that go beyond the models to structure and conduct the publication of data in semantic form, it is necessary to carefully study recovery models that make these data useful. The objective of this paper is to present the SPARQL protocol and its possibilities, as a key ally in the process of information retrieval in semantic environments. The methodology used in this research has a bibliographical and documentary research followed by an exploratory procedure around the use of the SPARQL protocol. As a result there was the availability of tools that make possible the effective use of the SPARQL protocol for information retrieval in semantic environments.

Keywords: Semantic Web. Information Retrieval. SPARQL. Linked Data.

1 INTRODUÇÃO

Além das “nuvens”: expandindo as fronteiras da Ciência da Informação, tema do Enancib 2014, retrata com muita clareza a situação atual das pesquisas e principalmente das perspectivas de pesquisa que envolvem a Ciência da Informação atualmente.

Enquanto grande parte dos chamados objetos de estudo da Ciência da Informação encontram-se em processo de mudança para o contexto digital, outros já estão a algum tempo encapsulados nas nuvens. Dentre estes estudos e pesquisas encontram-se os conceitos e tecnologias que envolvem a Web Semântica, proposta descrita pela primeira vez em 2001, quando Tim Berners-Lee, James Hendler e Ora Lassila, publicaram na revista *Scientific American* o texto “*The Semantic Web*”.

Desde a proposta inicial de 2001, a Web Semântica vem ganhando força e agregando novas tecnologias, funcionalidades e evoluindo para tornar real o processo de construção de ambientes semânticos. Tecnologias como RDF (*Resource Description Framework*), XML (*eXtensible Markup Language*), OWL (*Web Ontology Language*) e todos os conceitos que as envolvem, ganham novas versões, descritos com clareza no W3C (*World Wide Web Consortium*), e tornam possível a materialização do conceito da Web Semântica. As tecnologias citadas estão diretamente relacionadas ao processo de construção da informação e armazenamento das mesmas, constituindo assim ambientes que possam ter conjunto de dados ligados semanticamente.

O conceito de Web Semântica e de ligação entre dados deu origem ao *Linked Data*, um conjunto de dados em formato aberto, que se apresenta como um grande ambiente de informação através de dados ligados semanticamente entre si em vários lugares ao redor do mundo, caracterizando-se por utilizar a internet (a nuvem) como seu hospedeiro. É possível afirmar que o *Linked Data* é uma materialização dos conceitos da Web Semântica.

Não é necessário muito esforço para ter acesso a toda essa imensidão de tecnologias que tem tornado real a ideia da construção de uma Web Semântica. Apesar do mapa do *Linked Data* não ser atualizado desde 2011, o conjunto de dados disponíveis continua crescendo em escala logarítmica e é possível identificar muita informação pronta para acesso em vários *datasets* (nome dados as bases de dados no *Linked Data*) ao redor do mundo.

Essa grande disponibilidade sugere que utilizemos formatos e estruturas adequadas de pesquisa para encontrar e principalmente fazer uso de todo esse conjunto de dados.

Desta forma o objetivo deste trabalho é apresentar o protocolo *SPARQL* e suas possibilidades como um dos principais aliados no processo de recuperação da informação em ambientes semânticos. A metodologia utilizada nesta pesquisa conta com uma pesquisa bibliográfica e documental seguida de um procedimento exploratório, sendo que a parte da pesquisa que envolveu a metodologia de pesquisa bibliográfica e pesquisa documental foram realizadas em literatura da área de Ciência da Informação e documentação técnica disponível no site do *W3C*. A pesquisa de caráter exploratório foi realizada através de aplicações sucessivas de consultas a conjuntos de dados diversos em formato de arquivos *RDF* e também em ambientes de dados disponíveis no ambiente *Linked Data*, sempre utilizando-se o protocolo Sparql para recuperação de dados.

2 WEB SEMÂNTICA: ESTRUTURA, ARMAZENAMENTO E PUBLICAÇÃO DE DADOS

Após o ano 2001 quando Tim Berners-Lee, James Hendler e Ora Lassila publicaram na revista *Scientific American* o texto “*The Semantic Web*” algumas áreas de pesquisa, em especial dentro da Ciência da Informação e da Ciência da Computação, têm direcionado esforços para colocar em prática os conceitos abordados no texto citado.

Nos últimos anos vários elementos foram surgindo e ampliando o contexto da ideia original de Berners-Lee. O W3C iniciou um processo de publicar, efetivar e disseminar um conjunto de tecnologias que foram se agregando em busca da Web Semântica. Vários projetos ao redor do mundo também evoluíram de forma a constituir ambientes semânticos, tanto do ponto de vista de estrutura informacional quanto da possibilidade de recuperação semântica da informação.

O projeto da Web Semântica tem como ponto fundamental a criação de uma nova estrutura de armazenamento de dados. O ponto principal está na separação da apresentação do conteúdo e do conteúdo da estrutura, tratando as unidades atômicas de uma informação como componentes independentes.

O desafio da Web Semântica vem sendo, a cada dia, prover uma linguagem capaz de expressar ao mesmo tempo dados e regras, de forma a possibilitar a dedução de novos dados e regras a partir de qualquer sistema de representação de conhecimento a ser importado ou exportado na Web (SANTAREM SEGUNDO, 2010).

Diferente da chamada Web de Documentos (a primeira Web), tem-se promovido a construção da Web de Dados, uma iniciativa do W3C com grande apelo em todo o mundo. A Web de Dados propõe uma estrutura que sugere a ligação entre dados publicados em bases de dados abertas disponíveis livremente na Internet. Esses dados publicados estão associados a vocabulários padronizados, ontologias e aplicação de inferências (W3C, 2010).

Para disponibilizar dados numa estrutura semântica é necessário usar partes (ou todo) do modelo descrito por Berners-Lee em 2001, no chamado bolo de noiva (*layercake*), estrutura de camadas que apresenta a Web Semântica. Destaca-se neste quesito a linguagem RDF, também indicada para representação de dados abertos.

Um dos principais objetivos da linguagem RDF é justamente criar uma rede de informações a partir de dados distribuídos. Essa característica de constituição de redes se dá pela relação existente entre o modelo de construção de triplas do RDF e o modelo matemático conhecido como grafo.

Um grafo é um modelo matemático muito poderoso que pode ser aplicado na resolução de um conjunto de problemas. É composto por um conjunto de vértices e arestas/arcos (SANTAREM SEGUNDO, 2010).

De acordo com o W3C, o RDF é uma linguagem de uso geral para representar informações na Web. O RDF tem como princípio fornecer interoperabilidade aos dados, de forma que possa contribuir com a recuperação de informações de recursos na Web.

Segundo Lassila,

RDF é uma aplicação da linguagem XML que se propõe ser uma base para o processamento de metadados na Web. Sua padronização estabelece um modelo de dados e sintaxe para codificar, representar e transmitir metadados, com o objetivo de torná-los processáveis por máquina, promovendo a integração dos sistemas de informação disponíveis na Web (LASSILA, 1999).

O modelo RDF é constituído de três objetos básicos: recursos, propriedades e valores. Um recurso é uma informação (página web, livro, cd, pessoa, lugar, documento disponível em um repositório ou biblioteca digital) que pode ser identificada por uma URI (Universal Resource Identifier). Propriedades são as informações que representam as características do recurso, ou seja, são os atributos que permitem distinguir um recurso de outro ou que descrevem o relacionamento entre recursos. Os valores são os dados que representam a informação a ser descrita.

A declaração é a constituição da informação completa, que comprehende um recurso com suas propriedades e valores para as propriedades. Uma URI pode ser um local ou página na WEB como uma URL (*Unified Resource Locator*) ou ainda outro tipo de identificador único.

Além de representar graficamente uma informação através de grafos, o modelo RDF pode ser representado através da sintaxe XML para uso como artefato computacional.

O modelo básico de representação dos dados é o RDF, porém para que se possa publicar dados e principalmente torná-los disponíveis para serem recuperados com mais eficiência e eficácia, é necessário que o esquema lógico estrutural dos dados esteja sob uma ontologia, ou minimamente utilizando-se de vocabulários de representação de conhecimento padronizados e reconhecidos.

Utilizar ontologias e suas relações é uma das maneiras de se construir uma relação entre termos dentro de um domínio, favorecendo a possibilidade de contextualizar os dados, tornando mais eficiente e facilitando o processo de interpretação dos dados pelas ferramentas de recuperação da informação.

Para Guarino (1998, p.7), ontologia é “uma maneira de se conceituar de forma explícita e formal os conceitos e restrições relacionados a um domínio de interesse”. Numa visão mais tecnológica, o termo refere-se a um artefato computacional, que em uma visão simplista, pode ser descrito como uma hierarquia de conceitos relacionados entre si através de uma classificação de parentesco (hipernímia e hipônimo), também chamada de taxonomia.

A definição de Jacob aproxima-se muito do conceito de ontologia que mais se aplica à Ciência da Informação quando no contexto da recuperação semântica de informações:

Ontologias são categorias de coisas que existem ou podem existir em um determinado domínio particular, produzindo um catálogo onde existem as relações entre os tipos e até os subtipos do domínio, provendo um entendimento comum e compartilhado do conhecimento de um domínio que pode ser comunicado entre pessoas e programas de aplicação (JACOB, 2003, p.19).

Construir ontologias nem sempre é o melhor ou mais rápido caminho para disponibilizar dados em formato semântico. Em grande parte das vezes utilizar-se de ontologias prontas e universalmente conhecidas e reconhecidas por uma determinada comunidade, acelera o processo de publicação de dados, mas principalmente favorece o processo de recuperação da informação em ambientes semânticos.

Não utilizar esquemas lógicos como as ontologias e os principais vocabulários para descrever dados a serem publicados certamente é o pior caminho para publicação de dados em formato semântico, principalmente do ponto de vista da recuperação da informação.

Apesar da criação de etiquetas para representação das propriedades do RDF ser livre, há um grupo de vocabulários que são utilizados em larga escala nas principais ontologias conhecidas e também em grande parte dos exemplos de publicação de *datasets* em formato semântico disponíveis na Internet.

Dentre os vocabulários mais utilizados encontramos:

- *Friend-of-a-Friend* (FOAF) , o vocabulário para descrever pessoas.
- *Dublin Core* (DC) define os atributos gerais de metadados, baseado no padrão de metadados *Dublin Core*.
- Comunidades Online Semanticamente Interligados (SIOC) , vocabulário para representar as comunidades online.
- Descrição de um projeto (DOAP) , vocabulário para descrever projetos.
- Simples Sistema de Organização do Conhecimento (SKOS) , vocabulário para representar taxonomias, vocabulários controlados e conhecimento pouco estruturado.

- *Music Ontology*, prevê condições para descrever artistas, álbuns e faixas.
- *Creative Commons (CC)* , o vocabulário para descrever termos de licença.

Além dos listados, há ainda uma série de outros vocabulários disponíveis, assim como vários são os modelos de esquemas e ontologias disponíveis para serem utilizados e principalmente consumidos livremente. Uma das ideias materializada, que agrupa e congrega esses esquemas, vocabulários e principalmente dados, é chamada de *Linked Data*.

3 LINKED DATA

Pensando em um modelo associativo de publicação de dados estruturados na Web, foi constituído o *Linked Data* (dados ligados). Organizado por Tim Berners-Lee tem como característica principal o estabelecimento de links entre dados de fontes distribuídas.

Segundo Heath e Bizer (2011), “o Linked Data é um conjunto de melhores práticas para publicação e conexão de dados estruturados na Web, permitindo estabelecer links entre itens de diferentes fontes de dados para formar um único espaço de dados global”.

Para Berners-Lee (2006),

a Web Semântica não trata apenas de depósito de dados na web. Trata-se de fazer ligações, de modo que uma pessoa ou máquina possa explorar esse conjunto de dados. Com *Linked Data*, quando você tem um pouco de dados, você pode encontrar outros que estão relacionados.

A construção do *Linked Data* está baseada em quatro princípios publicados por Berners-Lee (2006):

- Usar URIs como nomes para os itens.
- Usar URIs HTTP para que as pessoas possam consultar esses nomes.
- Quando alguém consulta uma URI, prover informação RDF útil.
- Incluir sentenças RDF com links para outras URIs, a fim de permitir que itens relacionados possam ser descobertos.

O projeto tem crescido muito nos últimos anos, em 2007 o *Linked Data* era constituído de aproximadamente um bilhão de declarações RDF, interligados por 120.000 links RDF. Em 2011 eram 52 bilhões de declarações RDF. Hoje não há números expressos sobre o conjunto de ligações que não para de crescer.

Com certeza a base de maior tamanho e também maior destaque no *Linked Data* é a *DBpedia*.

A *DBpedia* é uma interface orientada a dados, constituída a partir de um esforço da comunidade para extrair informação estruturada da *Wikipédia* e tornar esta informação disponível na web. Atualmente está sendo constituída a *DBpedia Live* que é uma estrutura

funcional que tem como objetivo transferir diretamente os dados do *Wikipédia* para o padrão semântico do *DBpedia*.

Ressalta-se a grandeza e proporção que o *Linked Data* vem tomando, a base de conhecimento *DBpedia* atualmente descreve mais de 4 milhão de itens, das quais 3,22 milhões são classificados de forma consistentes por ontologias, incluindo 832 mil pessoas, 63 mil lugares, além de músicas, filmes, games, espécies, doenças entre outras informações. Os dados da *DBpedia* apresentam estes itens em até 119 idiomas diferentes. A base de conhecimento *DBpedia*, em todos idiomas, consiste em aproximadamente 2,5 bilhões de declarações RDF, dos quais 470 milhões foram extraídos da edição Inglês da *Wikipedia*. (*DBPEDIA*, 2014).

Outra base de destaque, principalmente porque praticamente 95% de seus dados tem precisão manual é a base Yago. Essa base relaciona 2 milhões de entidades (pessoas, organizações, cidades, etc.). São aproximadamente 20 milhões de fatos sobre essas entidades.

Destaca-se ainda a integração de dados geográficos, tais como nomes de lugares em vários idiomas, altitude, população e outras fontes publicados pelo banco de dados *Geonames* que está disponível para download gratuitamente sob uma licença *Creative Commons*. O dataset do *Geonames* contém mais de 10 milhões de nomes geográficos e é composto por 8 milhões de recursos exclusivos. Todos os recursos são classificados em uma das nove classes de recurso e ainda subcategorizados em um dos 645 códigos de recurso. Os dados são acessíveis gratuitamente através de um número de webservices e uma exportação de banco de dados diariamente. O webservice *Geonames.org* tem atendido aproximadamente 30 milhões de requisições diárias. (*Geonames*, 2013)

Outra característica a ser abordada nos dados publicados no *Linked Data* é a disseminação de instrumentos de controle de vocabulário e controle de autoridade. O *AGROVOC* é um tesouro desenvolvido a partir de 1980 que atualmente é referência para temas ligados à agricultura, piscicultura, silvicultura entre outros assuntos relacionados ao meio ambiente. O *AGROVOC* é utilizado mundialmente por pesquisadores, bibliotecários, gestores de informação e outros profissionais para indexar, recuperar e organizar dados em sistemas de informações agrícolas. Utilizando-se da terminologia *SKOS*, modelo apropriado para publicação de tesouros na web, o *AGROVOC* pode ser consultado assim como interligado com outros recursos para provimento de dados em formato semântico. (*AGROVOC*, 2014).

Há ainda uma grande diversidade de bases de dados, que vão desde informações sobre Ciências da Vida (*GeneID*, *PubMed*, *Geo Species*, *Gene Ontology*, etc), Dados Geográficos (Aeroportos, *Earth*, *Linked GeoData*, etc), Dados de Uso Geral (*Slideshare*, *Semantic Tweet*, *Delicious*, *Flickr*, etc), Mídia (*BBC*, *Music Brainz*, *New York Times*, *Last.FM*, etc), Publicações (*IEEE*, *ePrints*, *CiteSEER*, *theses.fr*, etc) e Dados Governamentais (*Patentsdata.gov.uk*, *researchdata.gov.uk*, *transportdata.gov.uk*, etc).

Todos estes chamados *datasets* estão listados no mapa do *Linked Data* publicado em 2011. Atualmente existe uma infinidade de outros *datasets*, que não faziam parte do mapa do *Linked Data* em 2011, com um volume imenso de declarações RDF sobre os mais variados tipos de dados.

4 PROTOCOLO SPARQL E A RECUPERAÇÃO DA INFORMAÇÃO

Os capítulos apresentados até agora definem bem a estrutura informacional que se construiu a partir das tecnologias da Web Semântica. Grande parte do conteúdo de pesquisa disponível em bases nacionais e internacionais cita os padrões e tecnologias abordados aqui e indicam o caminho das pedras no processo de construção de bases de dados ligados no modelo do *Linked Data*.

Com essa grande oferta de dados disponíveis é necessário um segundo passo, o desfrute desses *datasets*. É necessário transformar todo esse conjunto de informações em conhecimento útil e aplicável, de forma a mudar positivamente a vida das pessoas. Essa era a proposta inicial de Berners-Lee com a Web Semântica.

A busca por pesquisas já realizadas sobre recuperação da informação em ambientes semânticos nos leva a um conjunto de textos publicados, sendo que a maioria deles apresentam contextualizações teóricas ou ainda modelos de recuperação da informação que podem ser aplicados a dados ligados. Poucos desses textos apresentam informações precisas sobre como proceder ao lidar com dados como os disponíveis no *Linked Data*.

Nessa pesquisa, temos como foco apresentar e discutir o protocolo SPARQL (*Simple Protocol and RDF Query Language*) com um caráter mais pedagógico do ponto de vista de estímulo aos pesquisadores que desejam utilizá-lo.

Segundo Tim Berners-Lee, tentar utilizar o potencial da Web Semântica sem SPARQL é o mesmo que tentar utilizar um banco de dados relacional sem usar a linguagem *SQL*.

O SPARQL é um conjunto de especificações que fornecem linguagens e protocolos para consultar e manipular o conteúdo publicado em RDF na Web. O padrão compreende as seguintes especificações: uma linguagem de consulta para RDF; uma especificação que

define uma extensão do SPARQL *Query Language* para executar consultas distribuídas em diferentes terminais SPARQL; uma especificação que define a semântica de consultas SPARQL sob regimes de vinculação, como RDF Schema, OWL, ou RIF; um protocolo que define os meios para a transmissão de consultas SPARQL arbitrárias e solicitações de atualização para um serviço de SPARQL; uma especificação que define um método busca e descoberta e um vocabulário para descrever serviços SPARQL e um conjunto de testes, para avaliação da especificação SPARQL 1.1 (SPARQL, 2013).

O protocolo SPARQL 1.1 é uma versão com muitas alterações e substanciais evoluções em relação a primeira versão SPARQL 1.0, publicada pelo W3C como linguagem de consulta em Janeiro de 2008. Atualmente chamado de protocolo, a versão 1.1 é uma recomendação W3C desde março de 2013.

4.1 SPARQL: Como funciona

Ao estabelecer uma metodologia de pesquisa para que se possa utilizar de forma consistente o SPARQL, é importante dominar exatamente como acontecem as declarações RDF, que podem ser apresentadas em forma de gráficos, tabelas, arquivos XML, arquivos em formato *Turtle*, N3, entre outros tantos formatos conhecidos para RDF.

CÓDIGO 1 - Tripas RDF de uma agenda em formato *Turtle*

```
@prefix ag: <http://meusite.com.br/namespace/agenda#> .
ag:maria    ag:fone          "(11) 111-1111" .
ag:maria    ag:e-mail        "maria@meuemail.com.br" .
ag:jose     ag:fone          "(22) 222-2222" .
ag:jose     ag:e-mail        "jose@meuemail.com.br" .
ag:antonio  ag:fone          "(33) 333-3333" .
ag:antonio  ag:e-mail        "antonio@meuemail.com.br" .
ag:antonio  ag:e-mail        "toninho@outroemail.com.br" .
```

Fonte: próprio autor

No CÓDIGO 1, são apresentadas informações de uma agenda representada pelo formato *Turtle* do RDF. Considera-se neste caso o uso de um vocabulário estabelecido em <http://meusite.com.br/namespace/agenda#>. Observa-se que neste caso poderia ter sido utilizado o vocabulário FOAF ou algum outro vocabulário reconhecido internacionalmente, porém preferiu-se para o exemplo um vocabulário próprio, que estaria disponível no link apresentado. Considera-se também que as pessoas na agenda são recursos disponíveis no mesmo endereço. A TABELA 1 apresenta o mesmo conteúdo do CÓDIGO 1, porém em um formato diferente.

TABELA 1 - Apresentação em forma de tabela dos dados do CÓDIGO 1

Recurso	Propriedade	Valor
maria	fone	(11) 111-1111
maria	e-mail	maria@meuemail.com.br
jose	fone	(22) 222-2222
jose	e-mail	jose@meuemail.com.br
antonio	fone	(33) 333-3333
antonio	e-mail	antonio@meuemail.com.br
antonio	e-mail	toninho@outroemail.com.br

Fonte: próprio autor

Apesar do Protocolo SPARQL apresentar um grande conjunto de funcionalidades, entender como funciona o processo de consulta é talvez a parte mais importante para dominar o funcionamento do mesmo.

Uma consulta SPARQL é composta, em ordem, por:

- Declarações de prefixos, para abreviar URIs
- Definição do conjunto de dados, informando quais grafo(s) RDF estão sendo consultados
- A cláusula de resultado, identificando que informação deve ser retornada a partir da consulta
- O padrão de consulta, especificando o que consultar dentro do conjunto de dados
- Modificadores de consulta, limites, ordenação, e outros que podem modificar o resultado final

CÓDIGO 2 - Estrutura padrão de uma consulta SPARQL

```

# declarações de prefixo
PREFIX foo: <http://example.com/resources/>
...
# definição de conjunto de dados
FROM ...
# cláusula de resultado
SELECT ...
# padrão de consulta
WHERE {
...
}
# modificadores de consulta
ORDER BY .

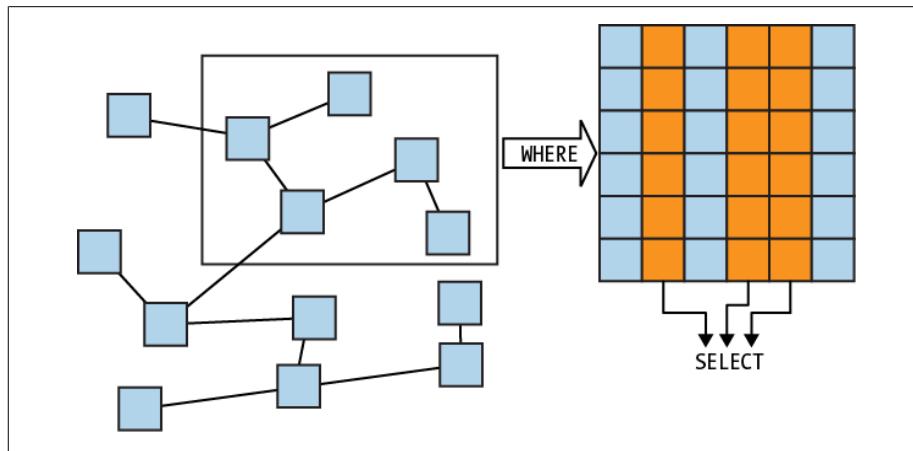
```

Fonte: SPARQL (2013).

No CÓDIGO 2 a cláusula “FROM” indica o conjunto de dados a ser pesquisado, que pode ser representado por um conjunto de arquivos ou de datasets publicados na internet através do *Linked Data*. A cláusula “ORDER BY”, como na linguagem SQL, serve como modificador de apresentação resultado, podendo alterar a classificação dos dados resultantes da consulta.

Ao observar o CÓDIGO 2 o motor principal de uma consulta SPARQL é realizado através da estrutura fundamentada pelas cláusulas “SELECT” e “WHERE”. Conforme pode ser observado na FIGURA 1, a cláusula “SELECT” define quais dados serão apresentados no resultado final da consulta, já a cláusula “WHERE” indica os dados que serão retirados do grafo RDF e trabalhados de forma a ser construído o conjunto de dados de resposta a consulta SPARQL.

FIGURA 1 – Cláusulas WHERE e SELECT em uma consulta SPARQL



Fonte: DUCHARME (2011, p. 4)

A submissão de uma consulta SPARQL a um conjunto de dados RDF apresenta resultados em forma de tabelas em grande parte das vezes.

Os exemplos a seguir foram aplicados sobre o conjunto de dados nomeados aqui como CÓDIGO 1 (nossa conjunto de dados).

O CÓDIGO 3 trata de uma simples consulta, na busca por identificar o e-mail do “José”.

CÓDIGO 3 - Consulta SPARQL

```
PREFIX ag: <http://meusite.com.br/namespace/agenda#>
SELECT ?email_do_ze
WHERE
{ ag:jose      ag:e-mail      ?email_do_ze . }
```

Fonte: próprio autor

Em uma consulta, toda vez que for utilizado o símbolo de interrogação antes de uma palavra ela deve ser entendida como uma variável, ou seja, valor a ser recuperado pela consulta. Neste caso após a cláusula “SELECT” encontramos a variável “?email_do_ze”, que deverá ser o único valor de retorno apresentado. A cláusula “WHERE” mostra três elementos, entre eles a variável “?email_do_ze” que é o valor a ser recuperado no grafo RDF.

É importante esclarecer o funcionamento da cláusula “WHERE”, visto que ela é a responsável por selecionar o conjunto de dados a ser buscado em um *dataset* ou em um arquivo de declarações RDF.

A cláusula “WHERE” é sempre construída por grupos de 3 elementos, que representam as informações de uma declaração RDF. Em publicações da área sobre SPARQL, encontram-se muitos exemplos da cláusula “WHERE ?s ?p ?o”, que representa sujeito (recurso), predicado ou propriedade e objeto (valor).

No CÓDIGO 3, a cláusula “WHERE” apresenta os 3 elementos (recurso, propriedade e valor), porém 2 deles são conhecidos dentro do conjunto de dados a ser buscado, ou seja o único elemento apresentado como variável é o “?email_do_ze”. No caso do CÓDIGO 3 entende-se que a consulta será submetida ao conjunto de dados do CÓDIGO 1, sendo que os dados a serem retornados ou recortados deverão obedecer a cláusula “WHERE” da seguinte forma: “retorne toda declaração RDF encontrada, que tenha como recurso - ag:jose -, tenha como propriedade - ag:e-mail - e tenha qualquer valor (por isso a variável) no elemento que representa o valor”.

Nesse caso a consulta teria como resposta o resultado apresentado na FIGURA 2, uma linha apenas com a informação do e-mail do José.

FIGURA 2 – Resultado da consulta SPARQL

```
MBS--2014_Enancib$ arq --data exemplo1.ttl --query exemplo2.rq
-----
| email_do_ze
=====
| "jose@meuemail.com.br"
-----
```

Fonte: próprio autor

Ainda sobre a cláusula “WHERE” e sobre o CÓDIGO 3, alterar o primeiro elemento de “ag:jose” para uma variável (que poderia ser representada por qualquer palavra antecedida por ?) mudaria o contexto do recorte no conjunto total de dados. Neste caso o conjunto de dados a ser recuperado deveria ter o seguinte contexto: “retorne toda declaração RDF encontrada, que tenha como propriedade - ag:e-mail -, independente do recurso ou do valor apresentado”. Neste caso a consulta buscaria por um conjunto de dados mais abrangente, porque deixa de ter um valor definido para o recurso.

Esta mudança na cláusula “WHERE” representa um possível conjunto maior de dados como resultado, dessa forma a informação do recurso também pode ser representada na cláusula “SELECT”, melhorando a informação de apresentação de resultado.

O CÓDIGO 4 e a FIGURA 3 apresentam essa nova situação, descrita nos parágrafos anteriores.

CÓDIGO 4 – Consulta SPARQL

```
PREFIX ag: <http://meusite.com.br/namespace/agenda#>
SELECT ?pessoa ?email
WHERE
{ ?pessoa     ag:e-mail      ?email . }
```

Fonte: próprio autor

FIGURA 3 – Resultado da consulta SPARQL

pessoa	email
ag:antoniu	"toninho@outroemail.com.br"
ag:antonio	"antonio@meuemail.com.br"
ag:jose	"jose@meuemail.com.br"
ag:maria	"maria@meuemail.com.br"

Fonte: próprio autor.

Nota-se que no CÓDIGO 4 fora alterada a variável antes utilizada, “?email_do_ze” para “?email”, visto que a nova consulta apresenta todo o conjunto de declarações que tenham a propriedade e-mail, independente do recurso.

Entender o conjunto de variações possíveis na cláusula WHERE é um fator de extrema importância no desenvolvimento das consultas escritas em SPARQL.

Ainda sobre o conjunto de dados apresentados no CÓDIGO 1, poderia propor várias outras alternativas na cláusula WHERE, de forma a recortar da base de dados um conjunto de dados diferentes. É possível pensar em buscar todos os dados de uma única pessoa, neste

caso deveriam ser utilizadas variáveis nos elementos que definem propriedade e valor, deixando o recurso com um valor definido. Poderia-se pensar também em buscar o telefone ao invés do e-mail, neste caso seria necessária a substituição da informação “ag:e-mail” por “ag:fone”.

4.2 SPARQL: experimentando implementações

A falta de publicações na área de Ciência da Informação que relatem implementações e uso do SPARQL pode estar associada a dificuldade em entender como pode ser realizado o processo de experimentação do protocolo.

O caminho mais utilizado é sempre o de operar os chamados *SPARQL EndPoints*, disponíveis na grande maioria dos *datasets* publicados na Internet. O DBPEDIA³⁰, o GEODATA³¹, entre outros vários *datasets* disponibilizam seus “pontos de acesso”. Nesse caso é possível consultar suas bases de dados em formato RDF e também cruzar dados com outras bases de dados, entretanto há de se utilizar sempre os dados públicos disponibilizados.

Um dos primeiros passos dessa pesquisa foi justamente implementar a estrutura funcional do SPARQL no próprio laboratório ou mais precisamente no equipamento de uso diário.

Pensando sempre em ambientes *opensource* e em soluções livres, característica de pesquisa deste pesquisador, o uso do Apache Jena³² é o mais adequado para iniciar os estudos sobre aplicações e implementações de Web Semântica e do protocolo SPARQL. O Apache Jena é um *framework* de código aberto para construção de Web Semântica e aplicações *Linked Data*. (APACHE, 2014).

O Apache Jena é constituído de um conjunto de aplicações com destaque para algumas delas: suporte a elaboração de projetos usando arquivos em formato RDF; implementação de banco de dados persistente do tipo *triplestore* (banco de dados específico para formato RDF); ambientes de consulta SPARQL; mecanismos de inferência e suporte a ontologias, entre outras implementações diversas.

Na FIGURA 4 é possível visualizar a estrutura funcional do Apache Jena e inclusive relacionar os recursos ao Bolo de Noiva de Berners-Lee criando associações para constituir

³⁰ <http://dbpedia.org/sparql>

³¹ <http://linkedgeodata.org/sparql>

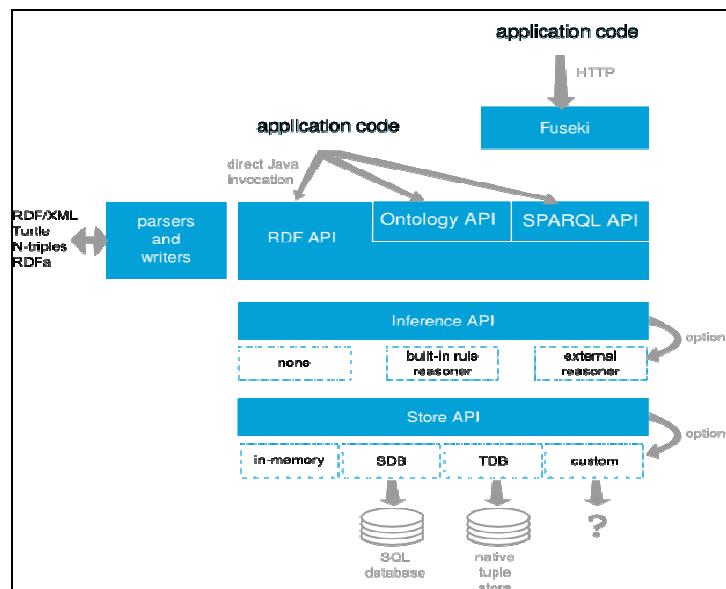
³² <https://jena.apache.org/>

implementações de Web Semântica. Na área de downloads do Apache Jena, é fácil encontrar, baixar e posteriormente instalar a ferramenta, que é multiplataforma.

Após a instalação do Jena ele fica disponível para uso na área de terminal (prompt de acesso) nos mais diferentes sistemas operacionais Linux, Windows ou MacOS, operando em processo parecido com o compilador Java.

Os CODIGOS e FIGURAS (resultados) que apresentaram implementações neste texto foram todos testados utilizando-se o Apache Jena.

FIGURA 4 – Estrutura Apache Jena



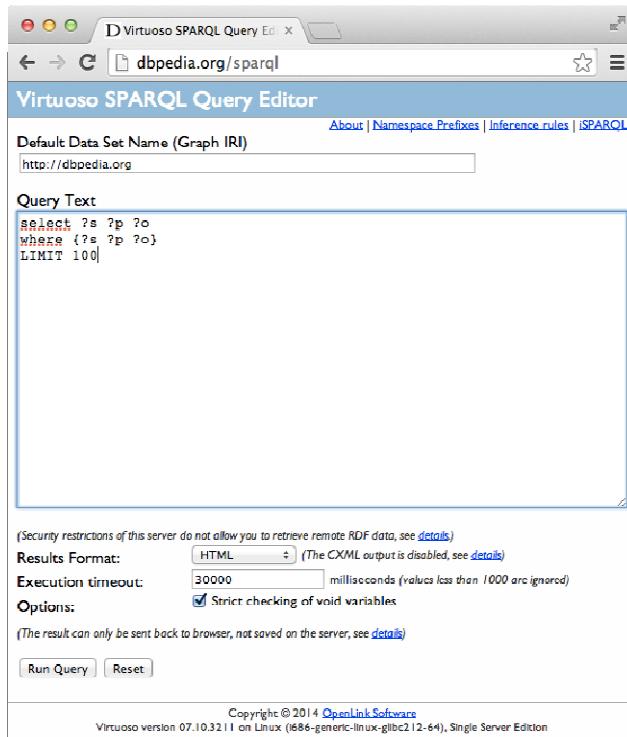
Fonte: APACHE (2014).

Se em um primeiro momento construir estruturas semânticas e utilizar o SPARQL in loco favorece e acelera o desenvolvimento de pesquisas relacionadas a Web Semântica, acessar e conhecer os dados disponibilizados em grandes *datasets* é o passo seguinte, porque permite vislumbrar não apenas o acesso a um determinado *dataset*, mas também o cruzamento de dados entre vários deles.

Os *datasets* disponíveis, em sua maioria, oferecem acesso via SPARQL *EndPoint*, que são interfaces online que permitem a realização de consultas diretamente por um endereço na web. A implementação dessa ferramenta dar-se-á pelo uso de bancos de dados do tipo *triplesstore*.

Bancos de dados do tipo *triplesstore* são diferentes dos bancos de dados relacionais, visto que permitem armazenar de forma nativa as declarações RDF e tem como característica permitir o acesso via protocolo SPARQL. Há uma lista imensa de ferramentas que implementam bancos de dados *triplesstore*, entre eles o Apache Jena, já citado.

FIGURA 5 – SPARQL EndPoint – DBpedia EndPoint³³



Fonte: DBPEDIA (2014)

Dentre os principais *datasets*, a ferramenta mais utilizada para implementar o banco de dados do tipo *triplestore* é o Virtuoso Open-Source Edition³⁴, que apesar de estar sob o domínio de uma empresa privada é uma plataforma *opensource*.

A interface do DBpedia, apresentada na FIGURA 5, possibilita o acesso aos dados disponibilizados no *dataset*. Para construir consultas é necessário conhecer a estrutura de dados disponibilizado pelo *DBpedia*, assim também será para utilizar outros *datasets*.

Através da interface do *DBpedia* é possível recuperar um conjunto diverso de informações. Um aficionado por eventos esportivos como os Jogos Olímpicos poderia ter interesse em saber todos os medalhistas de ouro da história ou de uma determinada modalidade olímpica; um fã de Ayrton Senna poderia ter interesse em saber as vitórias do mesmo em dias de chuva; um cinéfilo poderia se interessar por todos os filmes estrelados por *Morgan Freeman* ou então por filmes estrelados e dirigidos por *Clint Eastwood*.

Buscas por países que não fazem fronteira com o oceano, ou que tem períodos maiores que 18 horas de sol por dia e até musicas que duram mais de 60 minutos também são informações disponíveis no *DBpedia*, passíveis de serem consultadas.

³³ <http://dbpedia.org/sparql> (Usando Virtuoso OpenLink)

³⁴ <http://virtuoso.openlinksw.com/dataspace/doc/dav/wiki/Main/>

O objetivo deste trabalho não é discutir o uso das informações do *DBpedia*, aparentemente informal e com relevância específica para alguns contextos. Também não é proposta desta pesquisa analisar outros *datasets* disponíveis como de localização geográfica, dados de aeroportos, informações sobre publicações científicas que poderiam gerar muitos dados importantes para tomada de decisão do cotidiano das pessoas. Apresentar-se-á apenas a forma de se utilizar um *dataset* tendo como modelo o *DBpedia*.

Para se recuperar informações de um determinado *dataset* é importante que se conheça a estrutura semântica utilizada por ele, os vocabulários associados e todo o contexto de que envolve a construção da informação que está relacionada. Utilizando o contexto já citado, é possível verificar a pagina oferecida no *DBpedia* sobre o ator *Morgan Freeman* na FIGURA 6, ela apresenta as propriedades e valores associados ao recurso que identifica unicamente a pagina do ator.

FIGURA 6 – Recurso Morgan Freeman (DBpedia)

Property	Value
<code>dbpedia-owl:abstract</code>	<ul style="list-style-type: none"> Morgan Porterfield Freeman, Jr., é um ator, produtor cinematográfico e papel como o ex-lutador de boxe Eddie Scrap-Iron Dupris em <i>Million Dollar Baby</i>. 1984-01-01 00:00:00 (xsd:date)
<code>dbpedia-owl:activeYearsStartYear</code>	1997-06-01 (xsd:date)
<code>dbpedia-owl:birthDate</code>	dbpedia:Tennessee
<code>dbpedia-owl:birthPlace</code>	dbpedia:Memphis,_Tennessee
<code>dbpedia-owl:birthYear</code>	1937-01-01 00:00:00 (xsd:date)
<code>dbpedia-owl:occupation</code>	dbpedia:Morgan_Freeman_1
<code>dbpedia-owl:thumbnail</code>	http://upload.wikimedia.org/wikipedia/commons/thumb/c/c0/Morgan_Freeman.jpg/220px-Morgan_Freeman.jpg
<code>dbprop:title</code>	Awards for Morgan Freeman
<code>dbprop:wordnet_type</code>	http://www.w3.org/2006/03/wn/wn20/instances/synset-actor-noun
<code>dbprop:yearsActive</code>	1964 (xsd:integer)
<code>dc:description</code>	American actor
<code>is dbpedia-owl:portrayed_of</code>	dbpedia:Alex_Cross
<code>is dbpedia-owl:presenter_of</code>	dbpedia:Mississippi_Rising
<code>is dbpedia-owl:producer_of</code>	dbpedia:The_Maiden_Helst
<code>is dbpedia-owl:starring_of</code>	dbpedia:Under_Suspicion_(2000_film)
	dbpedia:Bruce_Almighty
	dbpedia:Chain_Reaction_(film)
	dbpedia:Kiss_the_Girls_(film)
	dbpedia:The_Big_Bounce_(2004_film)

Fonte: DBPEDIA (2014)

No caso específico desse recurso, é possível notar as propriedades e os valores que estão diretamente associados a ele. É possível notar propriedades oriundas do vocabulário DC (*Dublin Core*) além dos vocabulários próprios da *DBpedia* (*dbpedia-owl*). É possível também identificar que alguns valores associados as propriedades são outros recursos, o que efetiva a concepção de grafo ou de rede das declarações RDF.

O CÓDIGO 5 apresenta a consulta *SPARQL* constituída para recuperar os filmes estrelados por Morgan Freeman disponíveis no *DBpedia*.

CÓDIGO 5 – Consulta SPARQL em *DBpedia EndPoint*

```
SELECT ?filme
WHERE {
?filme dbpedia-owl:starring dbpedia:Morgan_Freeman
}
```

Fonte: próprio autor

A consulta apresentada no CÓDIGO 5 foi baseada em um estudo do próprio recurso visto na FIGURA 6. A cláusula “*WHERE*” identifica todos os recursos que tiveram como propriedade o termo “*dbpedia-owl:starring*” e como valor o termo “*dbpedia:Morgan_Freeman*”. Dessa forma o processador SPARQL deverá recortar de todo o conjunto de bilhões de declarações RDF do *DBpedia* apenas os que se enquadram nos quesitos descritos.

O CÓDIGO 6, agrega na cláusula “*WHERE*” um recorte sobre os filmes produzidos pelo mesmo *Morgan Freeman*. Neste caso é possível recuperar os filmes que foram estrelados e produzidos pelo mesmo.

Verifica-se neste caso que o “ponto” que liga as duas linhas que estão na cláusula “*WHERE*” serve para criar uma relação booleana do tipo *AND*, considerando uma intersecção entre o recorte produzido pelas duas linhas existentes na cláusula “*WHERE*”.

CÓDIGO 6 - Consulta SPARQL em *DBpedia EndPoint*

```
SELECT ?filme
WHERE {
?filme dbpedia-owl:starring dbpedia:Morgan_Freeman .
?filme dbpedia-owl:producer dbpedia:Morgan_Freeman
}
```

Fonte: próprio autor

Assim como fora apresentado o uso do *DBpedia*, qualquer outro *dataset* pode ser explorado e consumido utilizando-se *SPARQL*, desde que se conheça a estrutura lógica em que as informações estão disponíveis.

5 CONSIDERAÇÕES FINAIS

A Web Semântica é um fato, porém o uso dela passa inicialmente pela construção dos ambientes e posteriormente pela recuperação das informações. A área tem discutido de forma exaustiva o uso do padrão RDF e a construção de ontologias sólidas, porém ainda

existe uma necessidade efetiva de discussão sobre a construção de estruturas claras e funcionais de recuperação da informação em ambientes semânticos.

O *Linked Data* tornou-se a principal referência concreta de uso da Web Semântica, e sua proposta de dados ligados tem motivado as mais diversas organizações, instituições e setores a disponibilizar dados ligados em formato aberto.

O protocolo *SPARQL* apresentado e demonstrado de forma introdutória neste texto é a base para a recuperação da informação em ambientes semânticos, entretanto a maximização de seu uso efetivo está ligado diretamente ao domínio da sua grande diversidade de funcionalidades. Dentre as funcionalidades não apresentadas neste texto podemos citar ainda a atualização de dados e o uso de inferências.

Nos últimos anos a utilização do protocolo *SPARQL* já é possível devido a existência de um grande número de ferramentas que implementam não apenas especificações *SPARQL* mas também ambientes propícios de acesso e experimentação. O acesso a estas ferramentas, em grande parte *opensource*, facilita o desenvolvimento de pesquisas sobre recuperação semântica da informação. O uso do Apache Jena é um ótimo caminho para iniciar as atividades práticas com Web Semântica e *SPARQL*.

A disponibilidade de *SPARQL Endpoints* em grandes e conhecidos *datasets* é também um ótimo caminho para evoluir com as pesquisas sobre como o utilizar o protocolo *SPARQL* no construção de modelos de recuperação de dados ligados

REFERÊNCIAS

- AGROVOC. AGROVOC Linked Open Data.** 2014. Disponível em <<http://aims.fao.org/standards/agrovoc/linked-open-data>>. Acesso em: 10 jun. 2014.
- APACHE. Apache Jena.** 2014. Disponível em <<https://jena.apache.org/>>. Acesso em 20 jan. 2014.
- BERNERS-LEE T.; LASSILA, O.; HENDLER, J.** The semantic web. **Scientific American**, New York, v. 5, May 2001.
- BERNERS-LEE, T. Linked Data Principles.** 2006. Disponível em <<http://www.w3.org/DesignIssues/LinkedData.html>> Acesso em : 10 abr. 2014.
- DBPEDIA TEAM. DBpedia.** 2014. Disponível em <<http://thedatahub.org/dataset/dbpedia>>. Acesso em 02 jun. 2014.
- DUCHARME, B. Learning SPARQL.** Sebastopol: O'Reilly, 2011.
- GEONAMES. Geonames Ontology.** 2013. Disponível em : <<http://www.geonames.org/ontology/documentation.html>>. Acesso em: 10 jun. 2014.

GUARINO, N. Formal ontology and information systems. In: INTERNATIONAL CONFERENCE ON FORMAL ONTOLOGY IN INFORMATION SYSTEMS - FOIS'98, 1998, Trento. **Proceedings...** Amsterdam: IOS Press, 1998. p. 3-15.

HEATH, T.; BIZER, C. Linked Data: Evolving the Web into a Global Data Space. Morgan & Claypool, 2011.

JACOB, E. K. Ontologies and the semantic web. **Bulletin for the American Society for Information Science and Technology**, v. 29, n. 4, p. 19-22, Abr./Mayo 2003.

LASSILA, O. **Resource Description Framework (RDF) model and syntax specification 1.0.** 1999. Disponível em: <<http://www.w3c.org/TR/REC-rdf-syntax>>. Acesso em: 2 fev. 2014.

SANTAREM SEGUNDO, J. E. **Representação Iterativa: um modelo para repositórios digitais.** 2010. 224 f. Tese (Doutorado em Ciência da Informação) – Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília. 2010.

SPARQL **Sparql 1.1 Overview.** W3C, 2013. Disponível em <<http://www.w3.org/TR/2013/REC-sparql11-overview-20130321/>> Acesso em: 13 nov. 2013.

W3C. **Semantic Web.** W3C, 2010. Disponível em: <<http://www.w3.org/standards/semanticweb/>>. Acesso em: 10 de Abr. 2014.

DESIGN SCIENCE: FILOSOFIA DA PESQUISA EM CIÊNCIA DA INFORMAÇÃO E TECNOLOGIA

*DESIGN SCIENCE: PHILOSOPHY OF RESEARCH IN
INFORMATION SCIENCE AND TECHNOLOGY*

Marcello Peixoto Bax

Resumo: A *Design Science Research* (ou *DSR*) é uma meta-teoria que investiga a geração de conhecimento no processo de concepção de artefatos, i.e., sobre como métodos de *design* podem constituir pesquisa de caráter científico. O objetivo do artigo é apresentar uma análise do paradigma *DSR*, como quadro teórico-metodológico de fundamentação científica das pesquisas em tecnologia e gestão da informação e do conhecimento, importante campo da ciência da informação. Para analisar a *DSR* e sintetizar os resultados, o método usado foi o de rever na literatura pertinente os aspectos da *DSR* considerados mais relevantes para o campo. Foram analisados os artigos mais citados nos últimos 50 anos. Tem-se como resultado da análise, por um lado a constatação surpreendente do completo desconhecimento da *DSR* na literatura de ciência da informação em geral, tanto nacional quanto internacional e mais especificamente no campo citado. Conclui-se que, apesar de desconhecida na ciência da informação, a *DSR* deve ser vista como uma das mais apropriadas metodologias para orientar a condução de pesquisas científicas em tecnologia e gestão da informação e do conhecimento, em uma abordagem que alia a relevância da aplicação prática com o rigor científico.

Palavras-chave: ciência da informação; ciência de projeto; metodologia de pesquisa; filosofia da ciência; epistemologia.

Abstract: The Design Science Research (DSR) is a meta-theory that investigates the knowledge generation process while designing artifacts, i.e., how design methods can constitute research of scientific nature. The objective of this paper is to present an analysis of DSR paradigm as theoretical and methodological framework for doing scientific research in information and technology, major field of information science. To analyze the DSR and summarize the results, the method used was to make a literature review of the DSR's aspects considered most relevant to the field. The most cited papers over the last 50 years were analyzed. As a result we found a complete ignorance of DSR in the information science literature in general, both nationally and internationally and more specifically in that field. We conclude that, although unknown in information science, DSR should be seen as one of the most appropriate methodologies to guide the scientific research in information, technology and management, an approach that combines the relevance of practical application with scientific rigor.

Keywords: information science; design science research; research methodology; philosophy of science; epistemology.

1 INTRODUÇÃO

The inadequate articulation of its central conceptions as an orientation strategy³⁵ is also the crucial weakness of the information resources management movement [...]. The promise of the movement is that it is interested in both individual information use and its organizational an

³⁵ Discurso meta-teórico sobre a natureza e escopo de um campo de pesquisa que sugere orientações e estratégias para se conhecer os fenômenos que ocorrem no campo.

social context. [...]. That it has not articulated itself as an orientation strategy is due to the fact that it has developed an ideology but no basic concepts for the field (Pertti Vakkari, 1994, p.50).

No trecho acima Vakkari (1994) declarava, em 1994, que dos três paradigmas da ciência da informação – social, cognitivo e gerencial – este último, o gerencial³⁶ (e tecnológico), à época então conhecido como gerenciamento de recursos informacionais (ou GRI), constituía movimento promissor para a ciência da informação, mas que por não exercitar o auto-escrutínio de seus métodos, acabou se revelando muito mais uma “ideologia”, do que uma estratégia de orientação epistemológica (*i.e.*, de pesquisa rigorosa). No mesmo texto, Vakkari classifica a ciência da informação como uma “*design science*”, cuja missão é prover orientações (*guidelines*) para incrementar o acesso à informação.

The purpose to which information science is pledged is to facilitate access to desired information. [...] It is design science, whose mission it is to provide, with help from research, the guidelines through which access to information can be enhanced (Pertti Vakkari, 1994, p.47).

O autor concorda com Vakkari que a ciência da informação, como uma ciência social aplicada é, em grande parte, uma “ciência de projeto” ou *design science*; logo, a tarefa de explicitar a ontologia, epistemologia e metodologia de uma *design science* contribuirá muito para o amadurecimento desta ciência. Por falta de espaço, entretanto, nosso objetivo aqui ainda não será este³⁷. Uma constatação surpreendente do presente artigo é que, embora já em 1994, Vakkari classificasse a ciência da informação como uma *design science*, a natureza de uma tal ciência de projeto³⁸ ainda persiste por demais desconhecida da comunidade dos pesquisadores da ciência da informação. Com efeito, foi no campo da computação e dos sistemas de informação que a ciência do *design* foi primeiro proposta. Uma “ciência do artificial” como a batiza Herbert Simon (1996), em 1968. Pensada como uma ciência do desenvolvimento de artefatos, a *design science* foi inicialmente proposta por Simon, sob o olhar da filosofia da ciência, como um contraponto às ciências naturais e à hermenêutica. As disciplinas consideradas “tecnológicas” e a pesquisa relacionada, centram-se no desenvolvimento, compreensão e uso da TI para atender às necessidades práticas de indivíduos e organizações. Artefatos e em particular *software*, são a base operacional da pesquisa nesta área de investigação.

³⁶ Em geral esse paradigma é representado, na ciência da informação, pela linha de pesquisa gestão da informação e do conhecimento (GIC).

³⁷ Realizar uma reflexão mais aprofundada sobre os princípios filosóficos da design science research é trabalho futuro, a ser publicado em outro artigo.

³⁸ Alguns textos em português traduzem design science como ciência de projeto. Nós preservaremos, neste artigo, o termo na sua grafia original em língua inglesa.

Segundo Gregg (2001), embora a importância do desenvolvimento de *software* seja bem aceita, o que constitui pesquisa de qualidade na área ainda não é tão estabelecido. Talvez porque a mera aplicação de um método consolidado de desenvolvimento de *software*, por si, não constitua necessariamente uma atividade de pesquisa. Também não é trivial distinguir entre os esforços de desenvolvimento, aqueles que contribuem para fazer avançar os limites do conhecimento científico na área, daqueles que são simples implementações pontuais que solucionam problemas conhecidos reusando conhecimentos já consolidados na literatura. Essa dificuldade gera a visão comum e recorrente de que um projeto de concepção de um sistema de informação ou de outro artefato tecnológico qualquer dificilmente poderá caracterizar pesquisa científica. O argumento é de que tais projetos não ensejam o desenvolvimento de conhecimento novo, de caráter rigoroso.

A questão é análoga à discussão sobre como justificar o caráter científico das pesquisas realizadas no âmbito do paradigma gerencial e tecnológico da ciência da informação. E mais especificamente das pesquisas realizadas no âmbito do GT8 do Enancib³⁹, visto a preponderância desse tipo de pesquisa no GT. Contudo, a preocupação com o uso de métodos adequados de pesquisa na ciência da informação, mais especificamente em sua subárea de sistemas e informação, gestão e tecnologia (GT8, Enancib) é ausente na literatura nacional e o paradigma discutido neste artigo, a *Design Science Research* (DSR) é desconhecido da ciência da informação, como concluiu o esforço de pesquisa aqui relatado. Ora, a DSR é exatamente o tipo de meta-teoria que auxilia o pesquisador a criar conhecimento teórico durante os processos mesmos de concepção de artefatos, justificando como tais processo podem constituir pesquisa de caráter científico.

Outra forma de colocar a questão é quando se busca diferenciar, no contexto de projetos de pesquisa do GT8, entre um projeto de pesquisa, desenvolvimento e inovação (PD&I), de um projeto de consultoria e desenvolvimento. Trata-se de um difícil problema que diz respeito ao nível de complexidade do tema que se tem em mãos⁴⁰. A distinção é crucial entretanto, pois projetos de PD&I necessitam do envolvimento de pesquisadores ou instituição de pesquisa. Para esses últimos, em geral, não existem nem “boas” nem “melhores práticas” estabelecidas no mercado. Enquanto projetos rotineiros de desenvolvimento podem ser especificados e contratados de fornecedores de TI no mercado.

³⁹ Grupo de Trabalho Informação e Tecnologia (GT8) do Encontro Nacional de Pesquisa em Ciência da Informação no Brasil (Enancib).

⁴⁰ A esse propósito vale a pena consultar o arcabouço Cynefin (Kurtz e Snowden, 2003). O Cynefin é derivado de pesquisa em complexidade organizacional e troca de conhecimento, tomada de decisão, estratégia e formulação de políticas.

Neste último caso, “melhores práticas”, normalmente, já se encontram consolidadas. A dificuldade é que, obviamente, existe uma larga zona cinzenta entre esses dois extremos.

Contudo, vale notar que o objetivo aqui não é o de argumentar que a atividade, desde que rigorosa, de construção de um sistema de informação (artefato⁴¹) possa ser considerada, em muitos casos, uma atividade de pesquisa. Ou seja, uma atividade que envolve reflexão tanto sobre o processo de construção em si (*engenharia de software*), quanto sobre como tal processo, ao produzir o artefato, resulta ou gera conhecimento científico sobre como melhorar a eficácia e eficiência dos processos da organização. Tal é a essência do paradigma da DSR e isso já foi feito por inúmeros trabalhos, que definiram arcabouços⁴² ou quadros teórico-metodológicos, dos quais citamos: Simon (1966), Nunamaker (1994), Gregg *et al.* (2001), Vashinav (2001), Hevner *et al.* (2004) e Wieringa (2009). Embora não haja consenso sobre os fundamentos filosóficos da *DSR*, os autores citados a destacam como sendo singular e única, diferenciando-a e contrastando-a com os outros principais paradigmas positivista e hermenêutico, mas também concordando que ela compartilha dos fundamentos de uma ciência natural e esteja próxima do paradigma hermenêutico, interpretativo.

O objetivo deste artigo é destacar, na literatura, os pontos de atenção mais importantes da DSR para a ciência da informação, bem como comentar sobre a evolução do paradigma na visão de alguns de seus principais autores. Com isso espera-se apresentar a DSR como alternativa metodológica para o paradigma gerencial e tecnológico da ciência da informação de que dá falta Vakkari (1994), na citação no início desta introdução. A pretensão é colaborar para que sejam realizadas pesquisas de melhor qualidade na área, sobretudo aquelas no âmbito do GT8, do Enancib.

No intuito de fundamentar a análise do problema sugerido (e explicitado no objetivo declarado acima) e de sintetizar elementos de resposta, o método usado foi realizar uma revisão da literatura pertinente que analisa os aspectos epistêmicos mais relevantes de uma *design science* capazes de melhor fundamentar o conhecimento científico no campo gerencial e tecnológico da ciência da informação. Foram analisados os artigos mais citados, dos anos de 1960 até hoje, em periódicos internacionais, sobre o tema "*design science*" e "*desing science research*". As reflexões advindas desse trabalho fundamentarão no futuro a síntese de um quadro teórico-metodológico de cunho pragmático que possa orientar a realização de pesquisas de melhor qualidade no campo.

⁴¹ Artefato: é aqui definido como um objeto cuja construção segue métodos científicos.

⁴² Ou Research Frameworks.

Contrariamente ao positivismo comum que pleiteia a separação clara e precisa entre pesquisa de cunho científico de um lado, e de cunho tecnológico de outro⁴³, e corroborando Wieringa (2009), defende-se aqui que na pesquisa social aplicada há, em geral, uma nidificação mútua e sempre recorrente de problemas práticos e problemas teóricos (questões de conhecimento, no vocabulário de Wieringa). Assim, entende-se que a base que determina o nível de qualidade científica da pesquisa no campo do GT8 acaba sendo sobretudo de cunho metodológico, i.e., quando esses dois tipos de problemas são confundidos, o resultado comprometerá, quase sempre, a qualidade da pesquisa, justificando aquela estreita visão positivista comum. Com efeito, quando a metodologia falha em orientar o pesquisador, pode-se esperar por duas situações indesejáveis para uma ciência social aplicada: ou se produz o artefato e não se produz conhecimento, ou se produz conhecimento sem se produzir o artefato⁴⁴. Ou seja, perde-se no rigor científico ou na relevância da pesquisa; um resultado indesejado, mesmo sob a égide do compromisso apodítico entre rigor e relevância.

Finalmente, para concluir o artigo, ilustra-se através de um exemplo prático de projeto, como o quadro teórico e metodológico da DSR pode contribuir para justificar a diferença entre projetos de desenvolvimento e projetos de PD&I, como foi comentado acima. Como em qualquer pesquisa científica, pressupostos filosóficos são assumidos, mesmo que de forma inconsciente, pelo pesquisador. Considera-se que explicitar tais pressupostos, à luz da DSR, ajudará o pesquisador a se orientar na ciência de projeto (ou *design science*), do ponto de vista metodológico.

2 INTRODUÇÃO À DESIGN SCIENCE E À DESIGN SCIENCE RESEARCH

O paradigma DSR foi amplamente discutido nos últimos anos e agora ganha terreno como um quadro teórico ou uma estratégia de pesquisa capaz de orientar, tanto a construção do conhecimento, quanto aprimorar as práticas em sistemas de informação e de várias disciplinas relacionadas ao campo gerencial e tecnológico da ciência da informação. No paradigma DSR, o conhecimento e a compreensão de um domínio do problema e sua solução são alcançados graças à construção e aplicação de um artefato projetado. Segundo Hevner *et al.* (2004), numa abordagem pragmática, a DSR não anseia alcançar verdades últimas, grandes teorias ou leis gerais, mas procura identificar e compreender os problemas

⁴³ TAKAHASHI (2009) mostra que a diferença entre esses dois tipos de conhecimento está longe de ser trivial.

⁴⁴ Resultado de uma relação analítica pouco dialética entre relevância x rigor, incapaz de sínteses inovadoras.

do mundo real e propor soluções apropriadas, úteis, fazendo avançar o conhecimento teórico da área.

A DSR envolve construir, investigar, validar e avaliar artefatos tais como: construtos, arcabouços, modelos, métodos e instâncias de sistema de informações, a fim de resolver novos problemas práticos. Além disso, o estudo de métodos, comportamentos e melhores práticas relacionadas com a análise do problema e com o processo de desenvolvimento de artefato são abrangidos. Debates em curso sobre a natureza e o alcance dos arcabouços de pesquisa dominantes em DSR, no entanto, mostram que o paradigma ainda está emergindo. Seu núcleo, seus limites e sua interação com outras abordagens de pesquisa ainda estão sendo revelados e definidos (VAISHNAVI, 2009). Trata-se então de um momento propício para os pesquisadores da ciência da informação estudarem, compreenderem e aplicarem tal paradigma em suas pesquisas.

2.1 Histórico da *DSR* Como Pesquisa em Sistemas de Informação

A ciência de projeto (*design science*) aparece pela primeira vez em Fuller (1965) que a definiu como uma forma sistemática de projetar ou conceber coisas. O conceito foi retomado por Gregory (1966), que pontua a distinção entre a concepção pelo método científico (*design science research*) e o método de projeto ou concepção em si (*design*, simplesmente). Para Gregory, a concepção ou projeto (*design*) não era uma ciência, pois uma ciência da concepção deve ir mais fundo e referir-se à uma visão *científica* do processo de criação. Em seguida Simon (1996) no ano de 1968 populariza esses termos em seu argumento para o estudo científico do artificial (em oposição ao natural). De lá até hoje os dois termos têm sido confundidos ao ponto onde a DS passou a ter dois sentidos (*design* e *design science*), com o significado de estudo científico da concepção agora predominante. O tema foi mais recentemente rebatizado pela expressão “*design thinking*”, cf. Visser (2006). Na verdade, nota-se que é um mesmo conjunto de ideias, centrado na utilização de ferramental tecnológico contemporâneo, que está de algumas das mais conhecidas práticas de gestão moderna, tais como o *design thinking*, o desenvolvimento ágil, o *customer development* e o *lean startup*.

Vale notar que desde o início dos anos 1960 cientistas da computação vêm fazendo pesquisas no estilo DSR sem, entretanto, nomeá-las assim. Ou seja, realizando o ciclo racional implicitamente (planejar > conceber > avaliar), foram desenvolvidas novas linguagens de programação, compiladores, algoritmos, modelos e estruturas de dados e arquivos, sistemas de gestão de banco de dados, e assim por diante. Um observador de fora da computação pode se fazer a ideia de que não se discute internamente a epistemologia da

área; de que não se faz o “auto-escrutínio”, nos termos de Vakkari. E portanto não se faz ciência de primeira qualidade, por ausência de discussão dos fundamentos filosóficos. Contudo, a melhor explicação para o fato é exatamente a presença deste racional implícito, que advém de uma notável unidade paradigmática do campo da computação. Já no campo mais recente dos sistemas de informação, grande parte da pesquisa inicial foi focada em abordagens de desenvolvimento de sistemas e métodos, que também são representantes da pesquisa do tipo DSR.

Sistemas de Informação (SI) é um campo relativamente novo de estudo que investiga tecnologias de informação e comunicação em contextos organizacionais. Originalmente um ramo da ciência de gestão, tornou-se um campo independente já no final de 1960. Autores divergem quanto à evolução histórica do uso da DSR como pesquisa em SI. Para alguns só recentemente é que a DSR torna-se uma linha distinta de investigação na área. Para outros as origens da pesquisa na área seguiram a DSR, pendendo-a de vista contudo nos anos 1980 e 90 e somente agora retornando como uma abordagem metodológica válida para suas pesquisas. Kuechler e Vaishnavi (2008) detalham o surgimento da DSR na área de SI e descrevem seu estado atual.

Segundo Livari (2007), nos últimos 30 anos a pesquisa em SI perdeu de vista sua origem como DSR. A filosofia da pesquisa dominante passou a ser a de desenvolver pesquisas cumulativas, estritamente teóricas capazes de fazer prescrições. Afirma o autor que essa estratégia de pesquisa "teoria-com-implicações-práticas" não conseguiu produzir resultados de interesse prático, perdendo relevância. O que impulsionou, no início do século XXI, uma retomada dos métodos práticos de pesquisa, como a DSR (LIVARI, 2007, p.40).

Wieringa (2009) e vários outros autores seminais da área relatam que as primeiras ideias que, mais tarde inspirariam a DSR como é hoje conhecida, teriam iniciado com o artigo de Simon (1996). A partir do qual March e Smith (1995, p.253) introduzem a DSR como tentativas de criar soluções que servem aos propósitos humanos, de resolver problemas.

Mais tarde, vários pioneiros, dentre eles Hevner *et al.* (2004), forneceram um quadro teórico-metodológico para a DSR em que (1) as necessidades de negócios motivam o desenvolvimento de artefatos (validados) que atendam a essas necessidades, e em que (2) o desenvolvimento de teorias justificadas sobre esses artefatos produz conhecimento que pode ser adicionado à base de conhecimento compartilhada. Este quadro esclarece a interface de DSR com o seu ambiente social e com a base de conhecimento científico, e deixa claro que a DSR e a pesquisa em ciências naturais (viés comportamental) são

atividades intimamente relacionadas (*behavioral science* e *design science*); o que é um contraponto à afirmação de Vakkari (1994), de que a influência das ciências naturais na ciência da informação não é grande. No artigo, Vakkari (cf. Vakkari, 1994, p. 50) critica a linha de GIC⁴⁵ (na época chamada GRI⁴⁶) por falta de rigor científico. Não percebendo que a DSR poderia ser a metodologia científica adequada para transformar as “ideologias” (na visão daquele autor), que eram geradas pela linha, em teorias científicas efetivas.

No Brasil vale notar que em trabalho no âmbito do projeto “Gestão de Operações em Organizações Inovadoras”, financiado pela CAPES, desenvolvido com a participação de programas pós-graduações de várias universidades (PPGEPS/UNISINOS, PEP/COPPE /UFRJ, PEP-PE/UFPE, AI/INPI e Poli/USP), a DSR foi vista como alternativa concreta de método para a condução de pesquisas de cunho tecnológico (tecnologias de gestão, p.ex.), constituindo-se em uma abordagem que, quando bem aplicada, produz rigor científico efetivo (LACERDA, 2013).

2.2 O CICLO Regulador da DSR

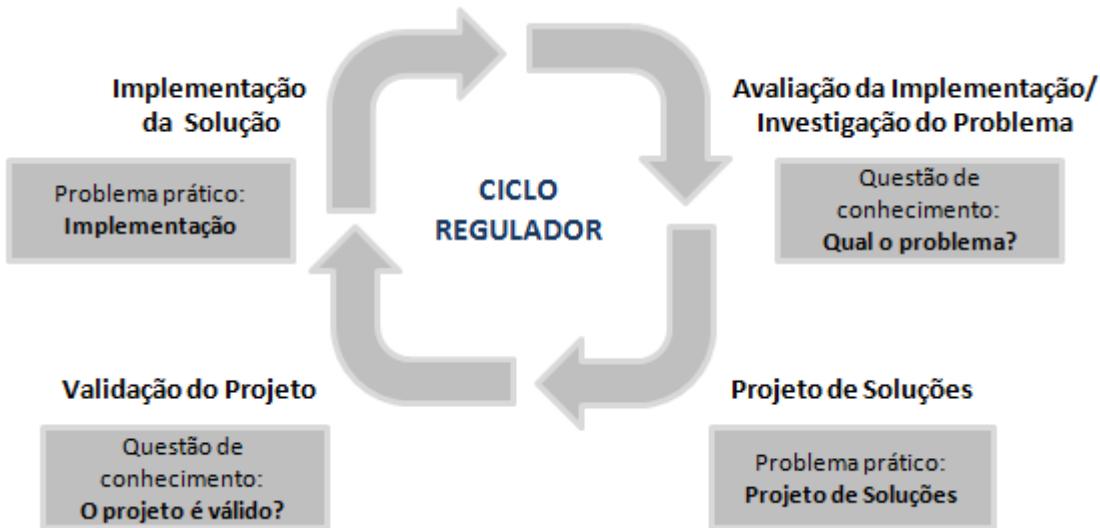
Wieringa (2009) chama a atenção para o fato de que a estrutura lógica usada na resolução de problemas consiste no tradicional ciclo regulador⁴⁷ que envolve as atividades de (1) Investigação do problema; (2) Projeto de soluções; (3) Validação da solução; (4) Implementação da solução e (5) Avaliação da implementação (com pequenas variações). O ciclo regulador pode ser visto na FIG. 1, que apresenta suas etapas caracterizando cada uma como um problema prático ou de conhecimento, ponto discutido em detalhes na Seção 4.

⁴⁵ Gestão da informação e conhecimento, linha de pesquisa da ciência da informação.

⁴⁶ Gerenciamento de recursos informacionais.

⁴⁷ Também análogo ao conceito de PDCA (de Deming e Shewhart), que por sua vez é baseado no método científico, tal como foi desenvolvido a partir do trabalho de Francis Bacon (Novum Organum, 1620). O método científico pode ser escrito como "hipótese" - "experimento" - "avaliação" ou planejar, fazer e verificar. Ou ainda problema > proposta > avaliação > novo problema.

FIGURA 2 – Ciclo regulador e a decomposição de um problema prático



Fonte: o autor, baseado em Wieringa (2009).

De forma a auxiliar na decomposição dos problemas e explicitar as oportunidades de geração de conhecimento científico durante a execução do ciclo regulador, Wieringa (2009) propõe uma classificação precisa dos problemas, distinguindo entre problemas práticos e teóricos (que ele denomina “questões de conhecimento”). Veremos a seguir que este ciclo regulador serve tanto para orientar a parte prática da pesquisa, i.e. resolução de problemas práticos, quanto a parte teórica, quanto trata-se sobretudo de gerar conhecimento respondendo a questões de conhecimento (*i.e.*, questões teóricas).

2.2.1 Tarefas do ciclo regulador

O ciclo regulador, proposto por diversos autores com pequenas diferenças, compiladas em Lacerda (2012), começa com uma investigação de um problema prático, ele mesmo resultado da resolução de problemas práticos anteriores; em seguida, especifica projetos de soluções, os valida, e implementa o projeto selecionado; cujo resultado, então, será avaliado, o que poderá ser o início de uma nova rodada do ciclo regulador.

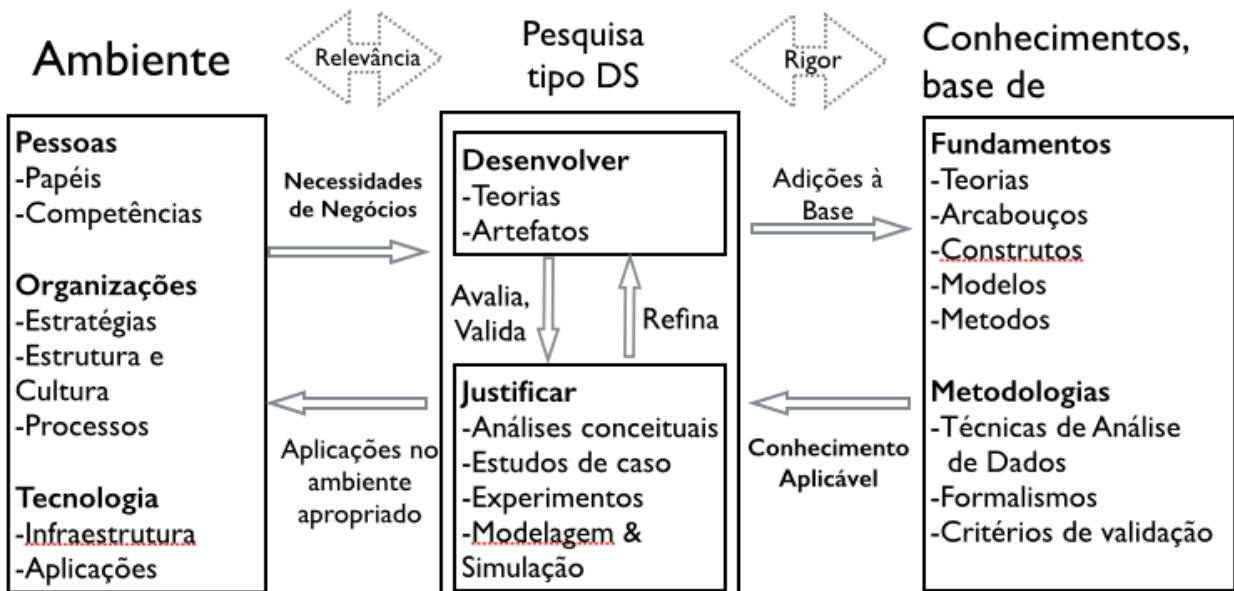
Wieringa (2009) cita pesquisadores de diversas áreas tais como: desenvolvimento de produto, engenharia de sistemas, arquitetura, engenharia mecânica e até na prática psicológica, que com algumas variações, identificam o ciclo regulador como representação da estrutura lógica da ação racional (tanto prática quanto teórica). O ciclo regulador é a estrutura geral de um processo racional de resolução de problemas: análise da situação atual e das metas de mudança novas ou em curso, propor possíveis alterações para atingir essas

metas, avaliar as alterações possíveis e selecionar uma, aplicar a alteração e, em seguida, reiniciar o ciclo.

3 O QUADRO TEÓRICO DA DSR DE HEVNER

Para Hevner *et al.* (2004), a DSR é fundamentalmente um paradigma para a solução de problemas. Usa-se a criação de artefatos, com bases nas leis naturais e nas teorias comportamentais, que são aplicados, testados, modificados e estendidos por meio da experiência, criatividade, intuição e habilidades de resolução de problemas de seus pesquisadores. Os artefatos podem ser dos tipos: construtos (entidades e relações), modelos (abstrações e representações), métodos (algoritmos e práticas) e instanciações (implementação de sistemas e protótipos), corroborando March e Smith (1995). Portanto os artefatos podem ser desde um *software*, uma teoria, até descrições informais em linguagem natural. Para os autores, artefatos auxiliam na definição de ideias, práticas, capacidades técnicas e produtos. O conhecimento e a compreensão do problema a ser resolvido são adquiridos durante a construção e o uso dos artefatos, quando estes são validados e avaliados.

Para Hevner a DSR pode ser vista como uma conjunção de três ciclos reguladores de atividades relacionadas (Hevner, 2007). O “ciclo de relevância” inicia a pesquisa com um contexto de aplicação que não só fornece como insumo os requisitos para a pesquisa, mas também define os critérios de aceitação para a avaliação dos resultados da investigação. O “ciclo de rigor” fornece o conhecimento científico passado ao projeto de pesquisa para assegurar a sua inovação. A boa condução deste ciclo depende do pesquisador realizar uma investigação o mais completa possível na base de conhecimento (*i.e.*, conjunto das bases de referência da literatura científica da área), fazendo referências aos trabalhos correlatos. O objetivo é garantir que os artefatos produzidos são contribuições de pesquisa efetivas e não projetos de rotina, baseados na aplicação de processos já conhecidos (pelos pesquisadores da área). Finalmente, o “ciclo central” itera entre as atividades principais de construção e avaliação dos artefatos de *design* e processos da pesquisa. Os três ciclos estão representados na FIG. 2.

FIGURA 2 – *Design Science Research Framework*

Fonte: o autor, baseado em Hevner *et al.* (2004).

De acordo com Hevner *et al.* (2004) a *design science* e a ciência comportamental são dois paradigmas relacionados que caracterizam a pesquisa em SI, e por consequência também no campo da informação, tecnologia e gestão. A DSR tem suas raízes em disciplinas de engenharia e como visto acima é uma jovem disciplina, enquanto a ciência comportamental tem suas raízes nos métodos de pesquisa científica naturais e tem uma história bem mais longa. A DSR procura criar inovações e visa a utilidade, enquanto a ciência comportamental foca em desenvolver e justificar a teoria. Por isso elas devem “andar juntas”.

Finalmente, vale notar que na pesquisa comportamental, a significância estatística é estabelecida como uma medida de rigor de seus resultados - mas a relevância destes resultados varia. Orientada para a atividade de concepção, a DSR tem como objetivo a construção de melhores soluções, e a utilidade na prática é estabelecida como uma medida de relevância de seus resultados - mas o rigor de sua construção e avaliação varia. Eis outra razão para distinguir claramente problemas práticos de questões de conhecimento: garantir o compromisso adequado entre rigor e relevância, que devem possuir todas as ciências sociais aplicadas, como a ciência da informação.

4 O QUADRO TEÓRICO DE DSR DE WIERINGA

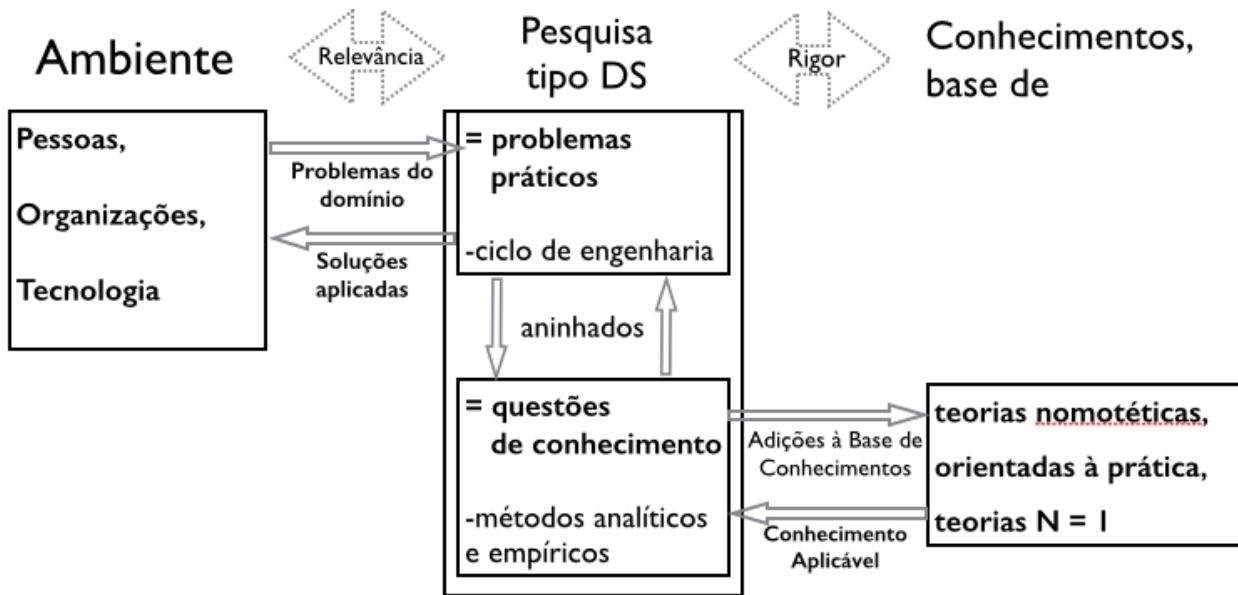
Até aqui já ficou claro que a ciência da concepção (*Design Science*) enfatiza a conexão entre conhecimento e prática, mostrando ser possível produzir conhecimento científico através da concepção de “coisas” úteis. Wieringa (2009) revela a existência de

“questões de conhecimento” que ocorrem durante as iterações dos ciclos. “Questões de conhecimento”, no jargão do autor, representam oportunidades de geração de conhecimento científico. Segundo o autor, com vistas a responder a questões de conhecimento pode-se procurar pela solução na literatura, consultar alguém que já resolveu questão similar ou mesmo usar o próprio conhecimento advindo de experiências anteriores.

No entanto, na ausência de diretrizes mais detalhadas, os pesquisadores tendem a identificar **problemas práticos** com **perguntas de conhecimento**, o que pode levar a projetos de pesquisa metodologicamente fracos. Portanto, é fundamental entender que para se resolver um problema prático, o mundo real é alterado para atender aos propósitos humanos, mas para resolver um problema de conhecimento, deve-se adquirir conhecimento sobre o mundo, sem necessariamente alterá-lo. Como argumenta Wieringa (2009), na DSR esses dois tipos de problemas são mutuamente aninhados. Ressalta o autor, contudo, que este aninhamento não deve nos cegar para o fato de que os métodos de resolução de problemas e de justificação das soluções são diferentes em cada caso.

Wieringa (2009) estende o quadro teórico de Hevner *et al.* (2004) de três formas (*cf. FIG. 3*). Em primeiro lugar, há uma maior elaboração (no ciclo da relevância) da investigação problema para se obter mais precisamente as necessidades de negócios. Segundo o autor a investigação pode ser “orientada a um problema”, a um “objetivo”, a uma “solução” e ao “impacto de uma solução” (WIERINGA, 2009, p.3). Este aspecto não recebeu atenção no quadro de Hevner *et al.* (2004), mas a investigação mais detalhada da natureza do problema é uma tarefa importante no caminho para se alcançar uma maior relevância da pesquisa. Em segundo lugar, a base de conhecimento no quadro de Wieringa é estratificada em níveis de generalização das teorias. Ele as classifica, da mais específica para a mais genérica. Organizando-as em teorias $N = 1$ (a mais específica), passando por teorias orientadas a classes de problemas (nível intermediário) e chegando em teorias nomotéticas, que são as mais gerais, dotadas de normas precisas que regulam os fenômenos. As DSs ocupam sobretudo o nível intermediário.

FIGURA 3 – Quadro teórico-metodológico da DSR de Roel Wieringa



Fonte: o autor, baseado em Wieringa (2009).

Em terceiro lugar, no quadro teórico de Wieringa, o ciclo regulador central da DSR é decomposto em uma estrutura aninhada de problemas práticos e questões de conhecimento. Questões de conhecimento podem ser respondidas pela aplicação do conhecimento a partir da base de conhecimento ou fazendo pesquisa original, usando a análise conceitual ou métodos empíricos tais como experimentos, estudos de caso, pesquisa de campo ou de modelagem e simulação. Os problemas práticos são resolvidos casando problemas e soluções durante os ciclos incrementais, em espiral.

Vê-se que as diretrizes básicas da DSR, tal como estabelecidas por Hevner *et al.* (2004), são suportados, comentadas e ajustadas pelo quadro teórico de Wieringa (2009). As diretrizes da DSR são: (1) O projeto de pesquisa produz artefatos; (2) problemas devem ser potencialmente relevantes e são dependente do contexto (organizacional); (3) propostas de solução devem ser validadas; (4) e devem produzir teorias que melhorem as práticas; (5) estas devem ser validadas com métodos científicos; (6) a pesquisa deve iterar, no ciclo de regulador, entre concepção e validação; (7) os resultados devem ser comunicados (publicados nas bases de conhecimento referenciadas).

O quadro proposto é também muito próximo daquele de Nunamaker e Chen (1990), que apresentam um processo de pesquisa de desenvolvimento de sistemas, na forma também de um ciclo regulador. As diferenças são que Wieringa distingue problemas práticos de questões de conhecimento e que ele dá maior atenção à investigação do

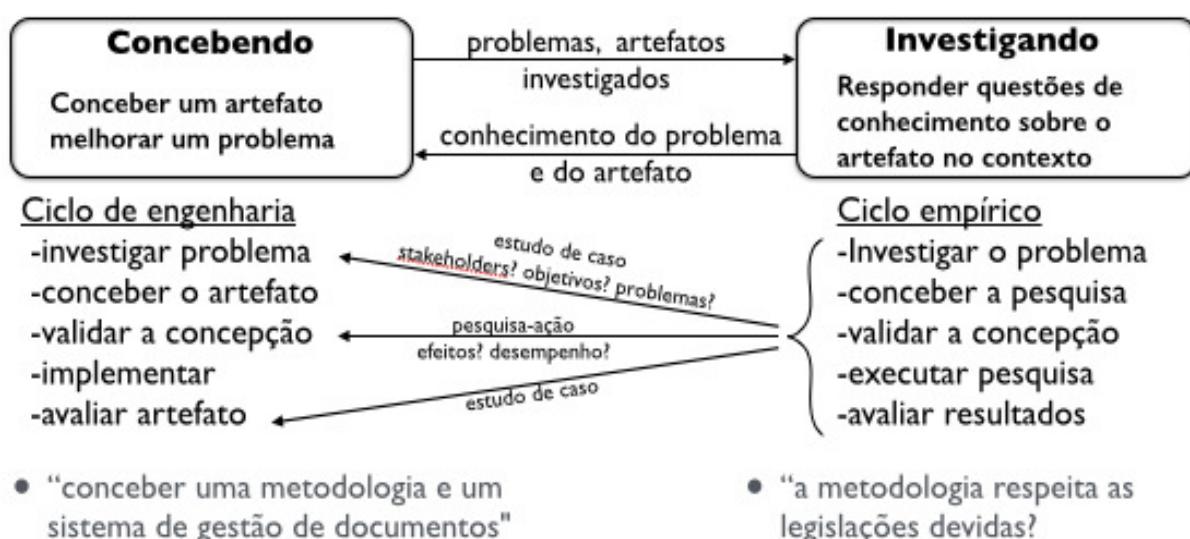
problema, além de mostrar o aninhamento teórico/prático sempre presente neste tipo de investigação científica.

5 DISCUSSÃO

Se Hevner *et al.* (2004) deixam claro que *design* e pesquisa são atividades intimamente relacionadas (*behavioral science* e *design science*), coube no entanto à Wieringa e Heerkens (2009) mostrarem que a interação dessas atividades não era profundamente analisada, e isso criava problemas metodológicos. Em particular, os problemas práticos e problemas de conhecimento em *design science* podem ser confundidos, e isso pode criar problemas metodológicos. Quando confundidos corre-se o risco de enfatizar-se apenas um deles, esquecendo-se do outro. Ou seja, pode-se por um lado resolver um problema prático sem que haja a criação explícita de conhecimento ou se pretender criar conhecimento sem nenhuma experimentação prática.

Wieringa identifica dois ciclos executados em paralelo durante uma iteração do ciclo regulador. Um ciclo de engenharia que investiga o problema, concebe a solução, valida, implementa e avalia. Um ciclo empírico (ou de conhecimento) que investiga o problema, concebe a pesquisa, valida, executa e avalia os resultados.

FIGURA 4 – Ciclos prático e teórico



O pesquisador interage sobre essas duas atividades

Fonte: o autor, baseado em Wieringa (2013).

Questões de conhecimento estão presentes no ciclo regulador. Para resolver um problema de conhecimento, fazendo pesquisa, tem-se que descobrir o que se quer saber, projetar a pesquisa com a qual se pretende adquirir este conhecimento, validar este projeto,

executar a pesquisa como projetado, e, finalmente, avaliar os resultados. Este ciclo de pesquisa tem a mesma estrutura que o ciclo de engenharia, e ambos têm a estrutura da ação racional, em que planejamos uma ação antes de agir, e na qual podemos avaliar o resultado de uma ação após o ato.

As tarefas de ambos os ciclos, de engenharia e de conhecimento são muito semelhantes, a distinção não sugere que sejam fenômenos separados, embora devam ser distintos durante a execução da pesquisa para melhor conceber e aplicar as metodologias e com isso melhorar os resultados. A diferença principal é de objetivos: enquanto um altera o estado do mundo e ganha conhecimento com isso, o outro altera o estado do conhecimento e o aplica no mundo para validar a alteração.

Wieringa (2009) apresenta a distinção entre os tipos de problemas a serem resolvidos: o problema prático e o de conhecimento. Ele define um problema prático como a diferença entre a forma como o mundo é experimentado pelas pessoas e como elas gostariam que fosse. Já o problema de conhecimento é a diferença entre o conhecimento das pessoas sobre o mundo e o que elas gostariam de saber. Os *problemas práticos* exigem uma mudança do mundo para que este melhor concorde com alguns objetivos das partes interessadas. *Problemas de conhecimento* por sua vez não demandam uma mudança do mundo, mas uma mudança do nosso conhecimento sobre o mundo (WIERINGA e HEERKENS, 2009). Essa distinção não deve porém sugerir que sejam problemas separados. A diferença principal é de objetivos. Enquanto um altera o estado do mundo e ganha conhecimento com isso o outro altera o estado do conhecimento e o aplica no mundo para validá-lo.

Como explica o autor, esta separação entre "problema de mundo" e "problema de conhecimento", termos sinônimos a "problemas práticos" e "questões de conhecimento", pode criar a falsa impressão de que é sempre possível gerar conhecimento sem mudar o mundo ou mudar o mundo sem gerar conhecimento. Para o autor é difícil gerar conhecimento sem mudar o mundo, e alguns métodos de pesquisa, como a pesquisa-ação e experimentos de laboratório, geram conhecimento através da manipulação do mundo, isto é, alterando-o. Os pesquisadores devem ainda demonstrar que estas mudanças no mundo não invalidam o conhecimento declarado, gerado pela pesquisa. Na outra extremidade do espectro, é também extremamente difícil mudar o mundo sem aprender alguma coisa com ele. Resolver um problema do mundo geralmente gera conhecimento. Assim, conclui o autor que a distinção entre os problemas do mundo e os problemas de conhecimento é antes de tudo uma questão de objetivos: quando o objetivo é resolver um problema de

conhecimento, faz-se alguma coisa para obter conhecimento e controla-se as mudanças criadas no mundo, de tal maneira a garantir que o conhecimento que se adquiriu é válido. Quando o objetivo é resolver um problema do mundo, muda-se o mundo e tem-se o conhecimento adquirido a partir dessa mudança, avaliando-a.

Estes dois problemas são ambos cidadãos de primeira classe na *DSR*, na medida em que podem ser igualmente desafiadores, mas eles exigem diferenças metodológicas importantes. As respostas para problemas de conhecimento são proposições que são declaradas e verificadas como verdadeiras ou falsas; soluções para problemas práticos são mudanças no mundo indicadas (prescritas) para cumprir as metas dos *stakeholders*. Como consequência, a resolução de problemas práticos envolve investigar os objetivos das partes interessadas e a avaliação de soluções envolve a aplicação de critérios de partes interessadas; mas responder a problemas de conhecimento envolve a aplicação de critérios de verdade, independentes de partes interessadas.

Finalmente, cabe notar que trata-se aqui de uma visão de mundo radicalmente diferente daquela expressa por Vakkari (1994), que parecia reinar na ciência da informação da época:

Practical objectives as well as practical problematics are typical for fields other than that of scientific inquiry. If science tries mainly to conceptualize and understand its object of interest, nonscience in this respect attempts foremost to find solutions to practical problems in everyday life. It implies that the methods used in science are scientific methods, not methods of operation in some practical field (Vakkari, 1994, p.5).

Na opinião do autor, Vakkari expressa aqui visão excessivamente positivista, que com sua envergadura de autor seminal, pode ter repercutido certa estagnação da ciência da informação; especialmente de suas linhas de pesquisa em informação e tecnologia e de gestão da informação e do conhecimento. Com efeito, enquanto linhas de pesquisa de uma ciência social aplicada, estas são inexoravelmente caracterizadas pela nidificação de problemas práticos e questões teóricas.

6 EXEMPLO ILUSTRATIVO DA APLICAÇÃO DA DSR EM UM PROJETO DE PD&I

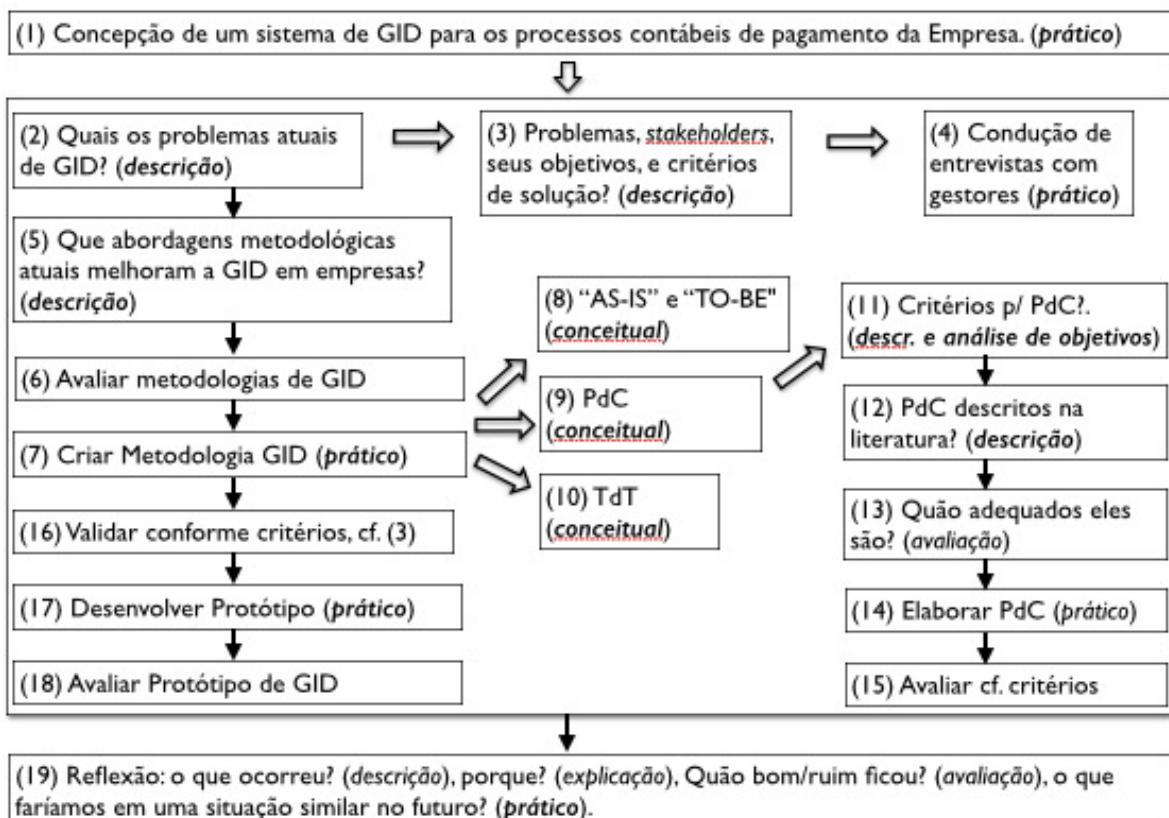
A FIG. 5 mostra a estrutura aninhada⁴⁸ característica de um projeto de PD&I em empresa do setor elétrico, coordenado pelo autor. O quadro maior apresenta os desdobramentos do problema prático, de concepção de uma solução de Gestão Integrada de

⁴⁸ Aninhamento entre problemas práticos e questões teóricas, revelado por Wieringa (2009).

Documentos (GID) para um determinado contexto organizacional. O problema é como deve-se conceber um sistema de Gestão Integrada de Documentos (GID⁴⁹) para gerenciar os registros (no sentido de documentos administrativos) envolvidos nos processos contábeis da Empresa. O sistema deve além de reduzir os custos operacionais, respeitar toda a legislação pertinente e garantir a segurança da informação. À luz dos conceitos da DSR apresentados, o problema de pesquisa caracteriza-se como um problema prático que, contudo, desmembra-se em vários subproblemas, que são eles mesmos de prática e teórica (de conhecimento): problemas práticos ou questões de conhecimento.

A decomposição, apenas em primeiro nível, do problema pode ser vista na FIG. 5. Cada subproblemas ou atividade, representada por um retângulo, é caracterizada na figura como uma atividade de natureza prática ou teórica. Os termos “descrição”, “análise conceitual”, “avaliação”, “prescrição”, “explicação”, denotam tipos de questões de conhecimento, conforme classificação inspirada de Wieringa (2009).

FIGURA 5 – Desdobramentos teóricos e práticos de um projeto PD&I



Fonte: o autor, inspirado de Wieringa (2009).

⁴⁹ “Integrada” aqui refere-se ao gerenciamento misto dos documentos em papel e formatos eletrônicos.

Conforme prega a DSR deve-se notar que em paralelo com o ciclo empírico existe um outro ciclo de engenharia (*cf.* FIG 4), encarregado de especificar os requisitos do sistema GID conforme os critérios estabelecidos em (3), desenvolver um protótipo (17), validá-lo (18) e em seguida avaliá-lo em condições de prática cada vez mais realistas (19). Avaliá-lo significa descrever o que ocorreu, encontrar razões, descrever os ajustes necessários. Observar que, na figura, as setas grossas indicam a decomposição das tarefas, as finas indicam a sua sequência.

Deve-se notar que os quadros-atividades (2 e 3) representam questões de conhecimento, seguidos de um problema prático (4). Nos quadros 2 e 3 descreve-se em detalhes quais são os problemas atuais com a GID nos processos contábeis de pagamento da Empresa. Levanta-se os problemas, os *stakeholders*, seus objetivos, e seus critérios de solução. O quadro-atividade (5) levanta os conceitos básicos oficiais (do Conarq, p.ex.) da disciplina de GID, tais como: documento de arquivo, documento eletrônico, documento arquivístico eletrônico, gestão de documentos arquivísticos (eletrônicos ou não) e muitos outros. Além disso é preciso conhecer qual é a legislação nacional (nos três níveis da federação) que regulamenta a prática da GID em organizações de capital misto (tipo de capital da organização estudada). Deve-se buscar conhecer também diversas outras questões como: se existem, quais são os princípios gerais internacionais (recomendados para a GID)? quais são as abordagens usadas hoje para melhorar o gerenciamento de documentos em grandes empresas? como levantar o acervo documental envolvido nos processos contábeis da Empresa? por meio da modelagem de processos (modelos “*as-is*” e “*to-be*”) ou outro? como conduzir oficinas de modelagem de processos? como elaborar um plano de classificação (PdC) e uma tabela de temporalidade (TdT) para esses documentos? como validar todos esses artefatos junto aos diferentes *stakeholders*?

7 CONCLUSÃO

Neste ensaio apresentou-se a *design science research* (DSR) e discutiu-se algumas de suas características mais relevantes para a sua aplicação em pesquisas no campo da ciência da informação. Procurou-se mostrar que a DRS é adequada como metodologia para orientar pesquisas nesse campo, mais especificamente na subárea da informação e tecnologia, abarcada pelo GT8 do Enancib.

O autor procurou defender ao longo do artigo que a DSR constitui a base epistêmica que Vakkari (1994, p.50) adverte faltar ao paradigma gerencial da ciência da informação. Portanto, munido de uma filosofia da ciência explicitada e cada vez mais bem estruturada e fundamentada, o paradigma gerencial passa a constituir, juntamente com o paradigma

cognitivo, uma parte muito expressiva da ciência da informação, conforme previa Vakkari no mesmo artigo.

Mostrou-se como na DSR é possível tratar um problema de pesquisa sem perder de vista a consideração de sua relevância e aplicabilidade, nem o rigor científico necessário à geração de conhecimento novo. Ou seja, com a DSR o pesquisador está munido de métodos que exploram os aspectos teóricos e práticos do seu problema de pesquisa, sempre aninhados entre si. Argumentou-se, contudo, corroborando Wieringa (2009), que a confusão dos dois pode levar a problemas metodológicos que inibem a qualidade e os resultados das pesquisas.

Na mesma linha de raciocínio elaborada em Wieringa (2009), ilustrou-se através de um exemplo prático, como a DSR guia a estruturação de um problema de pesquisa em subproblemas menores. Interessa observar como tal estruturação revela o aninhamento entre problemas práticos e questões de conhecimento; o que permite a orientação mais precisa na escolha dos métodos mais adequados de pesquisa para cada tipo de problema: sejam problemas do ciclo de engenharia ou do ciclo empírico, *cf. FIG. 4.*

O próximo passo da pesquisa relatada aqui será estabelecer um quadro teórico estruturado, baseado na DSR, que possa ser utilizado para guiar metodologicamente as pesquisas na área. O quadro deverá originar também critérios de avaliação da robustez metodológica e do rigor científico das pesquisas.

Outra iniciativa futura, que julga-se de forte interesse epistêmico para o desenvolvimento da ciência da informação, é a realização de um estudo crítico longitudinal (à luz do quadro teórico estabelecido anteriormente e de seus critérios de avaliação) dos artigos publicados no GT8. Acredita-se que ao explicitar a ontologia da DSR, bem como seus mecanismos epistemológicos, metodológicos e até axiológicos e usar tal quadro para realizar o auto-escrutínio das pesquisas publicadas, colabora-se significativamente para o amadurecimento da ciência da informação. Assim, pesquisa similar à que realizaram Wieringa e Heerkens (2006), analisando o corpus de artigos de periódicos do tema engenharia de *software* também ficará para um futuro artigo.

O autor concorda com Simon (1996) que a DS está em um constante estado de revolução científica (Kuhn, 1996), pois os avanços tecnológicos resultam de investigação científica criativa e *design* inovador, mesmo que realizados de forma inconsciente pelos pesquisadores. Para não serem irrelevantes nem arbitrárias essas pesquisas precisam estar atentas ao conhecimento pré-existente e às necessidades de negócios das organizações. Inovações, como as geradas na atual sociedade da informação tiveram impactos dramáticos

e, por vezes, não intencionais sobre a maneira como os sistemas de informação são concebidos, projetados, implementados e geridos. Consequentemente, as diretrizes da DSR estão, necessariamente, em constante adaptação.

REFERÊNCIAS

- FULLER, R; MCHALE, J. World Design Science Decade, 1965-1975. **World Resources Inventory**. Southern Illinois University, USA, 1965.
- GREGG, D. G.; KULKARNI, U. R.; VINZÉ, A. S. Understanding the philosophical underpinnings of software engineering research in information systems, **Information Systems Frontiers** v. 3, n. 2, p. 169-183, 2001.
- GREGOR, S.; HEVNER, A. R. Positioning and presenting design science research for maximum impact, **MIS Quarterly**, v. 37, n. 2, p. 337-356, 2013.
- GREGORY, S.A. **The Design Method**. Butterworths. 1966.
- HEVNER, A.R. A three cycle view of design science research, **Scandinavian journal of information systems**, v. 19, n. 2, p. 4, 2007.
- HEVNER, A.R.; CHATTERJEE, S. **Design science research in information systems**, Springer, 2010.
- HEVNER, A.R.; MARCH, S.T.; PARK, J.; RAM, S. Design science in information systems research. **MIS Quarterly**, v. 28, n. 1, p. 75-105, 2004.
- IIVARI, J. A paradigmatic analysis of information systems as a design science. **Scandinavian Journal of Information Systems**, v. 19, n. 2, p. 5, 2007.
- KUECHLER, W.; VAISHNAVI, V. The emergence of design research in information systems in North America. **Journal of Design Research**, v. 7, n. 1, p. 1-16, 2008.
- KUHN, T. **The Structure of Scientific Revolutions**. Chicago, Univ. of Chicago Press. 1996.
- LACERDA, D. P.; DRESCH, A.; PROENÇA, A; JÚNIOR, J. Design Science Research: método de pesquisa para a engenharia de produção. **Gest. Prod.**, São Carlos, v. 20, n. 4, p. 741-761, 2013.
- MARCH, S. T. and SMITH, G. F. Design and natural science research on information technology. **Decision support systems**, v. 15, n. 4, p. 251-266, 1995.
- NUNAMAKER Jr, J. F.; CHEN, M. Systems development in information systems research. **System Sciences, Proceedings of the Twenty-Third Annual Hawaii International Conference**, IEEE, p. 631-640, 1990.
- SIMON, Herbet. **The Sciences of the Artificial**. MIT Press. ISBN 0-262-69191-4. 1996.
- TAKAHASHI, Ricardo H. C. **A estrutura do conhecimento tecnológico do tipo científico**. Belo Horizonte: UFMG, 2009. 213p.
- VAISHNAVI, V.; KUECHLER, W. **Design Research in Information Systems**. 2009. Disp. em: <desrist.org/design-research-in-information-systems>. Acesso em: 18 jul. 2014.

VAKKARI, P. Library and Information Science: Its Content and Scope. In: GODDEN IRENE, P. (Org.). **Advances in librarianship**. San Diego, 1994.

VISSE, W., **The Cognitive Artifacts of Designing**, Hillsdale, NJ, USA, Lawrence Erlbaum Associates, 2006.

WIERINGA, R. Case Study Research in Information Systems Engineering. **Anais eletrônicos...** CAiSE, 2013. Tutorial. Disponível em: <<http://wwwhome.cs.utwente.nl/~roelw/case-studies-CAiSE-2013-Wieringa.pdf>>. Acesso em: 29 set. 2014.

WIERINGA, R. Design science as nested problem solving, Proceedings of the 4th int. conf. on design science research in information systems and technology, ACM, p. 8, 2009.

WIERINGA, R.; HEERKENS, J. The methodological soundness of requirements engineering papers: a conceptual framework and two case studies. **Req. Eng.** v. 11, n. 4, p. 295-307, 2006.

LIVROS DIDÁTICOS DIGITAIS EM AMBIENTES COMPUTACIONAIS NAS NUVENS

DIGITAL TEXTBOOKS IN ENVIRONMENTS CLOUDS COMPUTER

Charles Rodrigues
Angel Freddy Godoy Viera

Resumo: Apresenta um estudo sobre os benefícios e limitações do uso dos livros didáticos digitais em ambientes computacionais nas nuvens. Objetiva descrever e discutir o uso dos livros didáticos digitais em ambientes computacionais nas nuvens apresentando algumas experiências recentes. Trata-se de uma pesquisa bibliográfica, descritiva e analítica caracterizando-se no conjunto uma pesquisa qualitativa. Na fundamentação conceitual são abordados e discutidos os conceitos e características de livros didáticos digitais e de ambiente computacional nas nuvens e as experiências do uso dos livros didáticos digitais, em ambientes computacionais nas nuvens nos Estados Unidos, Coreia do Sul e no Brasil. Ainda foi concebido um *framework* dos benefícios funcionais e das limitações, este dividido em seis principais capacidades desejáveis em uma plataforma tecnológica de recursos educacionais centrados em livros didáticos digitais. Nas considerações constatou-se que o uso dos livros didáticos digitais em ambientes computacionais nas nuvens seja uma tendência, sobretudo, considerando-se: o número de conexões com a internet e dos dispositivos eletrônicos móveis, as novas possibilidades de uso das tecnologias da informação e comunicação para o desenvolvimento de novos produtos, conteúdos e aplicativos com foco na educação, que poderão intensificar ainda mais *mobile learning*. No entanto, por ser um fenômeno hodierno e em fase de aperfeiçoamento o uso dos livros didáticos digitais nas escolas enfrenta alguns entraves que necessitam mais aprofundamento e discussões, alguns de viés tecnológico, outros de questões de políticas públicas e ou de natureza legal.

Palavras-chave: Livros Didáticos Digitais. Ambientes Computacionais nas Nuvens. Tecnologias Educacionais. *Mobile Learning*.

Abstract: Presents a study on the benefits and limitations of the use of digital textbooks in environments cloud computational. Aims to describe and discuss the digital textbooks in environments cloud computational presenting some recent experiences in using this feature. This is a research bibliographic, descriptive and analytic in the set featuring a qualitative research. In conceptual basis are addressed and discussed the concepts and features of digital textbooks and environment clouds computing and the experiences of the use of digital textbooks in the environments cloud computing in the United States, South Korea and Brazil. Yet conceived a framework of functional benefits and limitations, this is divided into six main desirable skills in technological platform educational resources centered on digital textbooks. In the considerations it was found that the use of digital textbooks in environment clouds computational is a tendency, particularly considering: the number of internet connections and mobile electronic devices, the new possibilities of the use of information technologies and communication with the development of new products, content and applications with a focus on education may further intensify mobile learning. However, being a meeting today and phase enhancement phenomenon the use of digital textbooks in schools faces some obstacles that need more approaches and discussions, some technological bias, other public policy issues and other legal matters.

Keywords: Digital Textbooks. Environments Cloud Computing. Educational Technologies. *Mobile Learning*.

1 INTRODUÇÃO

O desenvolvimento dos livros didáticos em formato digital como material educacional e pedagógico constituinte do processo de ensino e aprendizagem destaca-se como relevante quando se observa o número de conexões móveis no mundo, em torno de sete bilhões, cento e cinco milhões em junho de 2014 (GSMA INTELLIGENCE, 2014). Adicionado a este número de conexões, a conversão por meio de digitalização de livros impressos e outros que já nasceram em formato digital, e paralelamente o aumento do número de usuários “nativos digitais”, isto, de forma progressiva poderá provocar, num futuro próximo, a eliminação dos livros didáticos em formato impresso.

A UNESCO (2013) em seu relatório sobre o futuro da *mobile learning* apresenta alguns avanços tecnológicos com maior probabilidade de impactar o processo de ensino e aprendizagem numa perspectiva global, a saber: 1) a *tecnologia será mais acessível, menor custo e funcional*, a evolução tecnológica melhorará a funcionalidade, a conectividade e a diminuição dos custos financeiros. A maior disponibilidade e penetração dos dispositivos móveis inteligentes com serviços baseados nas nuvens irão proporcionar possibilidades para a *mobile learning* em larga escala; 2) os *dispositivos serão capazes de coletar, sintetizar e analisar grandes quantidades de dados*, estes dados estão sendo coletados de diversas formas: desde os dados pessoais aos dados institucionais, estas informações estão sendo registradas pelos provedores de serviços nas nuvens. Na área educacional este conjunto de dados é proveniente de diversas fontes, como: perfis de trabalhos dos alunos; resultados de avaliação; registros de frequência; o tempo gasto em tarefas ou atribuições específicas; e informações produzidas ou utilizadas pelos alunos, incluindo textos, imagens, vídeos ou músicas. No futuro, os dispositivos ligados à nuvem terão a capacidade de sintetizar maiores quantidades de dados e poderão analisá-las por padrões; 3) *novos tipos de dados estarão disponíveis*, muitos dispositivos móveis já têm diferentes tipos de sensores que podem, por exemplo, captar o som através de um microfone, localizar por meio de um GPS, e sentir o movimento, a velocidade e a direção através de um acelerômetro. Atualmente estes sensores permitem analisar padrões de sono e monitorar os sinais vitais, como por exemplo, os batimentos cardíacos. Novos tipos de dados, combinados com os dados coletados por meio da tecnologia móvel, possibilitarão relações mais estreitas entre os alunos e os seus dispositivos. Esses irão "conhecer" as preferências de seus donos e estarão intimamente e inteligentemente conectados a eles, proporcionando uma aprendizagem mais personalizada e contextual; 4) *barreiras linguísticas serão quebradas*, os avanços recentes na área de processamento da linguagem natural, permitem o reconhecimento de linguagem

natural em dispositivos móveis e são capazes de traduzirem sem problemas, a linguagem falada, com um elevado grau de precisão. Assim, os alunos terão acesso a uma gama muito maior de recursos e conteúdos educacionais em diferentes línguas; 5) as *limitações de tamanho da tela irão desaparecer*, o tamanho da tela dos dispositivos serão diferentes, por exemplo, as tecnologias de visualização flexíveis permitirão que os usuários possam dobrar o tamanho de uma tela de um *tablet* ou *smartphone* transmitindo o conteúdo para visualização em um televisor de tamanho maior ou rolar uma grande tela em um pequeno cilindro de fácil portabilidade. Projetores ou óculos, como os óculos de realidade aumentada, serão capazes de exibir telas muito maiores do que aquelas fisicamente disponíveis em um dispositivo móvel, com excelente resolução e com capacidade em 3D; e 6) as *fontes e a capacidade de armazenamento de energia vão melhorar*, as limitações e capacidade de armazenamento podem ser barreiras para programas de *mobile learning*. Melhorias consideráveis serão feitas, como baterias menores, mais baratas, mais duradouras, mais eficientes e recarga mais rápida.

Em decorrência do desenvolvimento das tecnologias de informação e comunicação, as formas como as pessoas interagem e usam essas tecnologias vem alterando o ambiente educacional. A aplicação destas à educação pode melhorar os métodos didáticos e pedagógicos por meio de várias funções que antes não estavam disponíveis nas salas de aulas tradicionais. Pode, por exemplo, permitir aos alunos personalizarem experiências de aprendizagem, conforme, suas capacidades e seus interesses.

Em contextos de educação formal no mundo desenvolvido, a transição para livros didáticos digitais é uma das tendências mais estabelecidas em *mobile learning*. Como os dispositivos eletrônicos móveis e os aplicativos de leitura eletrônica continuam a melhorar, a experiência de leitura está se tornando mais agradável e propício para a aprendizagem. Novas abordagens para a conversão e criação de livros didáticos estão se afastando de meras reproduções digitais de textos impressos para interfaces visualmente mais ricas que podem incluir multimídia, elementos interativos e colaborativos (GSMA, 2011).

Há muitos indícios de que o uso da tecnologia pode melhorar a oportunidade para o acesso educacional, aumentar o envolvimento e o resultado dos alunos, e aprimorar a produtividade de aprendizagem. De acordo com o Departamento de Educação dos Estados Unidos e recentes estudos realizados pela Associação Nacional de Treinamento e Simulação do mesmo país, a instrução baseada em tecnologia pode reduzir de 30 a 80 por cento, o tempo necessário para os alunos atingirem um objetivo de aprendizagem. Em média, os estudantes que receberam aprendizagem *online* tiveram desempenho melhor do

que aqueles que receberam instrução face a face (FEDERAL COMMUNICATIONS COMMISSION, 2012).

É importante ressaltar, entretanto, que esses avanços tecnológicos não ocorrem de forma equânime entre os países do mundo e nem dentro dos países com alto nível de inovações tecnológicas. Como os países e regiões de todo o mundo têm políticas diferentes em relação aos livros digitais, para estabelecer padrões para a sua implementação em salas de aulas é necessário considerar que, não só os livros digitais diferem entre os contextos culturais, mas os impactos causados a este ambiente também são diferentes. Os livros didáticos digitais não só irão fornecer acessos aos livros impressos digitalizados, mas também, complementar o conteúdo educacional existente com recursos e ferramentas adicionais para a aprendizagem.

O ambiente computacional nas “nuvens” pode viabilizar o acesso a conteúdos pedagógicos, a livros didáticos digitais, além de outros recursos *online* de alta qualidade a alunos, corpo docente e funcionários da escola, em qualquer lugar, a qualquer hora, podendo promover aulas *online* à distância. As aulas presenciais também são beneficiadas na medida em que há possibilidades de acesso e compartilhamento ao amplo acervo de informação multimídia *online*, por meio de conexão com a Internet. Isso demarca as vantagens do ambiente computacional nas “nuvens” que possibilita que os conteúdos pedagógicos e didáticos não sejam armazenados apenas nos arquivos locais dos dispositivos, mas que possam ser acessados em múltiplos dispositivos eletrônicos, e permite que os padrões de uso de diversos usuários possam ser registrados para sugerir conteúdos a grupos de usuários com perfil de interesse similar baseado na experiência de outros usuários.

Assim este estudo objetiva descrever e discutir os livros didáticos digitais em ambientes computacionais nas nuvens apresentando algumas experiências recentes no uso deste recurso.

2 METODOLOGIA

O conteúdo apresentado neste artigo é parte dos estudos realizados no contexto do grupo de pesquisa Recuperação de Informação e Tecnologias Avançadas (RITA) e integrante dos fundamentos conceituais que estão sendo desenvolvidos para tese de doutoramento sobre a temática, no programa de pós-graduação em Ciência da Informação da Universidade Federal de Santa Catarina (PGCIn/UFSC). Trata-se do resultado decorrente de uma pesquisa bibliográfica, descritiva e analítica caracterizando no conjunto uma pesquisa de base qualitativa.

A pesquisa bibliográfica foi realizada para: conhecer quem está estudando o tema livros didáticos digitais, onde (instituições e regiões geográficas) e sob quais abordagens; conhecer os periódicos que se dedicam ao assunto; a produção científica disponível; verificar que novas abordagens poderão ser realizadas; e a construção de argumentos que poderão ser utilizados na fundamentação conceitual desse estudo (SOUZA, 1997).

Este estudo se caracteriza como descritivo e analítico, pois, visa delinear a natureza do objeto em estudo, no caso, livros didáticos digitais em ambientes computacionais nas nuvens apontando uma análise crítica das características, das correlações e descrição dos fatos e fenômenos observados, sem interferência no mesmo.

Pela situação e característica do objeto de estudo foi utilizado uma abordagem qualitativa. Neste caso, tentou-se descrever e discutir um processo recente e dinâmico, experimentado de um objeto social em constante mudança.

Durante o estudo foram desenvolvidas as seguintes etapas: (1) pesquisa bibliográfica, na base de dados do portal de periódicos da CAPES para identificação e reconhecimento inicial do fenômeno e das abordagens já realizadas; (2) construção de uma revisão de literatura (fundamentação conceitual), com desenvolvimento de estudos conceituais para compreender o objeto de estudo em seu contexto cultural; (3) seleção e interpretação de amostra da realidade observada; e (4) utilização da fundamentação conceitual para interpretar as informações coletadas na realidade observada.

3 FUNDAMENTAÇÃO CONCEITUAL

São apresentados e brevemente discutidos os conceitos relacionados à temática em estudo, com referências aos fundamentos, aos benefícios e limitações dos livros didáticos digitais em ambientes computacionais nas nuvens e as experiências em alguns países.

3.1 Livros Didáticos Digitais

O paradigma da educação está se movendo em direção ao *e-Learning*, o que exige a participação ativa dos alunos, a interação através da cooperação e compartilhamento de informações e o uso de recursos educacionais como os livros didáticos digitais baseados na Web 2.0.

Entende-se *e-Learning* como processo de ensino e aprendizagem apoiado por uma rede de computadores permitindo acesso a um conjunto de diferentes recursos didático-pedagógicos centrados no aluno. Nesse tipo de aprendizagem o ritmo de assimilação de conhecimento, o horário, e o local de acesso ao conteúdo é pautado pelo aluno. Importante

destacar, que o professor não será descartado, mas, seu papel será transformado, em mediador entre o conhecimento disponível e o aluno.

Neste contexto de mudança do ambiente educacional Jung (2009) conceitua livro didático digital como

um livro centrado no aluno que se destina a promover vários tipos de interações, e que permite ao aluno trabalhar de acordo com as suas necessidades e nível, seja na escola ou em casa, sem os limites de tempo e espaço, e que integra todo o conteúdo existente dos livros didáticos em papel, referências, livros de tarefas, e dicionários, e fornece um ambiente de aprendizagem multimídia que inclui vídeos, animações e realidade virtual (JUNG, 2009, p.4).

Além disso, oferece as funções de busca na Internet, tornando o ambiente de aprendizagem mais animado e orientado aos interesses dos alunos. Possibilitando ainda, a interação dos alunos com outros alunos, sem importar sua localização geográfica, por meio do uso das redes sociais como elemento fomentador do processo de ensino e aprendizagem colaborativo.

O livro didático digital é um livro voltado para o futuro, que pode ajudar os alunos a aprender de forma autodirigida a qualquer hora e em qualquer lugar. O livro didático digital é adequado para os nativos digitais que estão acostumados com o ambiente digital, desde a infância. Este permite que os alunos tenham não só o conteúdo de livro didático geral, mas também vários materiais de aprendizagem personalizados, com funções interativas que permitem aos alunos aprender de acordo com as suas aptidões, habilidades e níveis (JANG, 2014).

O uso dos livros didáticos digitais irá exigir de todos envolvidos (professores, tecnólogos, pais e até mesmo os decisores políticos) a pensar em como adequar o seu uso em sala de aula, uma vez que, num futuro próximo ele não irá substituir os livros impressos, mas sim, complementá-los, principalmente em regiões com baixa infraestrutura tecnológica digital. Nas salas de aula, professores e alunos deverão começar a valorizar a conveniência e acessibilidade desses recursos. Os tecnólogos devem trabalhar para expansão do uso através da criação de consciência da usabilidade dos livros didáticos digitais em diversos tipos de dispositivos móveis. Com isso, os pais serão expostos as mais recentes tecnologias de educação desenvolvidas, e poderão participar mais ativamente no processo de ensino e aprendizagem (EMBONG *et al.*, 2012).

Ao longo dos anos pode-se verificar que as tecnologias de informação e comunicação tornaram-se um ponto relevante entre os educadores como ferramentas eficazes tanto no processo de ensino e aprendizagem como de gestão dos recursos educacionais. Logicamente, que seu uso e perspectivas são bastante recentes, no entanto, a

sua utilização apresenta algumas vantagens que representam benefícios para os alunos, pais e gestores escolares.

Para os alunos os benefícios apresentados são: menor peso a ser carregado, pois, possibilita compactar em um único dispositivo eletrônico, uma grande quantidade de livros impressos; a possibilidade do processo de aprender tornar-se mais atraente, com a utilização de interface gráfica mais intuitiva e colorida, em muitos casos até personalizáveis, com uso de imagens, sons e vídeos, utilização de recursos para aumentar ou diminuir o tipo de letra, uso de *hiperlinks* para acessar mais conteúdo sobre um tema de pesquisa; também, permitir aos alunos personalizar sua experiência de aprendizagem conforme seus interesses e habilidades.

Para os pais uma das vantagens de uso dos livros didáticos digitais é a questão financeira, que com a produção em larga escala tornará os custos de aquisição ainda menores, e permitirá que os autores atualizem os conteúdos dos livros didáticos de forma mais fácil e rápida.

Para os gestores escolares o uso dos livros didáticos digitais poderá facilitar o processo de gestão escolar, uma vez que, permite monitorar, registrar e documentar o conteúdo acessado e estudado pelos alunos, contribuindo para o planejamento das aulas.

Por ser um fenômeno hodierno o uso dos livros didáticos digitais nas escolas ainda enfrenta algumas limitações, de caráter tecnológico, de questões políticas e culturais. Dentre outras, ressaltamos: os educadores e gestores precisam ser capacitados adequadamente para conduzir as aulas com os livros didáticos digitais; há a necessidade de ampla infraestrutura física e de rede lógica estruturada; a limitação da capacidade de armazenamento dos leitores de conteúdos didáticos eletrônicos para acesso *offline*; o fornecimento de conteúdo eletrônico na quantidade e no modelo necessário nas escolas; para alguns alunos os livros didáticos digitais podem não oferecer o mesmo prazer e comodidade de leitura comparada a um livro impresso; a dificuldade de aquisição/compra, pois, existem diferentes modelos de negócios; e outra dificuldade encontra-se em relação aos direitos de propriedade intelectual, devido, ao rigor do *Digital Rights Management* (DRM) implementados por alguns fornecedores de conteúdos.

Nas subseções seguintes são descritos o uso, os benefícios e limitações e experiências de alguns países no uso dos ambientes computacionais nas nuvens para suportar os livros didáticos digitais.

3.1.1 Livros Didáticos Digitais em Ambientes Computacionais nas Nuvens

A utilização dos ambientes computacionais nas nuvens como suporte para os recursos educacionais ainda são incipientes, em grande parte devido à falta de recursos para reforçar a infraestrutura física e tecnológica necessária para o desenvolvimento deste ambiente. No entanto, pelas suas características, benefícios e observando o amplo avanço do uso da tecnologia nas nuvens em outros setores da sociedade, o seu uso na educação provavelmente será ampliado.

Os dispositivos eletrônicos e/ou móveis, os livros didáticos digitais e outros conteúdos educacionais em formato digital viabilizam a promoção de aulas *online* à distância. Os benefícios do acesso via ambiente nas nuvens promovem e estimulam a ampliação e desenvolvimento de acervo de diferentes conteúdos armazenado nesse ambiente. Uma das vantagens desse ambiente é permitir que os conteúdos não sejam armazenados apenas nos arquivos locais dos dispositivos, mas também, possam estar acessíveis em múltiplos dispositivos e de forma remota.

O Instituto Nacional de Padrões e Tecnologia (NIST) do Departamento de Comércio norte-americano, em 2011 definiu “Computação nas nuvens” como

um modelo que permite acesso ubíquo, conveniente e sob demanda via rede a um agrupamento compartilhado e configurável de recursos computacionais (por exemplo, redes, servidores, equipamentos de armazenamento, aplicações e serviços), que pode ser fornecido e liberado com esforços mínimos de gerenciamento ou interação com o provedor de serviços (JULA; *et al.*, 2014, p. 3.810-3.811).

É uma plataforma de aplicação emergente e tem como objetivo hospedar, comunicar, compartilhar dados, conteúdos, cálculos e serviços entre os usuários.

As características essenciais do ambiente de computação nas nuvens para os recursos educacionais podem-se generalizar em: i) *autosserviço sob demanda*: o aluno deve ser capaz de alocar novos recursos automaticamente, sem interação humana com o provedor de serviços, ter acesso aos conteúdos educativos e as aulas com autonomia; ii) *acesso amplo via rede*: os recursos devem estar disponíveis por meio da rede e devem ser acessíveis por estrutura padrão, permitindo o uso por diferentes dispositivos, tais como computadores pessoais, *smartphones*, *tablets*, dentre outros; iii) *agrupamento de recursos*: os recursos computacionais do provedor de serviços devem ser agrupados para servir a múltiplos consumidores (alunos, corpo docente e gestores escolares), com recursos físicos e virtuais sendo arranjados e rearranjados dinamicamente conforme a demanda desses consumidores; iv) *elasticidade rápida*: os recursos devem ser alocados e liberados de forma elástica e em alguns casos automática, permitindo a rápida adaptação à demanda. Para o ambiente

educacional, é importante que os recursos estejam disponíveis em quantidades desejadas a qualquer momento, isto é, possibilitar a customização, conforme, o perfil do aluno, dos professores e das necessidades pontuais que ocorrerem; v) *serviços mensurados*: os serviços de computação nas nuvens devem controlar e otimizar os recursos de maneira automática. Deve ser possível monitorar, controlar e consultar o uso dos recursos, provendo transparência para o ambiente educacional e para o provedor dos serviços (JULA; *et al.*, 2014).

3.1.1.1 Benefícios e Limitações dos Livros Didáticos Digitais em Ambientes Computacionais nas Nuvens

As tecnologias são sustentadas e desenvolvidas num processo de construção e aperfeiçoamento contínuo, assim, também, ocorre com os livros didáticos digitais, os quais já apresentam vantajosos benefícios para o ambiente educacional, no entanto, ainda enfrentam enormes desafios e limitações em sua adoção, expansão e desenvolvimento.

Com base nas informações e argumentos apresentados por Ash (2011), Embong; *et al.* (2012), Jang (2014) construiu-se um *framework* dos benefícios funcionais e das limitações da postagem de livros didáticos digitais em ambientes computacionais nas nuvens. Este dividido em seis principais capacidades desejáveis em uma plataforma tecnológica de recursos educacionais centrados em livros didáticos digitais.

1) *Dispositivos de aprendizagem personalizável e interativa*

- Benefícios – podem oferecer lições individualizadas e personalizadas ao estilo e nível do aluno, permitir a retroalimentação em tempo real aos alunos, pais, professores ou tutores. Podem apresentar variados tipos de materiais auditivos ou visuais, incluindo voz, texto, música, animações em 3D, fotografias ou vídeos, isoladamente ou em diferentes combinações. Com *hiperlinks* a diversos conteúdos vinculados por meio da web para facilitar a aprendizagem autodirigida. Podem também, proporcionar flexibilidade no uso, permitindo aos alunos definir a velocidade da fala, decidir se o texto escrito também deve ser lido em voz alta, escolher o idioma apresentado no texto, ou repetir a apresentação em quantidade de vezes que o aluno quiser. A plataforma deve permitir aos alunos iniciarem discussões com os outros sobre conteúdos e atividades, de forma interativa. Pode possibilitar a pesquisa de mais informações sobre a temática da aula em outros livros didáticos digitais ou em outros repositórios educacionais. Como exemplos desses repositórios, pode-se citar: o Portal do Professor do Ministério da Educação (MEC) do Brasil, Banco Internacional de Objetos Educacionais, *OpenCourseWare* do Instituto de Tecnologia de Massachusetts, MERLOT - Multimedia Educational Resources for Learning

and Online Teaching e CK - 12 do Estado da Califórnia, nos Estados Unidos e KERIS da Coréia do Sul.

- Limitações – é necessária a formação e treinamento dos alunos, do corpo docente e funcionários no uso desta nova tecnologia. Definir que tipo de treinamento e suporte para promover a aceitação de novos modelos de ensino e aprendizagem. Também, é necessário criar estratégias para acelerar a criação de módulos interativos e personalizados de conteúdos de aprendizagem digital. É importante o desenvolvimento de metodologias para avaliar a eficiência desses recursos. E outra questão, a se refletir, é sobre a adesão a esses recursos, pois, para alguns alunos os livros didáticos digitais podem não oferecer o mesmo prazer de leitura em comparação com a leitura de um livro impresso.

2) Função de gestão de aprendizagem e ferramenta de avaliação

- Benefícios – conexão com ferramentas de avaliação dentro e fora do ambiente do livro didático digital; o sistema de gestão de aprendizagem deve permitir aos professores compreender os pontos fortes e fracos dos alunos, criando um programa de plano de estudo personalizado, para que eles possam trabalhar as habilidades em que possuem mais dificuldades. Outro benefício é a capacidade de apresentar informações e atividades em vários formatos, isto, significa que a plataforma tecnológica pode aceitar uma variedade de solicitação de formulário de entradas de dados. Dentre outras funcionalidades pode ser programado para verificar o trabalho de um aluno. Sendo altamente capaz de gravar e organizar informações e disponibilizar em vários formatos. Pode, por exemplo, gravar as respostas de todos os alunos de uma classe e, em seguida, comunicar imediatamente ao professor os erros cometidos por cada aluno, bem como os erros comuns de toda a classe. Em tarefas mais complexas, pode servir como dispositivo de registro e comunicação para os professores, ajudando a acompanhar o progresso dos alunos de maneira mais conveniente, permitindo que os professores criem conteúdos para reforçar as temáticas em que os alunos têm maior dificuldade.

- Limitações – também é necessária a formação e treinamento dos alunos, do corpo docente e funcionários no uso desta nova tecnologia. Além disso, outras questões merecem mais aprofundamento, como: deliberar sobre o desenvolvimento de modelos de ensino e aprendizagem; definir o papel do professor na sala de aula com a utilização dos livros didáticos digitais; e decidir até que ponto, os alunos terão a liberdade para direcionar suas próprias lições. E outro tema se refere à eficiência dos livros didáticos digitais, pois, são necessários estudos consistentes para verificar sua contribuição no desempenho dos alunos no processo de aprendizagem. Em relação às avaliações é necessário o

desenvolvimento de tecnologias que permitam a verificação da identidade do aluno que está desenvolvendo as atividades e avaliações propostas nos livros didáticos nas nuvens.

3) Melhoria da saúde dos alunos

- Benefícios – pode promover melhorias, tanto físicas como psicológicas. Um dispositivo eletrônico pode comprimir o conteúdo de vários livros didáticos impressos, isto, reduz o peso da carga que os alunos precisam carregar diariamente. Esta redução de peso, fisicamente, permite que os alunos cresçam de forma mais saudável, sem efeitos prejudiciais, tais como dores nas costas, má postura, deformidade da coluna vertebral e problemas de coluna na idade adulta. Psicologicamente, os alunos que se envolvem com os livros didáticos digitais podem achar o processo divertido devido às características atraentes do suporte (funções de fácil utilização; gráficos coloridos; possibilidade de modificar o tamanho do texto; plug-ins falantes, dentre outras). Estas características podem encorajar e motivar a criatividade e autonomia de aprendizagem dos alunos.

- Limitações – o uso intensivo de tecnologia pode provocar tendência à diminuição da interação social dos estudantes e provocar problemas de saúde como, a Síndrome VTD (Visual Display Terminal); a Insônia em adolescentes, cansaço ocular, tendinites e contraturas e ganho de peso por inatividade.

4) Processo de gestão escolar

- Benefícios – facilita o processo de gestão, na elaboração e execução da proposta pedagógica, na administração de pessoal e dos recursos materiais e financeiros, no monitoramento do ensino e aprendizagem, e no fortalecimento da comunicação entre professores, alunos, escola, pais e comunidade, proporcionando assim, uma maior integração. Ao permitir monitorar as atividades realizadas por cada aluno em sala de aula física e virtual, é possível registrar, documentar, categorizar, criar e acessar diversos relatórios. Assim, os professores e gestores podem tomar melhores decisões para aprimorar os métodos de ensino e aperfeiçoar o processo de gestão e de aprendizagem dos alunos em sala de aula.

- Limitações – como em outras capacidades é necessária a formação e treinamento dos alunos, do corpo docente e funcionários no uso desta nova tecnologia. O armazenamento de conteúdo em ambiente nas nuvens aumenta o risco de comprometimento da privacidade se comparado ao armazenamento *offline*. As questões que envolvem as estruturas físicas das salas de aulas presenciais, como mesas e cadeiras adequadas para o uso de dispositivos eletrônicos e o número de tomadas de energia em cada sala de aula para garantir o uso ininterrupto dos equipamentos são necessárias serem resolvidas. O

fornecimento insuficiente de dispositivos ou conteúdos didáticos eletrônicos em escolas, também, pode ser um entrave. Em termos de gestão de recursos materiais, as limitações impostas pelo DRM, muitas vezes impede que os livros didáticos digitais sejam deslocados de um dispositivo para outro são barreiras que podem impedir acesso a conteúdos de interesse dos alunos.

5) Economia de custos

- Benefícios - o uso em salas de aulas de livros didáticos digitais em ambientes computacionais poderá reduzir os custos com materiais pedagógicos. Os preços, geralmente, são menores do que dos livros didáticos impressos, além do mais, a vida útil do livro impresso é menor em comparação com o formato digital. Pela natureza do formato digital, a atualização de conteúdo poderá ser otimizada, ágil e de menor custo.

- Limitações – mais uma vez, em alguns casos, as limitações impostas pelo DRM podem tornar o custo de aquisição maior. Custos altos dificultam a distribuição dos conteúdos didáticos e o acesso à banda larga nas escolas e nas casas dos alunos mais carentes. Além disso, é necessário discutir novos modelos de negócio que permitam aos estudantes e professores utilizar os dispositivos e outras novas tecnologias, agora e nos futuros ciclos dos produtos. E construir meios de garantir o acesso e a disponibilidade de conteúdos a todos os alunos de uma rede de ensino.

6) Garantir a sustentabilidade dos recursos

- Benefícios – o processo de aprendizagem também deve apoiar e promover o desenvolvimento sustentável da sociedade. Os livros didáticos digitais hospedados em ambientes nas nuvens podem contribuir para este esforço por meio da maximização da disponibilidade de conhecimento, reduzindo o número de árvores derrubadas para produzir livros impressos.

- Limitações – a produção, o armazenamento e a utilização de conteúdos digitais, também, consomem muita energia. Além disso, ainda, é dependente de uma rede contínua de energia elétrica para o uso dos dispositivos eletrônicos ou para o recarregamento de suas baterias. Adiciona-se que o custo ambiental para geração de energia para os *datas centers* que viabilizam os serviços nas nuvens é enorme.

No QUADRO 1 são apresentados resumidamente os principais benefícios e limitações do uso dos livros didáticos digitais em ambientes computacionais nas nuvens.

QUADRO 1 - Benefícios e Limitações dos Livros Didáticos Digitais em Ambientes Computacionais nas Nuvens

Capacidades	Benefícios	Limitações
<i>Dispositivos de aprendizagem personalizável e interativa</i>	Lições individualizadas e personalizadas. Diversidade de materiais de aprendizagem em multimídias. Maior Interatividade.	Necessidade de: formação e treinamento; criação de módulos interativos e personalizados; novas metodologias; e estratégias de adesão.
<i>Função de gestão de aprendizagem e ferramenta de avaliação</i>	Ampliação de matérias de aprendizagem. Compreensão dos pontos fortes e fracos dos alunos. Programa de plano de estudo personalizado e individualizado. Capacidade de gerir informações sobre o aprendizado e avaliação.	Necessidade de: formação e treinamento; novos modelos de ensino e aprendizagem; definir o papel do professor; de decidir a autonomia dos alunos; de verificar a contribuição dos livros didáticos digitais no aprendizado dos alunos.
<i>Melhoria da saúde dos alunos</i>	Físicos – pode evitar dores nas costas, má postura, deformidade da coluna vertebral e problemas de coluna. Psicológicos - podem encorajar e motivar a criatividade e autonomia de aprendizagem.	Pode diminuir a interação social e provocar problemas de saúde como, a Síndrome VTD; a Insônia, cansaço ocular, tendinites e contraturas e ganho de peso por inatividade.
<i>Processo de gestão escolar</i>	Na elaboração e execução da proposta pedagógica, na administração de pessoal e dos recursos materiais e financeiros, no monitoramento do ensino e aprendizagem, e na comunicação e integração da comunidade escolar.	Necessidade de formação e treinamento. Maior risco de comprometimento da privacidade. Necessidade de infraestrutura física adequada, de rede de energia e de computadores. Limitações impostas pelo DRM.
<i>Economia de custos</i>	Redução dos custos de aquisição e de atualização dos materiais pedagógicos.	O DRM pode tornar o custo de aquisição maior. Custos altos dificultam a distribuição de materiais e o acesso à Internet de banda larga. Necessidade de novos modelos de negócio e de garantia de acesso e a disponibilidade de conteúdos.
<i>Garantir a sustentabilidade dos recursos</i>	Podem contribuir para a sustentabilidade dos recursos naturais, com a maximização da disponibilidade de conhecimento.	A produção, o armazenamento e a utilização de conteúdos digitais consumem muita energia. Dependência de rede de energia elétrica continua. O custo ambiental para geração de energia para os <i>data centers</i> .

Fonte: Dos autores (2014).

Os benefícios do uso dos livros didáticos suportados em um ambiente computacional nas nuvens são diversos, conforme descritos neste *framework*, no entanto, a operacionalização depende inicialmente de alto investimento financeiro em infraestrutura tecnológica e de redes de computadores. Como na aquisição de: computadores, acesso à

Internet de banda larga eficiente, mesas e carteiras adaptadas com tomadas elétricas e de conteúdos digitais pedagógicos. Além disso, existe a dependência permanente de rede de energia elétrica estável. Importante, também, que a plena expansão desses serviços para toda rede de ensino, depende, de políticas públicas focadas no combate à exclusão digital. Para que as atividades educativas complementares possam ocorrer nas residências e com a participação dos pais dos alunos.

Na seção seguinte serão expostas as experiências do uso dos livros didáticos digitais que estão ocorrendo em países como os Estados Unidos, Coreia do Sul e Brasil.

3.1.1.2 Uso dos Livros Didáticos Digitais em Ambientes Computacionais nas Nuvens – Experiências em Alguns Países

Em alguns países com alto nível de penetração tecnológica, o uso das tecnologias de informação e comunicação já são presentes no cotidiano dos sistemas de ensino, provocando mudanças nos hábitos e nas metodologias didáticas pedagógicas.

Nos Estados Unidos foi criado o *Digital Textbook Playbook*, um guia para ajudar os administradores e educadores na construção de experiências de aprendizagem digital. É resultado de um esforço conjunto de partes interessadas da indústria, funcionários de escolas e líderes de ONGs para incentivar a colaboração entre os atores desse ecossistema. Visando acelerar o desenvolvimento de livros didáticos digitais, para melhorar a qualidade e o nível de penetração da aprendizagem digital na educação pública.

Este programa propõe quatro etapas envolvendo gestores e educadores: i) *Fazer a transição*, ii) *Conectividade na escola*, iii) *Conectividade além da escola* e iv) *Perspectivas dos dispositivos*. Na etapa i) *Fazer a transição*, os distritos escolares necessitam de liderança compartilhada e dedicada, planejamento cuidadoso, envolvimento dos professores e da comunidade para criar um ambiente de aprendizagem digital bem-sucedido; na segunda etapa ii) *Conectividade na escola*, a chave para o fornecimento de conectividade suficiente está no estudo da estimativa atual e demanda futura do bairro, da escola e das salas de aulas. Isto irá assegurar que as escolas tenham acesso à Internet de banda larga suficiente para atender aos seus alunos, corpo docente e funcionários; na etapa iii) *Conectividade além da escola*, para realização de uma aprendizagem digital verdadeiramente ubíqua, os alunos devem ter a oportunidade de se conectar além das paredes das escolas. Isto pode ocorrer por meio da combinação de banda larga móvel, conectividade na comunidade, e acesso à banda larga em casa; e na iv) *Perspectivas dos dispositivos*, um verdadeiro livro digital é um conjunto interativo de conteúdo e ferramentas de aprendizagem acessados através de um *laptop*, *tablet*, *smartphone* ou outro dispositivo

avançado. Nesta perspectiva os dispositivos eletrônicos e a plataforma tecnológica devem ser interoperáveis para possibilitar a difusão dos livros didáticos digitais (FEDERAL COMMUNICATIONS COMMISSION, 2012).

Este guia está organizado em quatro etapas que seguem aproximadamente a cadeia cronológica necessária para a adoção dos livros didáticos digitais por uma rede de ensino. Apresentando preocupação com as questões pedagógicas, de infraestrutura física e analógica e questões sociais, como a inclusão digital. A intenção desse programa é que as articulações dessas etapas consigam atender aos desafios nos ambientes de aprendizagem digitais, tais como: conteúdos, conectividade, tecnologia, gestão pública e gestão escolar.

Recentemente a Coreia do Sul implantou o *Plano de Desenvolvimento do Livro Didático Digital*. O governo sul-coreano selecionou os livros didáticos digitais como a mídia de educação para liderar um novo paradigma na educação do país, a “*Smart Education*”. Apresentando três razões: primeira, os estudantes já se tornaram o que são referidos como "nativos digitais"; segunda, os livros didáticos em papel enfrentam limitações na sociedade baseada no conhecimento e informação, onde a tecnologia ubíqua é a norma; e em terceiro lugar, o país é equipado com uma infraestrutura de tecnologia de informação e comunicação de primeira classe mundial e da capacidade humana para utilizar a infraestrutura de forma apropriada. Em consequência desse desenvolvimento tecnológico, desde 2011, o Ministério da Educação, Ciência e Tecnologia tem preparado a estratégia de promoção da *Smart Education*, centrando-se no ensino e aprendizagem personalizada (JUNG, 2009).

A *Smart Education* mudou a forma como se percebe os livros didáticos. A transferência dos livros tradicionais em papel para livros digitais permitiu que os alunos deixassem de usar mochilas pesadas e possibilitou a exploração de um mundo para além da sala de aula. Os livros didáticos digitais, ricos em conteúdos, estão diretamente ligados à promoção de aulas *online*. As aulas *online* não só ajudam os alunos a compensarem as ausências, mas também facilitam os estudos daqueles que não podem estar presentes devido às deficiências ou a problemas de saúde. Além disso, aulas *online* irão garantir aos estudantes o direito a escolherem seus temas de aprendizagem, mesmo para os alunos em áreas rurais que foram previamente privadas deste direito, devido à falta de professores (JANG, 2014).

A *Smart Education* significa que a educação é autodirigida, motivada, adaptada, recursos enriquecidos e envolvidos em tecnologia. O significado de motivada destaca a forma como o aluno é incentivado a ter interesse em aprender. Enfatiza os métodos de

ensino e aprendizagem que promovem a resolução criativa de problemas e está centrada no processo de avaliação individualizado. As experiências de aprendizagem dos estudantes serão transformadas: de tipicamente baseadas no livro didático para baseadas na experiência dos alunos. O termo adaptado significa a forma da educação por meio de um sistema educacional personalizado, um ensino e sistema de aprendizagem personalizado. Fortalece a flexibilidade do sistema e facilita a aprendizagem personalizada, em conexão com interesses pessoais e futuras aspirações de carreira. Também ajuda as escolas a evoluírem de um lugar de entrega de conhecimento para a um lugar que apoie a aprendizagem personalizada de acordo com os níveis e aptidões dos estudantes (JANG, 2014).

O serviço de aprendizagem nas nuvens oferece acesso gratuito a ricos conteúdos desenvolvidos por instituições públicas e privadas e indivíduos envolvidos com a educação. Expande a utilização conjunta de recursos de aprendizagem interno e externo, e promove a aprendizagem colaborativa através da entrega de conteúdos em plataformas nas nuvens. Estas plataformas permitem que os estudantes aprendam a qualquer hora e em qualquer lugar, e por meio da construção de um ambiente educacional que incentiva a aprendizagem centrada no estudante, que contam com diversos métodos de aprendizagem sob medida e auto selecionadas de acordo com o interesse dos estudantes. A *Smart Education* é a aplicação institucionalizada das tecnologias da informação e comunicação à educação como um meio essencial de um futuro paradigma educacional para o século 21 (JANG, 2014).

No Brasil as experiências com o uso das tecnologias de informação e comunicação e dos livros didáticos digitais são muito recentes. Em 2012, o Ministério da Educação comprou cerca de 382 mil *tablets* para atender aos professores do ensino médio, com investimento de 117 milhões de reais.

Em 2014 a *Amazon*, empresa multinacional de comércio eletrônico começou uma parceria com o Ministério da Educação, a partir do lançamento de um edital de convocação, no ano de 2012. Que tinha como objeto a constituição de acordos de cooperação entre o Fundo Nacional de Desenvolvimento da Educação (FNDE) e instituições interessadas a participar da estruturação e a operação de serviço virtual. A fim de disponibilizar obras digitais e outros conteúdos educacionais digitais para professores, estudantes e outros usuários da rede pública de ensino brasileira. Com ênfase nos títulos do Programa Nacional do Livro Didático (PNLD), do Programa Nacional Biblioteca da Escola (PNBE) e de outras ações governamentais na área de material escolar (BRASIL, 2012).

No contrato firmado, a *Amazon* responsabiliza-se em converter, gerenciar e distribuir livros didáticos digitais para professores, usando a tecnologia própria

Whispercast. Este termo de cooperação com o governo tem vigência até fevereiro de 2015, até o momento (junho de 2014), mais de 40 milhões de exemplares foram convertidos e hospedados na plataforma da *Amazon*.

Cerca de 200 títulos de livros serão disponibilizados para professores de escolas públicas por meio do aplicativo gratuito de leitura *Kindle*, instalados em *tablets* com o sistema operacional *Android* fornecidos pelo governo brasileiro. Embora a tecnologia *Whispercast* utilize o formato *Kindle*, compatível com o leitor digital da *Amazon*, ela também roda em PCs e em *tablets* com sistemas operacionais *Android* ou *iOS* que tenham instalado o aplicativo *Kindle*.

Nos Estados Unidos, a tecnologia *Whispercast* já é usada em diversas escolas como ponto de acesso para compra e distribuição de livros e documentos para programas educacionais. Pelo aplicativo é possível ler, destacar, fazer anotações e utilizar o dicionário diretamente nos livros didáticos digitais, mesmo quando os *tablets* não estão conectados à Internet.

Mesmo se tratando de um importante começo, esta parceria mesmo em estágio inicial já apresenta algumas limitações e questões que requerem detalhamento e aperfeiçoamento, a saber.

O tamanho do acervo disponível muito restritivo, os livros didáticos digitais como elementos complementares aos demais recursos didáticos necessitam possibilitar acesso à quantidade de conteúdo diferente dos conteúdos dos livros didáticos impressos, para assim, realizar sua finalidade.

O prazo de dois anos de parceria pode-se considerar muito pequeno, qualquer empreendimento de grande porte, como é este caso, necessita de um tempo maior, para implementá-lo, consolidá-lo, aperfeiçoá-lo e para atender as demandas que irão surgir.

A tecnologia utilizada restringe o uso dos livros didáticos digitais ao app do *Kindle* da *Amazon*. O uso de *software* proprietário exige análise sobre: custo de manutenção, dependência do desenvolvedor, perda da capacidade de extensibilidade (dados e dispositivos) de interoperabilidade com outros dispositivos e tecnologias, reduzindo as barreiras dentro do ambiente de aprendizagem e em casa.

Outra questão é a interface de aprendizagem pouco amigável, e os recursos de interação limitados. Existem diversos leitores de PDF que permitem fazer exatamente a mesma coisa. Ou seja, essa não é uma tecnologia específica do *Kindle*. Não é nem mesmo uma tecnologia exclusiva de *e-book*.

Problemas relacionados à privacidade, neste contrato, o governo está disponibilizando uma enorme base de dados de usuários para a empresa privada, a custo zero. E ainda forçando as pessoas a assinarem esse serviço, caso contrário elas não terão acesso aos conteúdos dos livros.

Mesmo que envolva apenas livros didáticos, a aliança com a *Amazon* representa uma abertura do governo nacional para a adoção de tecnologias estrangeiras na nuvem, hospedando conteúdos em *data centers* fora do território nacional e o uso de *software* proprietário. Isso fere alguns princípios de preservação de conteúdos digitais em longo prazo e de interoperabilidade com outras plataformas tecnológicas.

4 CONSIDERAÇÕES

Percebe-se que o uso dos livros didáticos digitais em ambiente computacionais nas nuvens se encontra em diferentes estágios entre os países estudados. No entanto, pode-se inferir que a ampliação de seu uso seja uma tendência, sobretudo, considerando-se: o aumento do número de conexões com a internet e dos dispositivos eletrônicos móveis, as novas possibilidades de uso das tecnologias da informação e comunicação com o desenvolvimento de novos produtos, conteúdos e aplicativos com foco na educação poderão intensificar ainda mais a *mobile learning*.

O uso de recursos educacionais digitais exige a participação mais ativa dos alunos, proporciona a interação através da cooperação e compartilhamento de informações e conteúdos. Exige também, do corpo docente e dos funcionários treinamento e formação para o uso correto e aproveitamento de todas as potencialidades dos recursos fornecidos por esta tecnologia.

Os benefícios são enormes, alguns são ainda desconhecidos e outros estão em fase de desenvolvimento. Conforme, o *framework* já apresentado destaca-se: a capacidade dos dispositivos de aprendizagem ser personalizáveis e interativos; a contribuição na função de gestão de aprendizagem e de ferramenta de avaliação; as melhorias que podem promover a saúde dos alunos; o auxílio no processo de gestão escolar com as facilidades proporcionadas; as economias de custos financeiros as escolas e aos pais; e a colaboração em garantir a sustentabilidade dos recursos naturais.

A apresentação das experiências do uso dos livros didáticos digitais em países como Estados Unidos e Coreia do Sul que se encontram em estágios mais avançados na utilização desses recursos corrobora na medida em que podem direcionar melhores ações de etapas já experimentadas para outros países. Contribuindo assim, no sentido de poder pular etapas, já testadas e aprimoradas no processo de implantação.

Ficou evidente, no estudo desses dois países, primeiramente, a necessidade de criar uma infraestrutura de rede de computadores com acesso a Internet nas escolas e nas comunidades em torno desta, possibilitando assim, maior conectividade. Importante também, ter a noção do nível de habilidade e naturalidade no uso das ferramentas tecnológicas por parte dos alunos, ditos “nativos digitais” e do corpo docente das escolas. Além disso, é necessária uma infraestrutura de mesas e cadeiras adequadas e rede de energia elétrica confiável, com tomadas de energia suficientes em todas as salas de aulas. Aliado as políticas públicas de aquisição de recursos educacionais em formato digital.

No Brasil as experiências com o uso dos livros didáticos digitais, na rede pública, estão em estágio embrionário, necessitando de uma articulação mais realista, entre as infraestruturas (energia elétrica, rede de computadores e Internet de banda larga) existentes nas escolas e o que está estabelecido nos planos de aquisição de conteúdo educacionais digitais, como o edital público que constituiu uma parceira com a empresa norte-americana *Amazon*.

Por ser um fenômeno hodierno e em fase de aperfeiçoamento o uso dos livros didáticos digitais nas escolas enfrenta alguns entraves que necessita mais aprofundamento, alguns de viés tecnológico (plataformas tecnológicas e uso de *softwares* proprietários), outras de questões políticas (decisão sobre o uso de acesso aberto e plataformas abertas) e outros de natureza legal (como, direitos autorais e o DRM), dentre outras.

O uso de livros didáticos digitais é uma temática que está em pleno desenvolvimento, sendo assim, são necessários diversos outros estudos para aprofundamento do debate e para o aperfeiçoamento desses recursos. Destaca-se a necessidade de se verificar se o uso dos livros didáticos digitais pode melhorar o desempenho dos alunos. Importante estudar, qual será o papel, as competências e habilidades necessárias do professor na sala de aula com a utilização dos livros didáticos digitais. E a necessidade de desenvolvimento de novas metodologias de ensino utilizando estes recursos.

Ao final, é importante ressaltar que os recursos tecnológicos não irão substituir o professor, e sim, apoia-lo no processo de ensino e aprendizagem, permitindo que os professores socializem seus conhecimentos para um grupo maior de alunos que tenham acesso a essas plataformas educativas nas nuvens, como já esta demonstrado na experiência do *OpenCourseWare* do MIT por exemplo.

Conclui-se, que o aperfeiçoamento do uso de todos os recursos educacionais digitais, não somente os livros didáticos digitais requerem certa urgência, uma vez que,

estes recursos poderão contribuir para a formação de uma geração mais bem preparada intelectualmente e como cidadão. E para conviver numa sociedade globalizada e com altos níveis de competitividade e com enormes desafios sociais a ser enfrentado.

REFERÊNCIAS

- ASH, Katie. Publishers turn to cloud computing to offer digital content. ***Education Week***, Bethesda, v. 5, n. 1, p. 34-36, 2011.
- BRASIL. Ministério da Educação. Fundo Nacional e Desenvolvimento da Educação. **Edital de convocação 03/2012 – CGPLI**. Brasília, DF: DOU, 2012. Disponível em: <<http://www.fnde.gov.br/arquivos/category/165-editais?download=7624:editorial-de-convocacao-digital>>. Acesso em: 30 abr. 2014.
- EMBONG, Abd Mutualib; *et al.* E-books as textbooks in the classroom. ***Procedia - Social and Behavioral Sciences***, Louisville, v. 47, p. 1802 – 1809, 2012.
- FEDERAL COMMUNICATIONS COMMISSION. **Charting our transition to “interactive digital textbooks”**. Washington, DC: FFC, 2012. Disponível em: <http://transition.fcc.gov/Daily_Releases/Daily_Business/2012/db0329/DOC-313315A3.pdf>. Acesso em: 30 abr. 2014.
- FEDERAL COMMUNICATIONS COMMISSION. **Digital textbook playbook**: the digital textbook collaborative. Washington, DC: FFC, 2012. Disponível em: <<http://www.fcc.gov/encyclopedia/digital-textbook-playbook>>. Acesso em: 30 abr. 2014.
- GSMA INTELLIGENCE. **Definitive data and analysis for the mobile industry**. Londres, 2014. Disponível em: <<https://gsmaintelligence.com>>. Acesso em: 2 jun. 2014.
- GSMA. Mobile education landscape report. Londres, 2011. Disponível em: <<http://www.gsma.com/connectedliving/wp-content/uploads/2012/03/landscape110811interactive.pdf>>. Acesso em: 17 jul. 2014.
- JANG, Sanghyun. Study on service models of digital textbooks in cloud computing environment for SMART education. ***International Journal of u- and e- Service, Science and Technology***, Tasmânia, v. 7, n. 1, p.73-82, 2014.
- JULA, Amin; *et al.* Cloud computing service composition: a systematic literature review. ***Expert Systems with Applications***, Shreveport, v. 41, p. 3.809-3.824, 2014.
- JUNG, Sung-Moo. Leading future education: development of digital textbooks in Korea. In: WORLD TEACHERS' DAY IN THAILAND, 4., 2009, Bangkok; UNESCO-APEID INTERNATIONAL CONFERENCE QUALITY INNOVATIONS FOR TEACHING AND LEARNING, 12., 2009, Bangkok. **Anais...** Bangkok, Tailândia: 2009. 13p.
- SOUZA, Francisco das Chagas de. **Escrevendo e normalizando trabalhos acadêmicos: um guia metodológico**. Florianópolis, SC: Ed. da UFSC, 1997. 126p. (Série Didática).
- UNESCO. **The future of mobile learning**: implications for policy makers and planners. Paris: UNESCO, 2013. (UNESCO Working Paper Series on Mobile Learning).

ESTUDOS BRASILEIROS SOBRE INDEXAÇÃO AUTOMÁTICA NO AMBITO DA CIÊNCIA DA INFORMAÇÃO (1973-2012)

BRAZILIAN ESTUDIES ABOUT AUTOMATIC INDEXING IN INFORMATION SCIENCE AREA (1973-2012)

Remi Correia Lapa
Renato Fernandes Correa

Resumo: Apresenta um panorama dos estudos sobre a Indexação Automática por meio do mapeamento e análise da produção acadêmica e científica nacional da área de ciência da informação no período de 1973 a 2012. Como objetivos específicos, procura coletar um *corpus* de análise e caracterizar as publicações quanto aos objetivos e aspectos metodológicos. A metodologia consiste em um estudo bibliográfico aprofundado de caráter qualitativo e quantitativo, bem como análise de conteúdo da produção literária no Brasil a respeito da indexação automática de textos escritos no idioma português, dentre livros, artigos de periódicos científicos, anais publicados e literaturas cinzentas. Os principais resultados encontrados foram: 35% dos trabalhos realizaram revisão bibliográfica, enquanto 65% investigaram a indexação automática por meio de fórmula, método ou sistema, dos quais 23% realizaram proposição e aplicação, 20% a proposição, e 22% realizaram aplicação; os sistemas como o objeto de estudo mais pesquisado, e a comparação com a indexação manual como o método de avaliação mais usado; o texto completo como a natureza do *corpus* mais pesquisado; o trabalho científico como a tipologia do *corpus* mais estudada; a indexação semi-automática como procedimento mais aplicado na validação dos termos; o processo de atribuição como o meio mais adotado para identificar os termos; o texto não estruturado como a entrada de dados preferida nos sistemas; a linguagem natural como a natureza da linguagem, os termos compostos como a natureza dos termos mais pesquisados; a análise estatística como o método de seleção dos termos mais utilizado. Concluímos que há uma tendência em estudos sobre a indexação automática por meio dos sintagmas nominais e que com o uso de novas tecnologias procura-se desenvolver uma identificação automática dos termos por meio da atribuição.

Palavras-chave: Indexação Automática. Indexação. Recuperação da Informação. Sistemas de Recuperação da Informação. Ciência da Informação. Brasil

Abstract: This work presents an overview of studies about Automatic Indexing through the mapping and analysis of brazilian scientific and academic production in information science area over the period 1973-2012. Specific objectives, seeks to collect a corpus to analyze and characterize the publications by observing its goals and methodological aspects. The methodology consists of a detailed bibliographical study and contends analysis on literary production in Brazil about the automatic indexing of texts written in Portuguese language. The corpus of analysis for the realization of content analysis, consists of documents in the Portuguese language, such as books, journal articles, proceedings and gray literature. The most significant results show that : 35% of the publications performed a literature review , in relation to the researched systems, methods, and formulas of automatic indexation , 23% made application and proposition, 20% made the proposition and 22% made application; systems as the object of study more researched; the comparison with manual indexing as the most used method of evaluation; the full text as the most researched corpus' nature; scientific publication as the corpus' typology most studied; semi-automatic indexing as the more applied term of validation procedure; the attributing process as the more adopted to identify terms; the unstructured text as data input preferred of systems; natural language as the nature of indexing language, the compound term as the nature of terms most searched;

statistical analysis as the method most used to term selection. We conclude that there is a tendency in studies on automatic indexing by means of noun phrases and the use of new technologies seeks to develop an automatic identification of terms by assigning.

Keywords: Automatic Indexing. Indexing. Information Retrieval. Information Retrieval Systems. Information Science. Brazil.

1 INTRODUÇÃO

Neste artigo, abordamos a Indexação Automática como um processo circunscrito ao campo da Ciência da Informação (CI), que investiga a geração, coleta, organização, interpretação, armazenamento, recuperação, disseminação, transformação e uso da informação, com ênfase particular, na aplicação de tecnologias modernas nestas atividades (CAPURRO; HJØRLAND, 2007).

De acordo com Moraes (2002), os problemas relacionados com a recuperação da informação tornaram-se de interesse para Hans Peter Luhn, especialista da IBM e pioneiro na aplicação da análise estatística de vocabulário para executar uma indexação automática. Luhn procurou soluções práticas e de baixo custo, o que o levou a utilização de máquinas para resolvê-los, tornando-se um defensor da Indexação Automática (PALMQUIST, 1998, tradução nossa).

Com a intenção de reverter os efeitos negativos causados pela produção de grandes volumes de informações, e almejando obter a informação confiável, de fácil acesso, com um tempo de resposta reduzido e com um custo acessível, o tratamento físico e de conteúdo dos documentos assumem um papel fundamental, pois, analisam, traduzem e representam a forma e o assunto dos documentos com a finalidade de auxiliar na recuperação de informações (ALVES; CAFÉ, 2010).

Nesse sentido, segundo Fujita (2009, p. 22) “estudos vem sendo desenvolvidos acerca da teoria da indexação, sua natureza, procedimentos, estruturas e características de seu produto final, o índice”, visando melhor tratamento temático e recuperação da informação.

Segundo Robredo (2005), existe uma preocupação em oferecer um acesso mais rápido à literatura técnico-científica utilizando o computador no processamento de dados e informações. Sua aplicação advém da necessidade em indexar grandes volumes de informações, em um tempo curto para manter as bases de dados atualizadas, o que torna inviável pensar na indexação manual (humana ou intelectual) como única forma de analisar e codificar o conteúdo dos documentos (ROBREDO, 2005).

A **problemática** subjacente à este artigo está em descrever e analisar a produção científica sobre a indexação automática no Brasil entre os anos 1973 e 2012.

Destarte, esta pesquisa tem por **objetivo geral** apresentar o panorama da pesquisa no âmbito da CI no Brasil referentes aos estudos sobre a Indexação Automática no período 1973 – 2012. Para tanto este artigo possui como **objetivos específicos**: levantar um *corpus* contendo a produção brasileira da área da Ciência da Informação a respeito da indexação automática entre os anos de 1973 a 2012; e analisar os objetivos e aspectos metodológicos das publicações que compõem o *corpus* selecionado.

A **justificativa** para a realização de tal pesquisa está no valor da informação obtida através da análise do conjunto de documentos selecionados, permitindo, deste modo, distinguir as tendências e realizar projeções sobre futuras pesquisas.

2 INDEXAÇÃO AUTOMÁTICA NO BRASIL

A indexação automática pode ser definida como um conjunto de operações, basicamente matemáticas, linguísticas, de programação, destinadas a selecionar termos como elementos descritivos de um documento pelo processamento de seu conteúdo.

Na indexação automática por extração o processamento do conteúdo não é permeado pela interpretação de terceiros, pois os termos significativos são extraídos do texto e ordenados pela sua frequência de ocorrência (NASCIMENTO, 2008).

Outro tipo, é a indexação automática por atribuição, que segundo Lancaster (2004), consiste numa representação temática por meio de termos selecionados de um vocabulário controlado (tesauro ou lista alfabética), onde um programa de computador desenvolve para cada termo a ser indexado um “perfil” de palavras ou expressões.

As primeiras propostas de indexação automática ocorreram nos anos 60, segundo estudo desenvolvido por Cesarino e Pinto (1980), e eram totalmente baseadas em métodos estatísticos de ocorrência de palavras.

No Brasil, segundo Vieira (1988b), a aplicação da indexação automática tem seu início no final dos anos 60, com a utilização do programa KWIC (*Keyword In Context*) para elaborar os índices das bibliografias especializadas publicados pelo Instituto Brasileiro de Bibliografia e Documentação (IBBD), atual Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT).

Na década de 70, as pesquisas de indexação automática em território nacional ocorrem através de estudos individuais, realizados em cursos de pós-graduação, concentrando-se na análise de frequência (VIEIRA, 1988b).

Nos anos de 1980 surgem os estudos baseados em referenciais linguísticos, conjuntamente com uma abordagem estatística, como por exemplo, o estudo de Andreewski e Ruas (1983) que trata da adaptação do sistema francês *Système Syntaxique et Probabiliste*

d'Indexation et de Recherche d'Informaticos Textuelles (SPIRIT) para documentos em língua portuguesa (GIL LEIVA, 1997).

O uso de referenciais lingüísticos, mais exatamente de critérios sintático-semânticos, tal como a proposta de uso de sintagmas nominais como unidades de análise, estão presentes nos trabalhos de alguns autores brasileiros a partir da década de 90 (KURAMOTO, 1995; SOUZA, 2006; BORGES; MACULAN; LIMA, 2008).

3 PROCEDIMENTOS METODOLÓGICOS

Este estudo formou-se por meio do mapeamento e da discussão da produção acadêmica e científica sobre a Indexação Automática no campo da Ciência da Informação no Brasil através de uma abordagem qualitativa e quantitativa.

O estudo desenvolveu-se como pesquisa exploratória, pois tem como finalidade “proporcionar maior familiaridade com o problema, com vistas a torná-lo mais explícito ou a construir hipóteses” (GIL, 2010, p. 27). Sendo assim, se propôs em realizar uma revisão de literatura com a finalidade de analisar os diversos aspectos referentes ao tema estudado.

No que diz respeito aos procedimentos técnicos, se caracteriza como pesquisa bibliográfica, pois se trata do levantamento de toda bibliografia nacional já publicada no idioma português, seja no formato de livros, artigos de periódicos científicos, anais publicados em congressos e seminários, e literaturas cíntezas (MARCONI; LAKATOS, 2010), obtida por meio de pesquisa nas bases de dados virtuais BRAPCI⁵⁰, Google Acadêmico⁵¹ e PERI⁵², e na biblioteca da UFPE, onde os documentos foram localizados através das expressões de busca ‘indexação automática’, ‘automatização da indexação’ e ‘indexação semi-automática’.

Visando obter um panorama das pesquisas sobre o tema desenvolveu-se análise de conteúdo em relação aos documentos do *corpus* levantado. O *corpus* constitui-se de 69 documentos, que foram categorizados quanto ao objetivo, objeto de investigação, nome do sistema/método/fórmula, avaliação da indexação automática, natureza e tipologia do *corpus* processado, validação e identificação dos termos, forma que ocorreu a entrada dos dados, linguagem de indexação e método/processo de identificação/ponderação/seleção de termos.

4 RESULTADOS E DISCUSSÕES

4.1 Análise do objetivo dos documentos do *corpus*

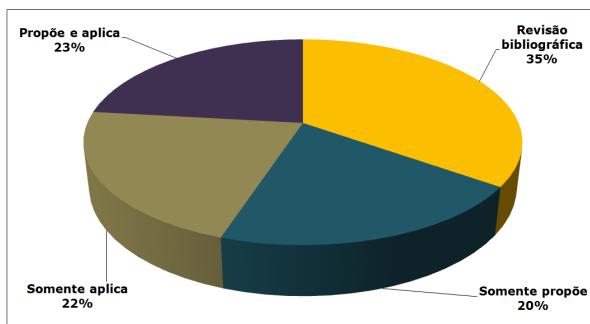
⁵⁰ <http://www.brapci.ufpr.br>

⁵¹ <http://scholar.google.com.br/schhp?hl=pt-BR&tab=ws>

⁵² <http://bases.eci.ufmg.br/peri.htm>

De acordo com o GRÁFICO 1 observamos que 24 trabalhos (35%) realizaram uma revisão bibliográfica.

GRÁFICO 1 – Objetivo dos documentos do *corpus*



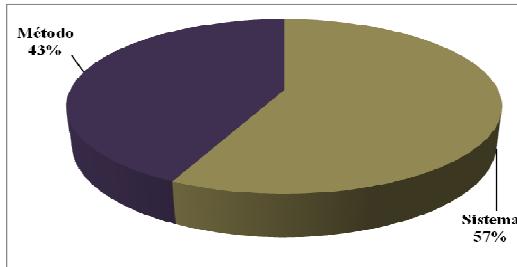
Fonte: desenvolvido pelo autor.

São 14 os trabalhos que ‘propõem’ algum método, sistema ou fórmula de indexação automática, correspondendo a 20% do total, os outros 31 trabalhos estão divididos entre aqueles que ‘aplicam’, 15 trabalhos (21%), e os que ‘aplicam e propõem’, 16 trabalhos (23%), algum sistema, método ou fórmula.

Procurando-se identificar como o tema da indexação automática é abordado nos trabalhos classificados na categoria Revisão Bibliográfica constatamos que a maioria dos trabalhos apresenta os fundamentos teóricos da indexação automática, sua evolução histórica e desenvolvimento teórico metodológico; um segundo grupo se concentra em abordar o embasamento filosófico e conceitual subjacente a Web Semântica e suas contribuições na automação da indexação na internet, por meio de motores de busca; o terceiro grupo é formado por trabalhos que discutem as vantagens e desvantagens do uso da indexação automática em comparação com a manual.

É providencial destacar três trabalhos individuais, pois são pontos de vista poucos explorados que podem estar surgindo para suprimir uma lacuna e representar base para o aparecimento de outras pesquisas: o primeiro de Guedes e Borschiver (2005), que apresenta a aplicação das leis e princípios da bibliometria, com foco nas palavras, na indexação automática; outro de Barreto (2007), que aborda a aplicação da indexação automática em vídeos; e Kochani, Boccato e Rubi (2011), que mencionam o desenvolvimento de uma política de indexação em sistemas automatizados.

GRÁFICO 2 – Distribuição dos trabalhos que realizaram proposição



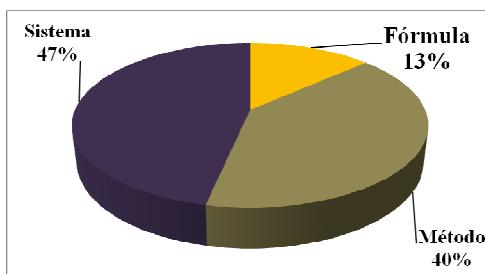
Fonte: desenvolvido pelo autor.

Conforme podemos averiguar no GRÁFICO 2, 14 trabalhos foram classificados no grupo dos que somente propõem algum método, sistema ou fórmula. Destes, nenhum trabalho propôs o uso de fórmula, seis trabalhos (43%), apresentaram a proposta de utilizar algum método, enquanto oito (57%) propuseram o uso de algum sistema de indexação automática.

Analizando trabalhos classificados na categoria proposição, verificamos que em relação ao sistema, o mais proposto foi o AUTOMINDEX/II; enquanto a extração dos Sintagmas Nominais foi o método mais proposto.

Dos 15 trabalhos (22%) que foram classificados como tendo o objetivo de somente aplicar fórmula, sistema ou método de indexação automática, dois (13%) aplicam fórmula; seis aplicam métodos (40%) e sete aplicam sistema (47%), ressaltados no GRÁFICO 3.

GRÁFICO 3 – Distribuição dos trabalhos que realizaram aplicação



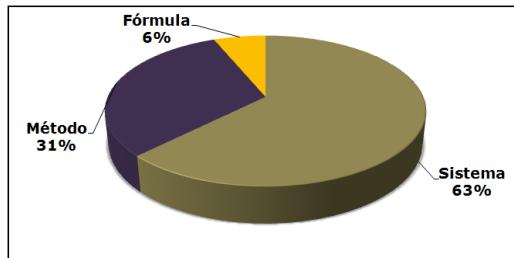
Fonte: desenvolvido pelo autor.

Levando em conta a aplicação, as fórmulas bibliométricas aplicadas foram as leis de Zipf e o ponto T de Goffman, em relação aos sistemas os mais aplicados foram o AUTOMINDEX/II e o SISA, já o método baseado na frequência de ocorrência foi o mais aplicado.

O GRÁFICO 4 ilustra a distribuição dos 16 trabalhos que realizam proposições e aplicações de fórmula, sistema ou método de indexação automática, onde um (6%) propõe e

aplica uma fórmula matemática; cinco (31%) estão relacionados com os métodos; e 10 (63%) propõem e aplicam sistemas de indexação automática.

GRÁFICO 4 – Distribuição dos trabalhos que realizaram proposição e aplicação



Fonte: desenvolvido pelo autor.

Em razão dos trabalhos que propõem e aplicam constatou-se um único trabalho referente à adaptação da fórmula de transição de Goffman, no que tange aos sistemas o PRECIS e o OGMA foram os mais propostos e aplicados enquanto aos métodos, todos fizeram menção aos Sintagmas Nominais.

4.2 Nome Sistema/Método/Fórmula

As fórmulas localizadas nas pesquisas são as Leis de Zipf e o Ponto T de Goffman, ambas consistem em fórmulas bibliométricas relacionadas com a frequência de ocorrência de palavras em textos, e aparecem como assunto em apenas três trabalhos.

Foram constatados 12 sistemas de indexação automática durante a análise dos trabalhos, e a relação entre o nome do sistema e sua frequência de ocorrência nos trabalhos está representado no GRÁFICO 5, de onde averiguamos que os três sistemas mais pesquisados para representar automaticamente os descritores aparecem empatados com quatro trabalhos (17%) cada um, são eles: o sistema SISA, o PRECIS e o AUTOMINDEX/II. Logo em seguida aparece o OGMA, sendo pesquisado por três trabalhos (13%).

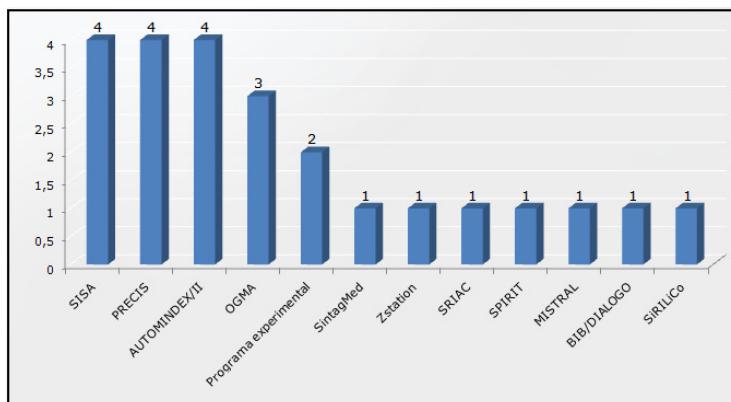
A quinta posição pertence a um sistema nomeado de ‘Programa experimental’, entretanto a verificação de um fato irá modificar este resultado, pois se constatou que este sistema experimental, registrado nos trabalhos de Haller (1983, 1985), foi desenvolvido em um computador Burroughs 6700 do CPD da UnB. Em Vieira (1988b), comenta-se que foi utilizado o sistema BIB/DIÁLOGO, implementado no Departamento de Biblioteconomia da UnB, para computadores Burroughs B6700, e terminais Burroughs, modelo TVA 800/10. Desta forma, o sistema outrora classificado como ‘programa experimental’ trata-se do sistema BIB/DIÁLOGO. E para finalizar este desfecho, Robredo (1991) explica que o sistema AUTOMINDEX/II, constitui-se num subsistema do sistema BIB/DIÁLOGO, o

qual já no início dos anos 80 é utilizado em estudos desenvolvidos por Robredo, então professor titular do Departamento de Biblioteconomia da Faculdade de Ciências Sociais Aplicadas da UnB.

Portanto, considerando que os sistemas ‘programa experimental’, BIB/DIÁLOGO e AUTOMINDEX/II fazem parte do mesmo sistema, e sendo classificados pelo sistema mais geral, o BIB/DIÁLOGO, este passa a ser o mais pesquisado com sete trabalhos.

Cada um dos seis sistemas restantes é proposto, aplicado, ou proposto e aplicado por apenas um trabalho. Desta forma os sistemas SintagMed, Zstation, SRIAC, SPIRIT, MISTRAL e SiRILiCo, representando 60% dos sistemas observados, não apresentam continuidade nas pesquisas e foram investigados por 25% dos trabalhos, enquanto os sistemas SISA, PRECIS, BIB/DIALOGO (programa experimental e AUTOMINDEX/II) e OGMA bancando os outros 40% foram pesquisados por 75% dos trabalhos, o que demonstra um certo prosseguimento nas pesquisas sobre estes sistemas.

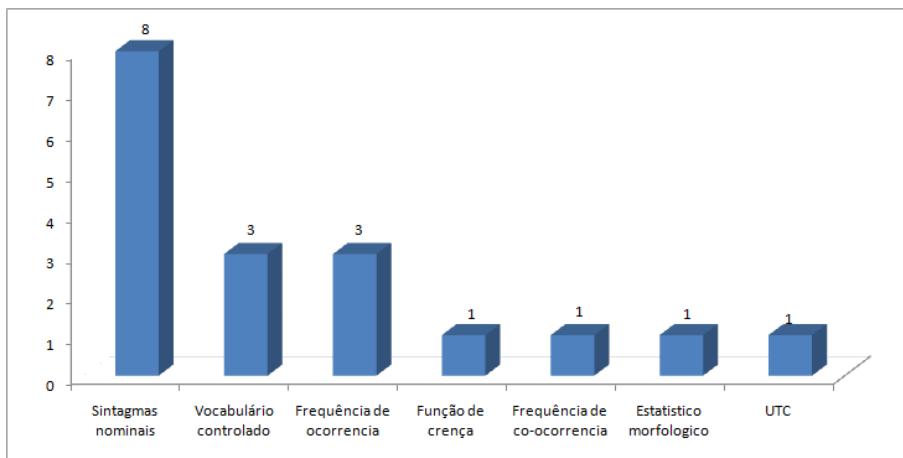
GRÁFICO 5 – Frequência de ocorrência dos sistemas nos documentos do *corpus*



Fonte: desenvolvido pelo autor

Avaliando os resultados referentes ao GRÁFICO 6 obtemos que o método mais pesquisado de indexação automática foram os sintagmas nominais, com 8 trabalhos (44%), dividido entre um trabalho que aplica; três que propuseram e aplicaram; e quatro que propuseram este método. Empatados na segunda colocação com três trabalhos (17%) cada um estão os métodos que utilizam o vocabulário controlado e a frequência de ocorrência. Este com três trabalhos que aplicam, e aquele com um trabalho que aplica e dois trabalhos que propõem e aplicam o método. Os demais métodos apareceram uma vez, são eles: função de crença, frequência de co-ocorrência, estatístico-morfológico e UTC.

GRÁFICO 6 – Frequência de ocorrência dos métodos nos documentos do *corpus*

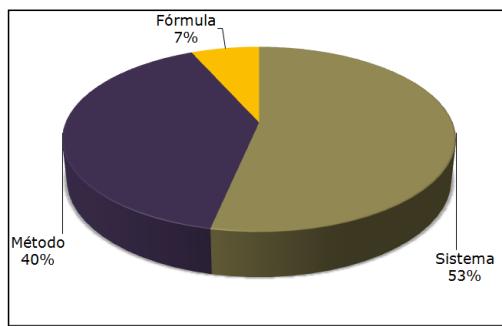


Fonte: desenvolvido pelo autor

4.3 A avaliação da indexação automática

De acordo com os dados contidos no GRÁFICO 7, 24 trabalhos (53%) avaliaram sistemas, destes seis aplicaram, oito propuseram e 10 propuseram e aplicaram; 18 trabalhos (40%) avaliaram métodos, sendo sete os que aplicaram, seis propuseram e cinco propuseram e aplicaram; e três trabalhos (7%) avaliaram fórmulas, onde dois aplicaram e um propôs e aplicou.

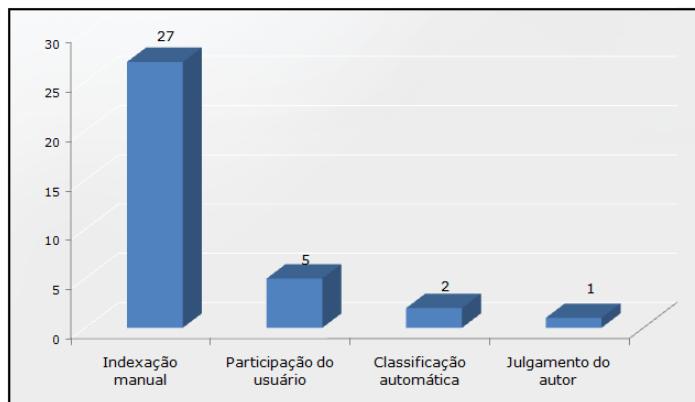
GRÁFICO 7 – Distribuição do objeto de avaliação



Fonte: desenvolvido pelo autor.

Para a grande maioria dos autores o método escolhido para avaliar suas pesquisas sobre o processo automático da indexação foi realizando uma comparação dos resultados obtidos pelo processo automático com os obtidos através do método intelectual de indexação, correspondendo a 27 dos trabalhos (71%), conforme GRÁFICO 8. Não foi possível identificar em dez trabalhos qual o método de avaliação empregado, mas nos 35 documentos analisados ficou evidente que a maioria optou como método de avaliação a comparação com a indexação manual.

GRÁFICO 8 – Frequência de ocorrência dos métodos de avaliação



Fonte: desenvolvido pelo autor.

Em segundo lugar, encontra-se o método de avaliar através da participação do usuário na comparação de índices ou Sistemas de Recuperação da Informação, com seis trabalhos (14%). Neste processo geralmente se aplica aos usuários, questionários e entrevistas estruturadas ou se avalia através de uma busca experimental comparada e simulada, com o objetivo de identificar as dificuldades e/ou facilidades através da reação dos usuários na utilização do índice. Esta avaliação pode ocorrer sobre dois pontos de vista, o do sistema e o do usuário.

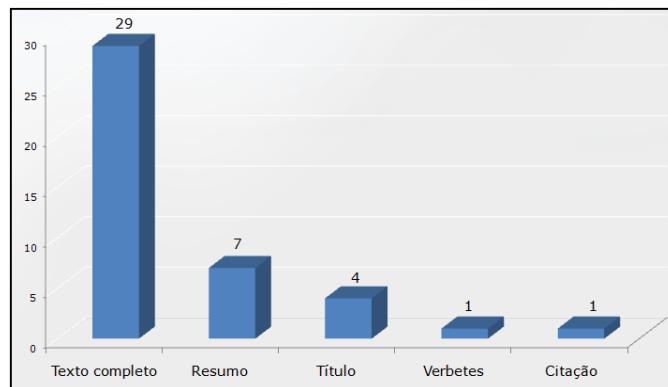
Dois trabalhos (11%) utilizam o método de classificação automática, que aplicam algoritmos de agrupamento e classificação para apresentar um valor percentual indicando quantos documentos foram classificados corretamente.

Um trabalho (1%) compara os resultados dos termos obtidos dos documentos através das modificações das fórmulas bibliométricas, onde o autor é quem fica encarregado de analisar se os termos obtidos são satisfatórios.

4.4 Natureza do corpus

Quanto à natureza do *corpus*, constatou-se que ocorreu uma preferência em se realizar pesquisas quanto à indexação automática do texto completo dos documentos, foram 29 trabalhos (69%). Não foi possível identificar em três trabalhos qual a natureza empregada.

GRÁFICO 9 – Natureza do corpus



Fonte: desenvolvido pelo autor.

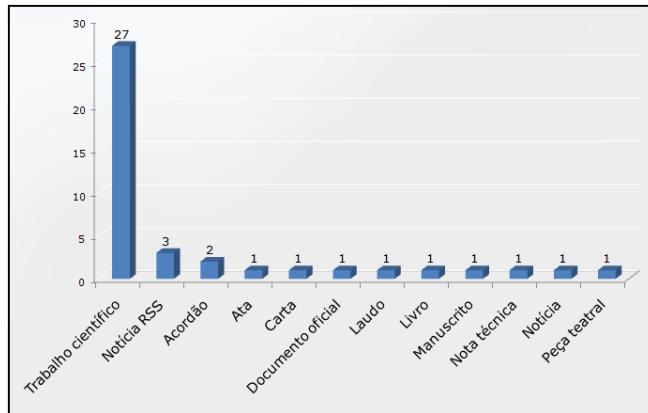
Os demais 31% estão divididos entre sete trabalhos que optaram utilizar os resumos como seu *corpus* de pesquisa, quatro trabalhos que utilizaram como *corpus* de análise os títulos. Tanto os verbetes quanto as citações constam como *corpus* de investigação em apenas um trabalho, cada. Os resultados estão ressaltados visualmente no GRÁFICO 9.

4.5 Tipologia do corpus

Observa-se que a distribuição referente à tipologia do *corpus* se comporta de acordo com o padrão das distribuições bibliométricas em geral: “poucos com muito e muitos com pouco”. Assim, no GRÁFICO 10 podemos constatar que poucas tipologias ocorreram muitas vezes, enquanto diversas tipologias ocorreram poucas vezes.

Os dados ilustrados no GRÁFICO 10 ajudam a compreender qual foi o comportamento da tipologia do *corpus* pesquisado nos trabalhos analisados. Verifica-se que a tipologia mais pesquisada com 27 trabalhos (66%) foram os trabalhos científicos, o que pode ser explicado por ser este tipo de material que mais interessa as instituições que normalmente desenvolvem e/ou financiam estas pesquisas, isto é, as Universidades Públicas, e por este motivo a importância da natureza do *corpus* incide sobre os trabalhos científicos que normalmente são produzidos na própria instituição. Não foi possível identificar em quatro trabalhos qual a tipologia do *corpus*.

GRÁFICO 10 – Tipologia do corpus



Fonte: desenvolvido pelo autor.

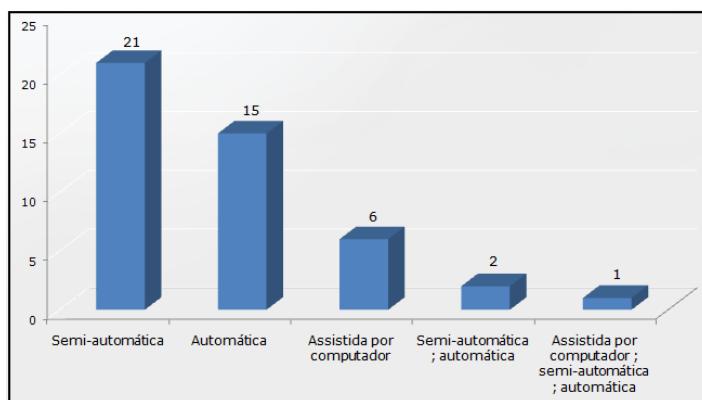
As demais tipologias, que juntas somam 14 trabalhos (34%), quase metade em relação à tipologia mais estudada, acabam refletindo realidades específicas caracterizando necessidades isoladas.

4.6 Validação dos termos

Analisando o tipo de indexação que os trabalhos aplicaram ou propuseram para validar os termos de suas pesquisas, observamos através do GRÁFICO 11 que em primeiro lugar, com 21 trabalhos (47%), estão aqueles que empregaram a indexação semi-automática, seguida pelos trabalhos que empregaram a indexação automática, com 15 trabalhos (33%), e um pouco atrás, se encontram seis trabalhos (13%) assistidos pelo computador.

Dois trabalhos (4%) declaram que a validação dos termos aconteceu tanto através de uma indexação semi-automática quanto automática. E um trabalho (2%) que a validação ocorreu através dos três critérios de análise.

GRÁFICO 11 – Validação dos termos



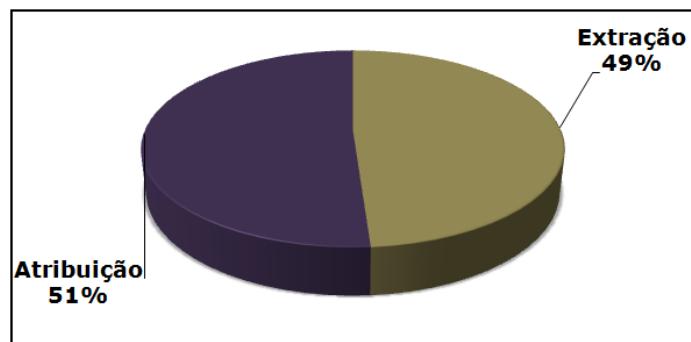
Fonte: desenvolvido pelo autor

4.7 Identificação dos termos

O GRÁFICO 12 demonstra que o processo de identificação foi realizado em 22 trabalhos (49%) por meio da extração, enquanto em 23 (53%) por meio da atribuição.

Esta pequena diferença pode ser explicada, pois apesar da dificuldade do computador em realizar o processo de obter um termo através da atribuição, a chegada de novas tecnologias e de pesquisas sobre aplicação de tesouros e vocabulários controlados motivou pesquisas sobre a atribuição.

GRÁFICO 12 – Identificação dos termos

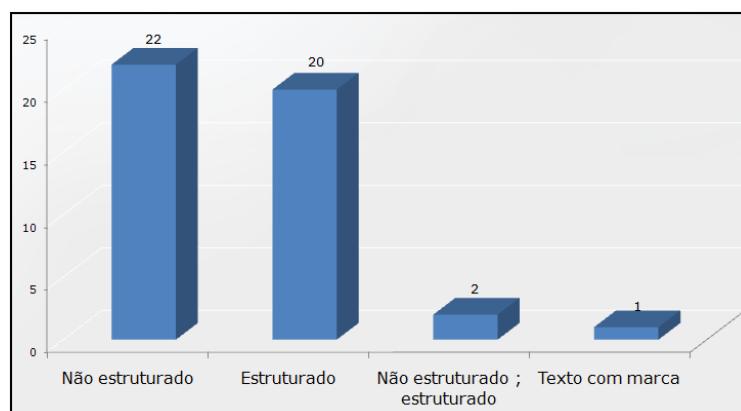


Fonte: desenvolvido pelo autor.

4.8 Forma que ocorreu a entrada dos dados

O GRÁFICO 13 ilustra como ocorreu à entrada dos dados nos sistemas, métodos e fórmulas investigados nos trabalhos que compõem o *corpus* deste artigo. Destes, 22 (49%) realizaram a entrada dos dados através de um texto não estruturado, 20 (44%) correspondem aos trabalhos que estruturaram o texto de alguma forma antes da inserção dos dados para análise, dois (aproximadamente 4%) foram os que alegaram trabalhar com a entrada de dados de forma não estruturada e em outro momento com o texto estruturado, e um trabalho (2%) cita utilizar a marcação nos textos.

GRÁFICO 13– Frequência de ocorrência da forma como ocorreu a entrada de dados

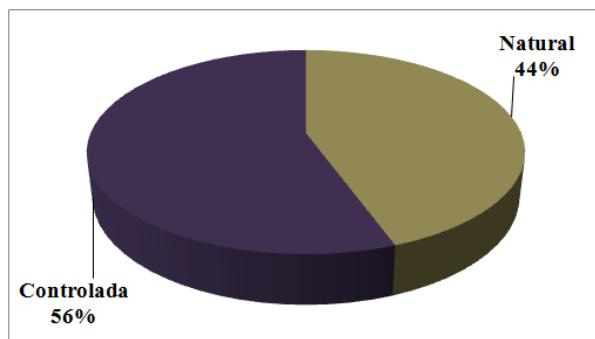


Fonte: desenvolvido pelo autor.

4.9 Linguagem de indexação

Nesta subseção procurou-se descrever os trabalhos quanto a natureza da linguagem de indexação (natural ou controlada) e dos termos (palavras isoladas ou termos compostos).

GRÁFICO 14 – Natureza da linguagem

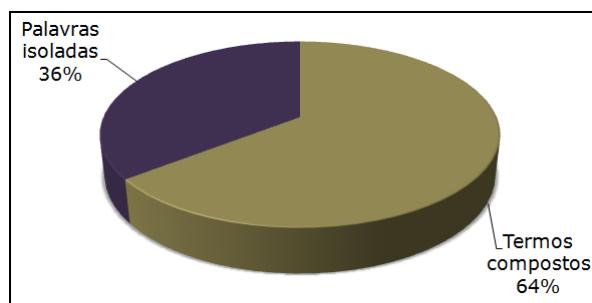


Fonte: desenvolvido pelo autor.

Quanto à natureza da linguagem notamos que ocorreu uma predominância pela linguagem controlada com 25 trabalhos (56%) em decorrência dos 20 trabalhos (44%) atribuídos à linguagem natural, visualizados no GRÁFICO 14.

Os dados relacionados aos termos estão ilustrados no GRÁFICO 15, que ressalta a superioridade numérica da extração de termos compostos com 29 trabalhos (64%) em relação à escolha apenas por palavras isoladas com 16 trabalhos (36%).

GRÁFICO 15 – Natureza dos termos



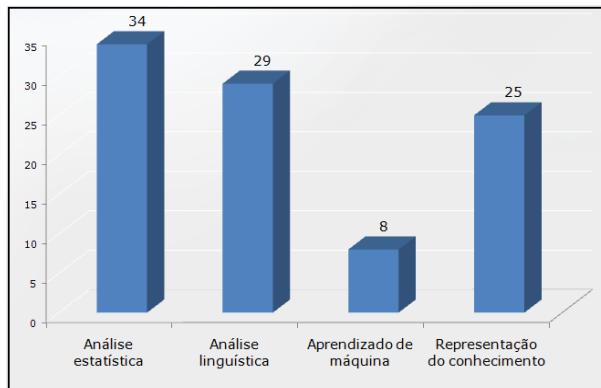
Fonte: desenvolvido pelo autor.

4.10 Abordagem utilizada na identificação / ponderação / seleção de termos

Quanto a abordagem ou tipo de técnicas utilizadas para identificar, ponderar ou selecionar os termos utilizados na indexação automática (em um total de 45 documentos que propuseram, aplicaram ou propuseram e aplicaram sistema/método/fórmula) foram observados quanto ao uso da análise estatística, análise linguística, aprendizado de máquina e representação do conhecimento, sendo classificados nas categorias e mensurados quanto à frequência em que apareceram nos trabalhos.

O resultado pode ser observado no GRÁFICO 16, onde o método mais pesquisado foi a análise estatística com 34 trabalhos (35%), o segundo método foi a análise linguística, aparecendo em 29 dos trabalhos (30%), a representação do conhecimento está em 25 trabalhos (26%), enquanto o aprendizado de máquina condiz com oito trabalhos (9%).

GRÁFICO 16 – Frequência de ocorrência do método/processo de identificação/ponderação/seleção de termos



Fonte: desenvolvido pelo autor.

5 CONSIDERAÇÕES FINAIS

Fundamentados nos resultados da análise de conteúdo dos documentos do *corpus* desta pesquisa, observamos que as categorias analisadas nos objetivos apresentaram percentuais próximos. Além disso, percebemos que a maior parte dos trabalhos classificados como revisão bibliográfica realiza uma abordagem da história da indexação automática, apontando seus fundamentos teóricos e evolução dos métodos.

Podemos concluir que os sintagmas nominais foram os métodos mais investigados. Em relação aos sistemas de indexação automática, quatro se destacam: o BIB/DIALOGO (incluindo o AUTOMINDEX), o SISA, o PRECIS e o OGMA.

Verificamos que a grande concentração dos trabalhos utilizou como método de avaliação a comparação com a indexação manual. Dessa forma, eles procuram avaliar se a implantação do sistema automático trará benefícios, obtendo resultados equivalentes em menos tempo.

Quanto à natureza e a tipologia do *corpus*, identificamos que a preocupação da maioria dos autores está concentrada em indexar o texto completo de trabalhos científicos.

Analizando o tipo de validação dos termos, percebemos a preferência pela aplicação da indexação semi-automática. O que pode ser justificado pelo fato dos processos totalmente automáticos ainda serem falhos e apresentarem limitações tecnológicas. Entretanto, a diferença em relação ao processo automático, na segunda posição, não é muito

grande, podendo ser interpretada como um esforço no desenvolvimento de uma indexação automática de qualidade.

Com relação à pequena diferença existente entre os trabalhos que pesquisaram a identificação dos termos por meio da extração (47%) e os que optaram pelo da atribuição (53%), há uma explicação provável. É que a chegada de novas tecnologias e de pesquisas sobre aplicação de tesouros e vocabulários controlados motivaram pesquisas sobre a atribuição, apesar da dificuldade em fazer com que o computador execute o processo de obter um termo através da atribuição.

Já a entrada dos dados apresentou um empate técnico entre textos não estruturados em relação ao texto estruturado. Em relação à linguagem de indexação de indexação, foi observada tanto a natureza da linguagem, que demonstrou uma preferência dos trabalhos pela pesquisa com a linguagem controlada, quanto a natureza dos termos, no qual a primazia encontra-se no estudo com termos compostos.

Quanto à categoria dos tipos de métodos de identificação, ponderação e seleção dos termos, chegamos à constatação de que o tipo de método mais pesquisado foi a análise estatística representando que uma grande parte dos trabalhos recorreram a, um ou mais dos seguintes processos: radicalização, eliminação de *stopwords*, análise de posição de ocorrência (localização), análise de frequência de ocorrência, análise de co-ocorrência, peso numérico, dicionário de raízes e/ou matriz binária.

O movimento ininterrupto da ciência continuará incentivando os pesquisadores a continuarem produzindo novas pesquisas. Consequentemente, eles identificarão e explicitarão outros caminhos (ou mesmo aqueles já trilhados, mas sobre uma ótica diferente), para que se chegue a um modelo automático de indexação de termos com qualidade igual ou superior a realizada pelo especialista humano quando realizam a mesma tarefa.

Em vista disso, este trabalho aponta trabalhos futuros na área da Ciência da Informação, sobre a indexação automática, que dariam continuidade ao trabalho desenvolvido nesta pesquisa.

Como sugestão para trabalhos futuros, apontamos:

- Investigar a análise de citação por permitir identificar características e mapear a comunicação científica;
- Realizar uma análise da produção internacional sobre a indexação automática;
- Mapear e discutir a produção acadêmica sobre a indexação automática em diversos campos do conhecimento (ciência da informação, ciência da computação,

linguística), diferentes fontes de informação, épocas e lugares, elaborando seu Estado da Arte.

REFERÊNCIAS

- ALVES, J. C.; CAFÉ, L. M. A. Análise focada em metadados sob a luz do padrão MTD-BR. **Em Questão**, Porto Alegre, v. 16, n. 2, p. 179-202, jul./dez. 2010.
- ANDREEWSKI, A.; RUAS, V. Indexação automática baseada em métodos linguísticos e estatísticos e sua aplicabilidade a língua portuguesa, **Ciência da Informação**, Brasília, v. 12, n. 1, p. 61-73, 1983.
- BARRETO, J. S. Desafios e avanços na recuperação automática da informação audiovisual, **Ciência da Informação**, Brasília, v. 36, n. 3, p. 17-28, set./dez. 2007.
- BORGES, G. S. B.; MACULAN, B. C. M. S.; LIMA, G. A. B. O. Indexação automática e semântica: estudo de análise do conteúdo de teses e dissertações. **Informação & Sociedade: Estudos**, João Pessoa, v.18, n.2, p. 181-193, maio/ago. 2008.
- CAPURRO, R.; HJØRLAND, B. O conceito de informação. **Perspectivas em Ciência da Informação**, v. 12, n. 1, p. 148-207, jan./abr. 2007.
- CESARINO, M. A. N.; PINTO, M. C. M. F. Análise de assunto, **Revista de Biblioteconomia de Brasília**, Brasília, v. 8, n. 1, p. 254-263, p. 32-43, jan./jun. 1980.
- FUJITA, Mariângela Spotti Lopes (Org.). **A indexação de livros**: a percepção de catalogadores e usuários de bibliotecas universitárias. Um estudo de observação do contexto sociocognitivo com protocolos verbais. São Paulo: Cultura Acadêmica, 2009.
- GIL, A. C. **Como elaborar projetos de pesquisa**. 5. ed. São Paulo: Atlas, 2010.
- GIL LEIVA, I. **La automatización de la indización, propuesta teórico-metodológica**: aplicación al área de Biblioteconomía y Documentación. 1997. 268f. Tese – Universidad de Murcia, Murcia, España, 1997.
- GUEDES, V.; BORSCHIVER, S. Bibliometria: uma ferramenta estatística para a gestão da informação e do conhecimento, em sistemas de informação, de comunicação e de avaliação científica e tecnológica. In: ENCONTRO NACIONAL DE CIÊNCIA DA INFORMAÇÃO, 6., 2005, Salvador. **Anais...** Salvador: ICI/UFBA, 2005.
- KOCHANI, A. P.; BOCCATO, V. R. C.; RUBI, M. P. Política de indexação para sistemas automatizados de coordenadorias de comunicação em ambientes universitários. In: CONGRESSO BRASILEIRO DE BIBLIOTECONOMIA, DOCUMENTAÇÃO E CIÊNCIA DA INFORMAÇÃO, 24., 2011, Maceió. **Anais...** São Paulo: FEBAB, 2011.
- KURAMOTO, H. Uma abordagem alternativa para o tratamento e a recuperação de informação textual : os sintagmas nominais, **Ciência da Informação**, Brasília, v. 25, n. 2, p. 1-18, 1995.
- LANCASTER, F. W. **Indexação e resumos**: teoria e prática. 2. ed. ver. atual. Brasília: Briquet de Lemos, 2004.

MARCONI, M. A.; LAKATOS, E. M. **Metodologia do trabalho científico:** procedimentos básicos, pesquisa bibliográfica, projeto e relatório, publicações e trabalhos científicos. 7. ed., 5. reimpr. São Paulo: Atlas, 2010.

MORAES, A. F. de. Os pioneiros da ciência da informação nos EUA. **Informação & Sociedade:** estudos, João Pessoa, v. 12, n. 2, 2002.

NASCIMENTO, G. F. C. L. **Folksonomia como estratégia de indexação dos bibliotecários no Del.icio.us.** 2008. 104f. Dissertação (Mestrado em Ciência da Informação) – Programa de Pós-Graduação em Ciência da Informação, Universidade Federal da Paraíba, João Pessoa, 2008.

PALMQUIST, R. A. **Class lecture notes:** Luhn and automatic indexing – references to the early years of automatic indexing and information retrieval. Organizing and providing access to information – LIS 391D.2 – Spring, 1998.

ROBREDO, J. **Documentação de hoje e de amanhã.** 4. ed. rev. ampl. Brasília, DF: Ed. Do Autor, 2005.

SOUZA, R. R. Uma proposta de metodologia para indexação automática utilizando sintagmas nominais. **Enc. Bibli: R. Eletr. Bibliotecon. Ci. Inf.**, Florianópolis, n. esp., p.42-59, 1. Sem. 2006.

VIEIRA, S. B. Indexação automática e manual: revisão de literatura. **Ci. Inf.**, Brasília, v. 17, n. 1, p. 43-57, jan./jun. 1988.

DESENVOLVENDO UMA PUBLICAÇÃO CIENTÍFICA AMPLIADA NA PLATAFORMA E-QUILT

DEVELOPMENT OF AN ENHANCED SCIENTIFIC PUBLICATION IN THE E-QUILT PLATFORM

Adriana Carla Silva de Oliveira

Guilherme Ataíde Dias

Mariana Cantisani

Resumo: A pesquisa aborda o atual cenário da comunicação científica apresentando uma nova modalidade de publicação científica denominada *Enhanced Publication*. Esse tipo de publicação é uma inovação científica com possibilidades de convergência, conectividade, compartilhamento de dados e colaboração entre pesquisadores. Discussões se dão também em torno das lacunas da Propriedade Intelectual e das Tecnologias Digitais da Informação e Comunicação. Nessa perspectiva, esta pesquisa traz como objetivo geral levantar padrões de estruturação de metadados adotados em *Enhanced Publication* para o desenvolvimento de um protótipo de artigo científico com características ampliadas. Adotou-se o método quadripolar constituído de 4 (quatro) polos: epistemológico, teórico, morfológico e técnico, sendo a abordagem de caráter exploratório e aplicado. Na condução da pesquisa, utilizou-se como subsídio uma pesquisa da área de saúde do Departamento de Clínica e Odontologia Social da Universidade Federal da Paraíba, Campus I, João Pessoa, Paraíba. Dentro do escopo da pesquisa, utilizou-se como viés do experimento o projeto intitulado Levantamento Epidemiológico em Saúde Bucal de Escolares do Município de Caaporã - PB, Brasil, cadastrado no Grupo de Pesquisa em Odontopediatria e Clínica Integrada. O protótipo criado para o experimento foi denominado e-Quilt. Como resultado desta fase inicial foi possível identificar os padrões de estruturação de metadados adotados pelas iniciativas de *Enhanced Publications* e a agregação de dados de pesquisa ao artigo. Utilizaram-se o software *Open Journal Systems* para publicação aberta, o *Cascading Style Sheets* e a linguagem de marcação *Hyper Text Markup Language*. Conclui-se que o protótipo *e-Quilt* é viável como uma publicação ampliada e será aprimorado na segunda fase do experimento.

Palavras-chave: *Enhanced Publication*. Protótipo *e-Quilt*. Periódico Eletrônico. Dados de Pesquisa. Padrão de Metadados.

Abstract: This piece of research deals with the current scientific communication scenario presenting a new modality of scientific publication named *Enhanced Publication*. This type of publication is a scientific innovation with convergence, connectivity, data sharing possibilities and collaboration among researchers. Discussions also occur in regard to the gaps of the Intellectual Property and the Digital Technologies of Information and Communication. In this perspective, the general objective of the research is to survey the structuring metadata standards adopted in enhanced publications for the development of a prototype of scientific article with enhanced characteristics. Hence, it was adopted the quadripolar method which comprises 4 (four) poles: epistemological, theoretical, morphological and technical. The research approach is of experimental and applied nature. In conducting the research, a piece of research in the health area, from the Department of Clinics and Social Dentistry of the Paraiba Federal University, campus I, João Pessoa, Paraíba, was used as a subsidy. In relation to the research scope, the project entitled Epidemiological Survey in Oral Health of the students from the Caaporã municipality – PB, Brazil, registered in the Research Group in Pediatric Dentistry and Integrated Clinics was used as bias of this experiment. The prototype created for the experiment was named *e-*

Quilt. As result of this initial phase, it was possible to identify the metadata structure patterns adopted by the initiatives of the Enhanced Publications and the gathering of research data for the article. The open source publication software Open Journal Systems, the Cascading Style Sheets and the Hyper Text Markup Language were used. It was concluded that the e-Quilt prototype is viable as an enhanced publication and it will be improved in the second phase of the experiment.

Keywords: Enhanced Publications. E-Quilt Prototype. Electronic Journal. Research Data. Metadata Standard.

1 INTRODUÇÃO

A infraestrutura global de informação já é uma realidade. A possibilidade de que governos, negócios, instituições, comunidades e pessoas pudessem estar conectados em um mundo interligado pelas tecnologias digitais, através de redes de comunicação, era inimaginável épocas atrás, o que hoje se configura como uma realidade concreta. No contexto da comunicação científica, o *e-Science* é o atual cenário que emerge com novas modalidades de publicações utilizando-se das Tecnologias Digitais de Informação e Comunicação (TDIC) para a geração de novos fluxos informacionais de dados e pesquisas.

O termo *e-Science* ou e-Ciência refere-se à utilização de dados de pesquisa, por meio de computação intensiva, para lidar com grande volume de dados. O *e-Science* é entendido como “Ciência avançada, e-ciência, é um dos muitos termos usados para descrever as transformações recentes no empreendimento científico”⁵³ (JANKOWSKI, 2007, p. 549, tradução nossa).

O cenário do *e-Science* fez surgir a necessidade de repensar a comunicação científica nos moldes tradicionais para um entendimento de não linearidade da publicação, com possibilidades de expansão e conexão. A mudança não está apenas na forma de publicar e no acesso entre as diversas fontes (dados, pesquisa, *links*, resultados, mídias, simulações) e públicos (pesquisador, produtor, usuário), mas também na possibilidade de convergência, conectividade, compartilhamento de dados e colaboração coletiva. A modalidade denominada de *Enhanced Publication*⁵⁴ se apresenta como uma inovação nesse novo cenário da pesquisa científica. Discussões se dão também em torno das lacunas da Propriedade Intelectual (PI) para o compartilhamento e reutilização de dados de pesquisa. Respostas às questões de cunho autoral e tecnológico ainda estão sem uma definição integral.

⁵³ Enhanced science, e-science, is one of many terms used to describe recent transformations in the scientific enterprise (JANKOWSKI, 2007, p. 549).

⁵⁴ Publicação Ampliada, Expandida ou Agregada.

Nesse sentido, estabelece-se a pergunta de pesquisa: quais padrões de estruturação de metadados estão sendo adotados para a ambiência de um artigo científico com características de *enhanced publication*? Com base nessa problemática, propõe-se como objetivo geral deste trabalho levantar os padrões de metadados adotados em *enhanced publication*. Para atender à proposição geral, definiram-se como objetivos específicos entender uma *enhanced publication*; apresentar casos de *enhanced publication* no cenário internacional; e, desenvolver um experimento a partir de um protótipo de artigo científico com características ampliadas.

A pesquisa adotou, para efeito desse estudo, o método quadripolar que consiste de 4 (quatro) polos: epistemológico, teórico, morfológico e técnico. A abordagem da pesquisa possui caráter exploratório e aplicado. O instrumento utilizado se deu por meio do experimento com a criação de um protótipo.

O trabalho está estruturado em seções. A seção inicial é composta pela introdução. A segunda seção abordará uma explanação sobre *enhanced publication* e as iniciativas no cenário internacional das *enhanced publications* e seus padrões de metadados. A terceira seção apresenta um experimento de protótipo de artigo científico com características ampliadas. E por fim, as conclusões do experimento e as referências norteadoras do estudo.

2 ENTENDENDO ENHANCED PUBLICATION

O atual cenário da comunicação científica exige interatividade e convergência de informações entre as publicações eletrônicas, permitindo que, além da geração e do uso da publicação, ocorram a possibilidade de reuso bem como o compartilhamento de dados e resultados de pesquisas. Jankowski (2007, p. 549, tradução nossa) afirma que uma “variedade de termos está em voga para descrever mudanças contemporâneas na condução da ciência. Os mais prevalentes incluem: a ciberciência, a ciberinfraestrutura e a e-ciência⁵⁵”.

Com base na expectativa de uma nova infraestrutura de publicação científica, surge o fenômeno do *e-Science* no universo acadêmico-científico. Quando se fala em comunicação científica, a primeira ideia é a geração de informações e conhecimentos sobre novas descobertas e avanços. Contudo, o acesso aos resultados e à troca de conhecimentos gerados depende do compartilhamento destes. O atual cenário da comunicação científica

⁵⁵ A variety of terms are in vogue to describe contemporary changes in the conduct of science. The most prevalent include: cyberscience, cyberinfrastructure and e-science (JANKOWSKI, 2007, p. 549).

propõe que esses conhecimentos possam ser compartilhados e reutilizados proporcionando uma grande geração, integração e interação de dados.

Veernooy-Gerritsen (2009, p. 21, tradução nossa) afirma que “no mundo digital dos estudantes, o acesso *on-line* é fornecido para artigos, com referência de *hiperlinks* e dados complementares. [...] a relação com outros materiais, por exemplo, multimídia e contexto semântico, por exemplo, XML, não é amplamente percebida no momento”⁵⁶.

Os padrões convencionais de publicações já não atendem aos novos fluxos e anseios da pesquisa científica. Nesse sentido, Sales, Sayão e Souza (2013, p. 1) afirmam que:

A comunidade científica vem sentindo a necessidade de novos modelos de publicações que possam expressar a complexidade e a dinâmica da pesquisa científica contemporânea. Os pesquisadores demandam por modelos que explicitem mais claramente a pesquisa desenvolvida, seus métodos, seu material, os dados e conhecimentos gerados; que revelem com clareza a intenção do pesquisador; e que os resultados sejam avaliados de forma mais eficaz e interativa.

Esta complexidade reflete o atual cenário acadêmico-científico. O padrão da publicação em um ambiente eletrônico traz consigo as mesmas características semânticas, lineares e estáticas de um documento impresso. Seu formato linear não permite a interatividade entre a publicação, o pesquisador e o usuário, pois conforme afirmam Aalbersberg, Dunham e Koers (2011, p. 2, tradução nossa): “O artigo atual é geralmente o mesmo dos primeiros periódicos científicos do século XVII. [...] uma nova infraestrutura para disseminação e compartilhamento do conhecimento científico que novos modelos de publicação, que utilizam o máximo potencial tecnológico, vêm sendo propostos”⁵⁷.

A infraestrutura digital que apoia as publicações científicas eletrônicas baseia-se em plataformas tecnológicas, padrões de metadados e protocolos de interoperabilidade cuja finalidade está centrada no armazenamento e na disseminação, disponibilizando o documento nos mais variados formatos, não permitindo a convergência, a colaboração e o compartilhamento entre as partes e os pares.

Ainda atual e pertinente, afirma Dias (2003, p. 21):

No momento em que acessamos alguma revista científica eletrônica, disponibilizada na web, não é difícil constatar que a vasta maioria das

⁵⁶ In the digital world of scholarly publishing online access is provided to articles, hyperlinked reference and supplementary data. [...], relation with other materials, e.g. multimedia, and semantic context, e.g. XML, is not realized widely at present. (VEERNOOY-GERRITSEN, 2009, p. 21).

⁵⁷ The current article is generally the same as the first scientific journals of the seventeenth century. It is so with the intention of establishing a new infrastructure for dissemination and sharing of scientific knowledge that new publishing models, using the maximum technological potential, have been proposed (AALBERSBERG; DUNHAM; KOERS, 2011, p. 2).

mesmas não faz bom uso dos recursos passíveis de implementação para o padrão web. O que encontramos, muitas vezes, são meras transcrições de periódicos já existentes de forma impressa para o meio eletrônico, sem nunca terem tido nenhum ancestral no formato impresso, não implementam as possibilidades inerentes ao meio eletrônico de forma satisfatória.

A necessidade de mudanças nos modelos das publicações se configura, de maneira generalizada, em todas as áreas e em escala mundial. O cenário da comunicação científica está em ebulação e evolução, não só em sua forma e acesso entre fontes (dados, pesquisas, *links*, resultados, mídias) e públicos (pesquisador, produtor, usuário ou indivíduo), mas também na sua convergência, conectividade e interatividade.

Tenopir *et al* (2011, p. 1, tradução nossa) complementam esse entendimento com:

Os dados são a infraestrutura de ciência [...]. Além disso, “a ciência está se tornando em dados intensivos e colaborativos” [1]. A quantidade de dados coletados, analisados, reanalisados e armazenados tem aumentado enormemente devido à evolução na simulação e modelagem computacionais, aquisição de dados automatizados e tecnologias de comunicação [2]. Segundo os paradigmas de pesquisa anteriores (experimentais, teóricos e computacionais), esta nova era tem sido chamada de “o quarto paradigma: descoberta científica intensiva de dados” onde “toda a literatura científica é *online*, todos os dados da ciência estão *online* e eles interagem uns com os outros” [3]. Os dados digitais não são apenas as saídas de pesquisa, mas fornecem subsídios para novas hipóteses, permitindo novos conhecimentos científicos e conduzindo à inovação [4]⁵⁸.

É possível perceber claramente que a pesquisa tradicional está em transição para uma modelagem de publicação com dados agregados de pesquisa. Os dados são potencializados e há, então, a possibilidade de esses serem reaproveitados para gerar novos estudos, pesquisas e conhecimentos científicos.

Borgman (2010, p. 2, tradução nossa) corrobora esse entendimento afirmando que:

Nas esferas científicas, as expectativas em torno de um mundo rico são imensas [...]. A relevância dos dados no contexto das “grandes ciências”, [...] conduziu não somente ao surgimento de novos modelos de ciência -

⁵⁸ Data are the infrastructure of science [...]. Moreover, “science is becoming data intensive and collaborative” [1]. The amount of data collected, analyzed, re-analyzed, and stored has increased enormously due to developments in computational simulation and modeling, automated data acquisition, and communication technologies [2]. Following the previous research paradigms (experimental, theoretical, and computational), this new era has been called “the fourth paradigm: data-intensive scientific discovery” where “all of the science literature is online, all of the science data is online, and they interoperate with each other” [3]. Digital data are not only the outputs of research but provide inputs to new hypotheses, enabling new scientific insights and driving innovation [4] (TENOPIR *et al*, 2011, p.1).

coletivamente chamados de “Quarto paradigma científico” ou “eScience” [...]⁵⁹.

É possível constatar não só a ideia de um novo paradigma revolucionário no cenário da comunicação científica, mas também na complexidade de produção, compartilhamentos, curadoria e propriedade intelectual dos dados e das publicações geradas nesse contexto.

Assim, “uma publicação ampliada é uma nova forma de comunicação em ciência, onde os pesquisadores tornam as publicações disponíveis *online* em conjunto com outros materiais. Ao adicionar os dados subjacentes e modelos em um artigo, torna-se mais fácil verificar os resultados, reproduzir e reutilizar”⁶⁰ (DANS, 2014, *online*, tradução nossa).

Reunir os dados e diferentes nuances, nesse novo cenário, requer pensar em uma nova modalidade de publicação científica. Nessa perspectiva do *e-Science*, a publicação ampliada, conhecida internacionalmente como *Enhanced Publication* (EP), pode ser compreendida como uma maneira totalmente nova de publicação da sua forma tradicional (um livro, um artigo ou um relatório), enriquecida com informações adicionais.

O termo *Enhanced Publication* ainda não é usual na comunicação científica, muito embora existam investigações datadas de 2001, que apresentam as primeiras manifestações de mudanças na publicação científica. Estudos se concentraram na possibilidade de convergência e entrelaçamento de produtos de *e-pesquisa* que se encontram distribuídos, expandindo o conceito tradicional de documento científico (AALBERSBERG, DUNHAM e KOERS, 2001; KIRZ, 2001, 2002; VEERNOOY-GERRITSEN, 2009; SAYÃO e SALES, 2012) para uma publicação ampliada.

Consideram-se também as novas percepções e ações nas atividades de cooperação entre pesquisadores, publicações e seus componentes na sociedade em rede. A FIGURA 1 ilustra o modelo de publicação na sua versão original, idealiza a concepção inicial de *Enhanced Publication* com os objetos agregados à publicação tradicional, em diferentes suportes e disponíveis em vários formatos de mídias, emergindo, assim, as possibilidades de convergências, interações e conexões em uma mesma publicação.

⁵⁹ In scientific spheres, expectations around a rich world are immense [...]. The relevance of the data in the context of the “big sciences” [...], led not only to the emergence of new models of science - collectively called “Fourth scientific paradigm” or “eScience” [...]. (BORGMAN, 2010, p. 2).

⁶⁰ An Enhanced Publication is a new form of communication in science, where researchers make publications available online in conjunction with other material. By adding the underlying data and models in an article, it becomes easier to verify the results, reproduce and reuse. (DANS. Disponível em: <<http://www.dans.knaw.nl/en/content/categorieen/nieuws/enhanced-publications-accessible-narcis>>. Acesso em: 20 maio 2014).

FIGURA 1 – Modelo tradicional de *Enhanced Publication* (EP)



Fonte: Openaire (2013, *online*).

Muitos são os desafios que cercam a publicação ampliada, dentre eles a infraestrutura para a interoperabilidade e compartilhamento entre dados, pesquisadores e publicações que são destaques neste artigo.

Estudo realizado por Veernooy-Gerritsen (2009, p. 21, tradução nossa), na Universidade de Amsterdam, aponta para a necessidade de construção de infraestrutura informacional e tecnológica quando afirma que:

Em uma publicação ampliada, o padrão de *link* deve apoiar e refletir a relação entre a publicação e todos os relevantes objetos como dados, *Web sites*, comentários. Este padrão pode tornar-se bastante complexo e vai além do que pode ser capturado na estrutura linear ou sequencial do modelo de publicação tradicional⁶¹.

A complexidade envolvida requer uma infraestrutura que abrange uma gama de tecnologias, técnicas, padrões e pesquisadores que possam desvendar as diferentes nuances desse contexto.

Borgman (2003, p. 4, tradução nossa), em seus estudos, aponta para uma infraestrutura informacional global, quando afirma que: “o acesso à informação está entre os argumentos fundamentais para a construção de uma infraestrutura informacional global.

⁶¹ In an Enhanced Publication, this link pattern should support and reflect the relation between the publication and all relevant objects like data, Web sites, commentaries. This pattern can become quite complex and goes beyond what can be captured in the linear or sequential structure of the traditional publication model (VEERNOOY-GERRITSEN, 2009, p. 21).

Recursos informacionais são essenciais para todo manejo de um acontecimento humano, incluindo comércio, educação, pesquisa e políticas governamentais”⁶².

Os recursos informacionais que permeiam essa infraestrutura são compostos por objetos digitais de diferentes formatos e variadas mídias, bem como os relacionamentos de agregação existentes entre estes. Na seção a seguir, descrevem-se os objetos digitais que compõem as *Enhanced Publications*.

2.1 Dados de pesquisa e objetos digitais

Antes de definir quais elementos podem compor uma *Enhanced Publication*, é importante entender as nomenclaturas entre objetos digitais e dados de pesquisa. Esses termos são usuais e podem ser sinônimos.

Em uma definição concisa, Borgman (2010, p. 3, tradução nossa) afirma que “os dados são fatos, números, letras e símbolos que descrevem um objeto, ideia, condição, situação ou outros fatores. A noção de dados pode variar consideravelmente entre os colaboradores e, mais ainda, entre as disciplinas”.⁶³

Desta forma, comprehende-se que os dados de pesquisa são todos os elementos que foram produzidos e gerados a partir de uma pesquisa. Elementos esses que podem ter valores semânticos diferentes, com formatos e mídias variadas. Os formatos e as diferentes mídias de dados e informações geradas compõem os objetos digitais de uma *Enhanced Publication*.

Ainda conforme a mesma autora, “o compartilhamento de dados é a liberação de dados de pesquisas para o uso de outros” (BORGMAN, 2012, p. 1060, tradução nossa).⁶⁴ Nessa perspectiva, trata-se do reuso de dados de pesquisas para continuidade de outros estudos. Este é o cerne do *e-Science*, disponibilizar dados de pesquisa a partir da agregação, compartilhamento e interação entre dados e pesquisadores. A agregação dos diferentes dados digitais, gerados na pesquisa ou em outras pesquisas, reúne objetos digitais em um único espaço, proporcionando a expansão do acesso e o compartilhamento dos diferentes objetos independentemente de sua localização.

⁶² Information access is among the primary arguments for constructing a global information infrastructure. Information resources are essential for all manner of human affairs, including commerce, education, research, government policy (BORGMAN, 2003, p. 4).

⁶³ Data are facts, numbers, letters and symbols that describe an object, idea, condition, location or other factors. The notion of data can vary considerably among employees, and even more between disciplines (BORGMAN, 2003, p. 4).

⁶⁴ Data sharing is the release of research data for use in other (BORGMAN, 2012, p. 1060).

O padrão de ligação dos objetos é um aspecto importante da infraestrutura “em uma publicação ampliada, este padrão de ligação deve apoiar e refletir a relação entre a publicação e todos os objetos relevantes, como dados, *sites*, comentários. [...] bastante complexo e vai além do que pode ser capturado na estrutura linear ou sequencial do modelo de publicação tradicional”⁶⁵ (VEERNOOY-GERRITSEN, 2009, p. 21, tradução nossa).

Essa infraestrutura deve obedecer a padrões que possibilitem relações não lineares e agregue os dados à publicação, como também à interoperabilidade entre os diferentes objetos para o seu compartilhamento e reuso. A agregação dos dados de pesquisa não tem uma regra definida. O pesquisador e a natureza da pesquisa delineiam os objetos digitais que serão anexados ao artigo principal. Contudo, é nesse ambiente que surge a necessidade de adotar padrões que estruturem os diferentes níveis de objetos.

2.2 Padrões de metadados

A complexidade de estruturação de vários elementos contidos nos diferentes suportes de informação gera a necessidade de adotar padrões de metadados enriquecidos e legíveis que possam ser recuperados e compartilhados em uma publicação ampliada. Nesse sentido, pode-se afirmar que:

Em cada modelo, tenta-se adicionar uma estrutura utilizando metadados em níveis muito diferentes [...]. Esta estrutura não é importante apenas para a leitura e compreensão humanas, mas também deve ser legível por máquina para fins de mineração de dados. Uma das dificuldades é que a adição de metadados e/ou estrutura rica é muito demorada. Quase todos os pesquisadores mencionam a importância da criação de ferramentas para auxiliar o autor. Além disso, (Lynch, 2007) menciona a importância de ferramentas: “Espero que vejamos uma nova geração de ferramentas de visualização e anotação implementadas, supostamente trabalhando em representações semanticamente ricas no documento XML”. Parece que a maioria dos formatos de metadados, como Dublin Core, MARC e METS, é muito limitada para Publicações Ampliadas⁶⁶ (VEERNOOY-GERRITSEN, 2009, p. 31, tradução nossa).

⁶⁵ In an Enhanced Publication, this link pattern should support and reflect the relation between the publication and all relevant objects like data, Web sites, commentaries. This pattern can become quite complex and goes beyond what can be captured in the linear or sequential structure of the traditional publication model (VEERNOOY-GERRITSEN, 2009, p. 21).

⁶⁶ In every model it is tried to add a structure by using metadata on very different levels [...]. This structure is not only important for human reading and comprehension, but should also be machine-readable for data mining purposes. One of the difficulties is that adding metadata and/or rich structure is very time consuming. Almost all researchers mention the importance of creating tools to assist the author. Moreover, (Lynch, 2007) mentions the importance of tools: “I hope that we will see a new generation of viewing and annotation tools deployed, presumably working on semantically rich XML document representations.” It appears that most metadata formats, like Dublin Core, MARC, and METS, are too limited for Enhanced Publications.

Os padrões de metadados existentes e adotados, atualmente, ainda estão sendo aprimorados para o contexto de compartilhamento de dados e das *Enhanced Publications*.

Muito embora algumas iniciativas desenvolvidas já apresentem soluções de estruturação, descrição dos metadados e agregação de objetos nesse contexto, “a atual infraestrutura de repositórios, com base em OAI-PMH, XML, *Dublin Core*, DIDL ou DDI, é capaz de lidar com os padrões mais complexos associados às publicações ampliadas” (VEERNOOY-GERRITSEN, 2009, p. 22, tradução nossa).

Corroboram-se essas iniciativas por meio do padrão METS, desenvolvido para atender ao contexto das Bibliotecas Digitais, o qual tem a finalidade de estruturar os metadados de diferentes objetos e suas relações de agregações. Acerca desse padrão, Rodrigues (2008, p.3) afirma que, “o padrão METS é uma linguagem de marcação baseada em XML que provê uma estrutura capaz de registrar metadados descritivos, administrativos e estruturais relativos aos objetos de uma biblioteca digital”. O padrão METS é expresso através de um XML Schema, e um documento XML, criado com base nesse padrão, é denominado de documento METS.

Nesse sentido, outra iniciativa se pauta na “*Nature Publishing Group* (2014, tradução nossa)⁶⁷, uma editora internacional que publica periódicos acadêmicos, revistas, bases de dados *on-line* e serviços no domínio da ciência e da medicina”. Disponibiliza, por meio do periódico *Nature*, uma iniciativa de publicação ampliada que se apoia em descritores detalhados para o uso e reuso de dados.

Descritores de dados irão fornecer descrições detalhadas das experiências e procedimentos envolvidos na geração de conjuntos de dados importantes, incluindo informações essenciais necessárias para os cientistas avaliar a qualidade técnica dos dados, reproduzir os principais métodos ou fluxos de trabalho de análise e, finalmente, voltar a utilizar os dados para lidar com importantes questões de pesquisa. Além disso, todas as publicações de dados científicos serão apoiadas por metadados que descrevem propriedades-chave das experiências e dados resultantes, que serão verificados por um curador interno e lançados no formato ISA-guia, e esperamos que em outros formatos padrões no futuro. Estes metadados ajudarão a mineração de dados, e vão ajudar os cientistas a encontrar e reutilizar os conjuntos de dados de alta qualidade armazenados em vários repositórios de dados⁶⁸ (NATURE, 2014, *online*, tradução nossa).

⁶⁷ Nature Publishing Group (2014) is an international publishing company that publishes academic journals, magazines, online databases, and services in science and medicine.

⁶⁸ Data Descriptors will provide detailed descriptions of the experiments and procedures involved in generating important datasets, including essential information needed for scientists to assess the technical quality of the data, reproduce key methods or analysis workflows, and ultimately reuse the data to address important research questions. In addition, every publication at *Scientific Data* will be supported by metadata describing key properties of the experiments

A *Nature* adota o padrão *ISA-tab* para a descrição detalhada dos metadados de todos os objetos agregados aos dados de pesquisas que irão compor a *Enhanced Publication*. Esses são imprescindíveis para possibilitar a estruturação, curadoria e interoperabilidade dos dados e publicação. Os padrões atuais ainda estão sendo customizados para a estrutura de *Enhanced Publication*. O XML deve proporcionar a agregação e as relações entre os diversos objetos.

Nesse sentido, Veernooy-Gerritsen (2009, p. 31, tradução nossa) afirma que “a informação que precisa ser gravada não pode ser mapeada facilmente para uma estrutura XML hierárquica. Estamos lidando com objetos, propriedades e tipos de objetos e relações entre objetos. [...]. Para esta nova infraestrutura, precisaremos usar o modelo OAI-ORE”⁶⁹.

A afirmação sugere que o modelo OAI-ORE seja utilizado para a descrição e troca de agregações entre objetos digitais. O *Open Archives Initiative* (2014, *online*, tradução nossa) é,

A Iniciativa de Arquivos Abertos para Reuso de Objetos e Intercâmbio (OAI-ORE) define padrões para a descrição e troca de agregações de recursos da Web. Essas agregações, às vezes chamadas de objetos digitais compostos, podem combinar recursos distribuídos com vários tipos de mídia, incluindo texto, imagens, dados e vídeo. O objetivo dessas normas é expor o conteúdo rico nestas agregações para aplicativos que suportam a criação, depósito, troca, visualização e preservação⁷⁰.

Os estudos apontam para a adoção do modelo OAI-ORE, em publicações ampliadas, para possibilitar que os objetos digitais de pesquisa sejam agregados à publicação permitindo seu compartilhamento e reuso em diferentes mídias e suportes de informação.

and resulting data, which will be checked by an in-house curator and released in the ISA-tab format, and hopefully other standard formats in the future. These metadata will aid data mining, and will help scientists find and reuse high-quality datasets stored across multiple data repositories. (NATURE, 2014, *online*. Disponível em:<<http://blogs.nature.com/scientificdata/2013/07/23/scientific-data-to-complement-and-promote-public-data-repositories>>. Acesso em: 23 jun. 2014).

⁶⁹ The information that needs to be recorded cannot be mapped easily to a hierarchical XML structure. We are dealing with objects, properties and types of objects, and relationships between objects [...]. For this new infrastructure, we will need to use the OAI-ORE model (VEERNOOY-GERRITSEN, 2009, p. 31).

⁷⁰ Open Archives Initiative Object Reuse and Exchange (OAI-ORE) defines standards for the description and exchange of aggregations of Web resources. These aggregations, sometimes called compound digital objects, may combine distributed resources with multiple media types including text, images, data, and video. The goal of these standards is to expose the rich content in these aggregations to applications that support authoring, deposit, exchange, visualization, reuse, and preservation. (OPEN ARCHIVES INITIATIVE (OAI), *online*. Disponível em: <<http://www.openarchives.org/ore/>>. Acesso em: 28 jul. 2014).

Para efeito desse artigo serão abordados apenas os aspectos relacionados aos padrões de metadados. Na seção seguinte serão demonstradas as iniciativas do cenário internacional.

3 CASO DE ENHANCED PUBLICATION NO CENÁRIO INTERNACIONAL

Esta seção aponta para uma iniciativa no cenário internacional de *Enhanced Publication*, que norteou a construção do protótipo de artigo científico com características ampliadas associadas a esta pesquisa. Algumas iniciativas e muitas discussões estão sendo desenvolvidas no cenário internacional no que tange ao *e-Science*, especialmente nos contextos do *Scientific Data*⁷¹, *Data Sharing*⁷² e *Enhanced Publications*.

Foi delimitada, para efeito deste estudo, uma iniciativa de *Enhanced Publication* desenvolvida no cenário internacional pela *Nature Publishing Group*.

3.1 Periódico *Nature*

O periódico *Nature* apresenta uma experiência de convergência de objetos digitais com o artigo tradicional, fornecendo possibilidades de acesso ao conjunto dos dados oriundos da pesquisa. A partir da combinação da publicação científica com as informações estruturadas em *datasets*, é possível permitir o acesso aos dados de pesquisa. A *Nature* é uma plataforma que hospeda no sítio *nature.com*, aproximadamente, 80 revistas publicadas pela *Nature Publishing Group*, com inúmeros cientistas de acesso aberto e com um programa de publicação *online* contínua.

Na perspectiva do *e-Science*, a *Nature*⁷³ menciona que “os dados científicos objetivam se referir à crescente necessidade de tornar os dados de pesquisa mais disponíveis, citáveis, detectáveis, interpretáveis, reutilizáveis e reproduzíveis”. (SCIENTIFIC DATA, 2014, *online*, tradução nossa).

Dessa forma, a *Nature* desenvolveu uma nova perspectiva de acesso aberto dos seus periódicos adicionando ao artigo, descritores que possibilitem acesso a dados primários e permitindo as funcionalidades de comunicação e interação com a publicação, conforme ilustrado na FIGURA 2.

Os dados científicos, disponíveis na *Scientific Data* da *Nature*, disponibilizam os artigos científicos e permitem agregar diferentes dados de pesquisa, vinculando também,

⁷¹ Dados Científicos.

⁷² Compartilhamento de Dados.

⁷³ *Scientific Data* aims to address the increasing need to make research data more available, citable, discoverable, interpretable, reusable and reproducible (SCIENTIFIC DATA. Disponível em: <<http://www.nature.com/sdata/about>>. Acesso em: 23 jun. 2014).

recursos e ferramentas *web* em uma única interface. Informações sobre os autores, contribuições e afiliações também estão disponíveis por meio de *hiperlinks*.

A FIG. 2 apresenta ainda funcionalidades de acesso a redes e mídias sociais. No acesso ao artigo, há a possibilidade de serem disponibilizados o texto em PDF, as métricas, citações e direito de permissões.

FIGURA 2 – *Homepage* dos Dados Científicos da *Nature*



Fonte: *Nature* (2014).

Isso é viável a partir da adoção de um padrão de estruturação de metadados, descrito na seção a seguir.

3.1.1 Padrão de Estruturação da *Nature*

O padrão de estruturação de metadados da *Nature* é o protocolo *ISA*. O *ISA* é de autoria de diversos colaboradores em mais de 30 (trinta) organizações científicas em todo o mundo. Como resultado, o projeto beneficia a colaboração, evitando a repetição desnecessária dos dados de pesquisa.

A partir da descrição desses dados e metadados, esse padrão estrutura as informações de toda a pesquisa, além de utilizar abordagens da *web* semântica para tornar o conhecimento existente disponível para a realização de conexões que permitam que os dados sejam recuperados de suas fontes originais.

Por descritores, a *Nature*⁷⁴ define que: “Descritores de dados são uma combinação de conteúdo de publicação científica tradicional e informações estruturadas na curadoria interna e são projetados para maximizar a reutilização e permitir a pesquisa, vinculação e mineração de dados” (SCIENTIFIC DATA, 2014, *online*, tradução nossa).

O protocolo utilizado pela *Nature* para as suas publicações ampliadas é o *ISA-tab*⁷⁵. Para operacionalizar e usar amplamente os dados, a comunidade científica precisa adotar tecnologias e mecanismos que possam apoiar a interoperabilidade desses dados de pesquisa. O *ISA-tab* é um *framework*⁷⁶ que descreve e estabelece os pré-requisitos apresentando um ecossistema de soluções para o compartilhamento e reutilização dos dados de pesquisa.

A FIG. 3 ilustra as interligações que a estrutura *ISA-tab* utiliza: os conjuntos de dados de interesse para cada comunidade e como os sistemas de captação são alimentados pela suíte de softwares *ISA*.

O *framework*⁷⁷ ISA está estruturado de maneira hierárquica, conforme apresentado na FIGURA 3. Essa estrutura do formato de arquivo *ISA-tab* possui código aberto e foi desenvolvida sobre as seguintes categorias de metadados: uma sobre investigação (*investigation*) que trata do contexto do projeto; a segunda refere-se ao estudo (*study*) que define uma unidade de investigação; e, por último, o ensaio (*assay*) que realiza a medição analítica (dados quantitativos e qualitativos).

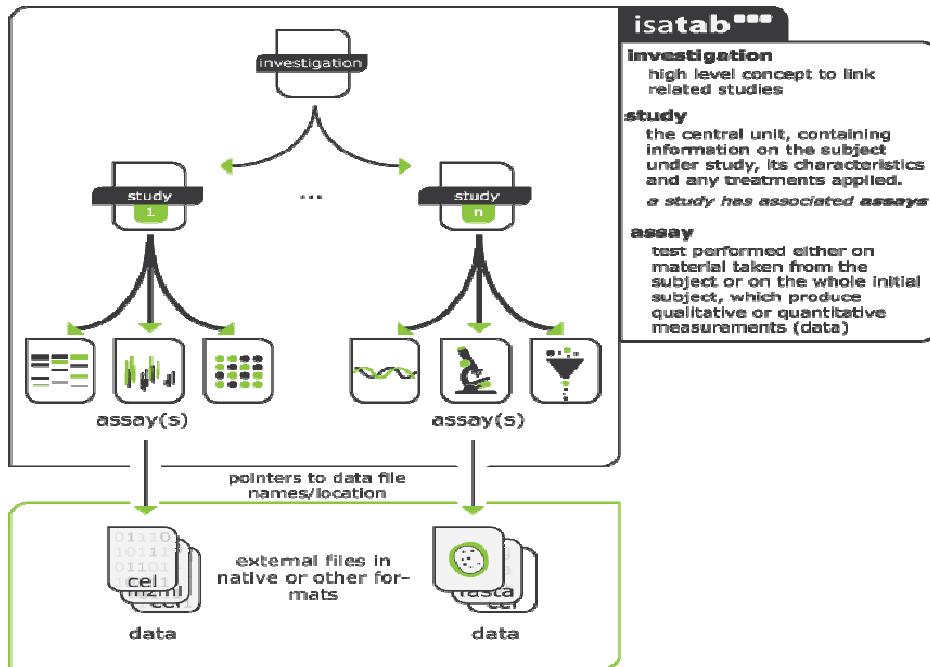
⁷⁴ Data Descriptors are a combination of traditional scientific publication content and structured information curated in-house, and are designed to maximize reuse and enable searching, linking and data mining (SCIENTIFIC DATA. Disponível em: <<http://www.nature.com/sdata/archive>>. Acesso em: 23 jun. 2014).

⁷⁵ Tradução de Guia-ISA. Disponível em: <<http://www.isa-tools.org/>>. Acesso em: 10 jul. 2014.

⁷⁶ Tradução de Estrutura.

⁷⁷ Disponível em: <http://www.nature.com/ng/journal/v44/n2/fig_tab/ng.1054_F2.html>. Acesso em: 15 jul. 2014.

FIGURA 3 – Categorias de metadados da estrutura *ISA-tab*



Fonte: *ISA-Tab* (2014).

Além da estrutura *ISA-tab* existe a *ISA Commons*⁷⁸: uma comunidade crescente que utiliza os metadados ISA para facilitar a criação de normas compatíveis para a gestão, colaboração, curadoria e reutilização de conjunto de dados em uma ambiência, cada vez mais, diversificada de domínios das ciências.

4 PROTÓTIPO DE ARTIGO AMPLIADO: UM EXPERIMENTO DE PESQUISA

Esta seção apresenta a primeira fase do protótipo que compõe o experimento de pesquisa da tese de doutorado, ora em andamento. Dessa forma, o protótipo objetivou a elaboração de um artigo científico, com características ampliadas, cuja pretensão foi alcançar a essência de uma *Enhanced Publication* e desenvolver um padrão de estruturação de metadados. Nessa primeira fase, recorreu-se à experiência da *Nature Publications* para identificar o padrão de metadados utilizado na estruturação dos elementos e na agregação dos objetos de pesquisa.

A denominação dada ao protótipo foi o *e-Quilt*⁷⁹. Esta denominação reflete o estado da arte do experimento, ilustrado por meio do protótipo do artigo com características

⁷⁸ Disponível em: <<http://isacommons.org/>>. Acesso em: 10 jul. 2014.

⁷⁹ Definição dos autores (2014): *Quilt* = Colcha de Retalhos, o *e-Quilt* é uma derivação da palavra *Quilt* para o ambiente digital no *e-Science*.

ampliadas, que apresentam, na sua essência, uma publicação ampliada composta pelo artigo tradicional agregando dados e objetos de pesquisa associados.

FIGURA 4 - Logo *e-Quilt*



Fonte: Elaborado pelos autores (2014).

Para a realização dessa fase, utilizou-se, como subsídio, uma pesquisa da área de saúde, do Departamento de Clínica e Odontologia Social da Universidade Federal da Paraíba, *Campus I*, João Pessoa, Paraíba. Dentro do escopo da pesquisa, o projeto intitulado Avaliação de Proposta de Qualificação de Recursos Humanos e Programação no Programa Saúde da Família derivou várias pesquisas, dentre elas: o Levantamento Epidemiológico em Saúde Bucal de Escolares do Município de Caaporã - PB, Brasil, cadastrado no Grupo de Pesquisa em Odontopediatria e Clínica Integrada, no CNPq, utilizado para o viés deste experimento.

O protótipo desenvolvido foi realizado com o propósito de agregar dados de pesquisa, gerados no campo da pesquisa, para o artigo tradicional. Para tanto, foi necessário definir uma sistemática e um *framework* para a estruturação do artigo na plataforma *e-Quilt* que vislumbrasse, nessa primeira fase, a adoção de critérios necessários para a arquitetura, agregação e linkagem dos objetos digitais relacionais da publicação.

A sistemática para a estruturação do artigo ampliado constituiu-se de: a) elaboração do artigo; b) estruturação dos elementos para padrão de metadados; c) formatação normativa do artigo; d) instalação do *Apache OpenOffice 4* para o pré-teste; e) criação de uma estrutura HTML; f) geração da página HTML com os conteúdos; g) agregação dos objetos da pesquisa (*hyperlinks* e âncoras); h) instalação da plataforma OJS; i) customização da plataforma *e-Quilt*; j) *upload* dos arquivos na plataforma *e-Quilt*.

Essa sistemática foi desenvolvida, simultaneamente, com a estruturação dos metadados do artigo, conforme a seção seguinte.

4.1 Estruturação dos metadados *e-Quilt*

Com base no caso de *Enhanced Publication* da *Nature*, foram adotados para o experimento *e-Quilt* dois tipos de metadados, a saber: descritivos e administrativos.

Santos *et al.* (2014, p. 148- 149) afirmam que,

No ambiente digital, os metadados influenciam tanto no acesso quanto na recuperação da informação e são utilizados como elementos de busca para facilitar o acesso aos dados descritivos e a localização de recursos na web, e para que tenham aceitação internacional e auxiliem na interoperabilidade eles devem possuir uma padronização.

Os metadados descritivos são usados para descrever, identificar e representar recursos de informações. Esses fornecem os elementos para descrição do título, autor, imprensa, data, resumo, palavras-chave que podem ser *hiperlinks* e anotações sobre usuários. E os metadados administrativos são usados no gerenciamento e administração dos recursos de informação. Esses, por sua vez, fornecem informações sobre a data de criação dos recursos, tipos de arquivos, formas de acesso, controle de direitos e reproduções e dos registros legais.

Esses dois níveis de metadados foram utilizados para padronizar a formatação do artigo ampliado. Para cada tipo de metadados, foram descritos elementos oriundos do conteúdo e dos dados agregados à pesquisa e outros elementos e ferramentas adicionadas pela plataforma do periódico. Entre cada seção do artigo, *hyperlinks* foram adotados para ligar os conteúdos do texto às suas respectivas seções.

A seguir, descreve-se brevemente a plataforma de gestão e linguagem de programação adotada na ambiência do *e-Quilt*.

4.2 Plataforma de gestão eletrônica do *e-Quilt*

Como etapa inicial do protótipo, a plataforma OJS (*Open Journal System*), na versão 3.0-beta, foi utilizada para a gestão eletrônica do artigo. Inicialmente, foi criada uma nova interface gráfica do usuário baseada na estrutura padrão desenvolvida pela *Public Knowledge Project (PKP)*⁸⁰. Em seguida, foi gerado um arquivo *Cascade Style Sheets (CSS)*⁸¹ definindo as propriedades visuais de cada elemento da interface. Foi realizado o *upload* do arquivo CSS, no CMS (*Content Management System*) da plataforma OJS, disponível na (URL: <<http://wrco.ccsa.ufpb.br/escience/>>).

Para a formatação do artigo científico ampliado, também se adotaram a linguagem de marcação HTML e CSS na plataforma OJS. Na estruturação dos conteúdos dos dados de pesquisa relacionados e informações adicionais, referentes ao periódico e ferramentas, uma

⁸⁰ PKP é uma iniciativa de multiuniversidade em desenvolvimento de software livre de código aberto e realização de pesquisas para melhorar a qualidade e alcance de publicação acadêmica. Disponível em: <<https://pkp.sfu.ca/>>. Acesso em: 27 jul. 2014.

⁸¹ Cascading Style Sheets ou simplesmente CSS que é uma linguagem de folhas de estilo utilizada para definir a apresentação de documentos escritos na linguagem de marcação HTML, por ser uma linguagem que possibilita o acesso através de diversos dispositivos (smartphones, tablets, desktop, etc.) e diferentes sistemas operacionais.

interface própria foi criada para a navegabilidade do *e-Quilt*, que foi desenvolvido especialmente para esse protótipo.

A interface foi dividida em três partes: menu esquerdo, central e menu direito. No menu esquerdo, agruparam-se os metadados descritivos que estão relacionados ao conteúdo do artigo, tais como: sumário, resumo, autores, filiações, método e os resultados oriundos da pesquisa em andamento. No centro, agrupou-se o artigo principal com os metadados descritivos e seus respectivos *hiperlinks*. E no menu direito, estruturaram-se os metadados administrativos que estão relacionados com as informações suplementares do artigo, as ferramentas do periódico disponíveis pela plataforma OJS, bem como as permissões referentes à propriedade intelectual conforme ilustrado na FIGURA 5.

Dessa forma, para a interface do artigo ampliado adotaram-se os objetos digitais compostos (formatos e extensões) agregados ao artigo principal. A primeira versão do protótipo *e-Quilt* pode ser visualizada na FIGURA 6 e acessível pela URL: <http://designlivre.net/artigo_ampliado/teste.html#FILE-001>.

FIGURA 5 - Interface do protótipo *e-Quilt*

The screenshot shows the e-Quilt prototype interface. At the top, there's a navigation bar with links for CAPA, SOBRE, ACESSO, CADASTRO, PESQUISA, ATUAL, and ANTERIORES. Below this is a secondary navigation bar with links for 'e-Quilt' (selected), 'e-Quilt', 'enhanced | queries | users | interoperability | law | technology'. The main content area displays an article titled 'Perfil epidemiológico da cárie dentária, condições periodontais e higiene oral em escolares do município de Caaporá - PB, Brasil'. The article is attributed to Mariana Leonel Martins, George Max de Oliveira Carvalho, and Wilton Willney Nascimento Padilha. It includes sections for 'Autores', 'Informações', and 'Resumo'. To the right of the article, there's a sidebar for 'OPEN JOURNAL SYSTEMS' with login fields for 'Login' and 'Senha', and options for 'Lembrar usuário' and 'Acessar'. Another sidebar for 'NOTificações' has links for 'Visualizar' and 'Assinar/Cancelar assinatura de notificações'. A third sidebar for 'CONTÉUDO DA REVISTA' has a search field and a 'Pesquisar' button. A fourth sidebar for 'PROCURAR' lists search options: 'Por Edição', 'Por Autor', 'Por Título', and 'Outras Revistas'. A fifth sidebar for 'TAMANHO DA FONTE' has three font size buttons. A sixth sidebar for 'INFORMAÇÕES' lists links: 'Para leitores', 'Para Autores', and 'Para Bibliotecários'.

Fonte: Elaborado pelos autores (2014).

Para vincular os diferentes objetos digitais compostos (formatos de mídia e múltiplas extensões), no conteúdo do protótipo do artigo ampliado, foram inseridos inúmeros *hyperlinks* e âncoras. As tags `<a>` `` são utilizadas para definir um *hyperlink*, podendo estar relacionadas a diversos atributos (`href`, `target`, `accesskey` e `title`).

FIGURA 6 - Interface do artigo científico do protótipo de periódico *e-Quilt*

Indice

- Resumo
- Introdução
- Seção 1
- Seção 2
- SubSeção 1
- SubSeção 2
- SubSección 3
- Sección 3
- SubSección 3.1
- SubSección 3.2
- SubSección 3.2.1
- SubSección 3.2.2
- SubSección 3.3
- Sección 4
- Sección 5
- Sección 6
- Sección 7
- Sección 8

Perfil epidemiológico da cárie dentária, condições periodontais e higiene oral em escolares do município de Caaporá - PB, Brasil

Autores

Informações

Artigo recebido em: 23 Junho 2014
Artigo aceito em: 02 Julho 2014
Artigo publicado em: 30 Julho 2014
Nome da publicação: Plataforma e-Quilt – Enhanced Publications
Identificador: 00001

Resumo

[Link para outro site \(externo\)](#) Curabitur mollis aliquam massa vehicula consequat. Vestibulum luctus ultrices faucibus. Nam venenatis mollis pellentesque. Mauris diam justo, tincidunt eget placerat sit amet, vehicula non nisl. Aliquam sagittis congue quam; sit amet congue libero feugiat eu. Maecenas hendrerit blandit augue, vel aliquam odio lobortis interdum. Nulla eleifend vehicula lacinia. Proin tristique ultrices posuere. In egestas feugiat ipsum sed vulputate. Mauris lobortis iaculis ipsum, in tristique ligula luctus nec. Nullam egel nunc non eros aliquam vestibulum. Curabitur odio a nunc tincidunt sagittis ac ac risus. Donec eget turpis est. In cursus, massa congue imperdiet consequat, tellus magna vulputate lacus, eu hendrerit dolor mauris eu lacus.

Áudio 1: Arquivo de áudio de exemplo.

Introdução

Curabitur sollicitudin purus placerat odio ultricies iaculis. Sed ac mi eros ([Referência 1](#)). Pellentesque gravida odio at accumsan sodales. Nunc dapibus laoreet mauris, id laoreet lorem molestie ac. Etiam ut tortor tempor, malesuada turpis non lobortis eff. Curabitur tempus arcu nulla, at porta tortor sollicitudin at. Ut nec mauris sed lectus luctus sagittis. Aliquam id pellentesque nulla. Vestibulum luctus, leo sit amet lacinia sodales, nibh est dapibus purus, vitae sodales risus quam eu leo. Sed suscipit a neque nec adipiscimus. Fusce sed faucibus felis. Pellentesque habitant morbi tristis senectus et netus et malesuada fames ac turpis egestas.

Seção 1

Fonte: Elaborado pelos autores (2014).

O uso comum das âncoras tem a finalidade de tornar as páginas acessíveis aos usuários. Isso proporciona uma melhor naveabilidade através da nomeação de uma região do conteúdo principal onde foi incluído um *hiperlink* no início da página que remete o leitor, diretamente, para o conteúdo nomeado, evitando que tenha que passar por cada uma das seções do artigo. A interface em HTML do protótipo *e-Quilt* está em fase de aprimoramento e agregação dos objetos, cumprindo a fase 1 do experimento. O protótipo está sendo aprimorado, para a segunda fase do experimento, com a complementação dos padrões de metadados e os protocolos de interoperabilidade adotados na ambiência da *Enhanced Publication*. A fase 2 será desenvolvida no período de agosto a novembro de 2014.

5 CONCLUSÕES

As mudanças no cenário da publicação científica estão em evidência. O protótipo *e-Quilt* criado com o intuito de apresentar um padrão de estruturação para publicação de artigo científico, com características ampliadas, atendeu ao objetivo inicial. A partir do caso da *Nature*, foi possível entender e identificar o padrão de metadados *ISA-tab* adotado na ambiência das suas publicações ampliadas no cenário internacional, motivando, assim, a estruturação preliminar do protótipo de artigo científico com características ampliadas.

Criaram-se, nessa fase, uma interface própria e uma logomarca para o *e-Quilt* com uma comunicação visual adequada à essência de uma *Enhanced Publication*. A adoção do Software OJS de publicação aberta para a hospedagem do artigo, o CSS como linguagem de

folhas de estilo, a linguagem de marcação HTML e o recurso de *hyperlink* foram essenciais para tornar o experimento acessível, com possibilidades de agregação entre os dados de pesquisa, objetos digitais e o conteúdo do artigo. Quanto ao artigo, vários objetos e dados de pesquisa foram adicionados, possibilitando, dessa feita, usar formatos com diferentes extensões, oriundos da pesquisa que foi utilizada como subsídio para o experimento. Dentre os diferentes formatos, destacam-se o resumo em áudio, dados de calibração advindos de *software*, arquivos em planilha *Excel* e outros documentos gerados para a pesquisa.

Por fim, concluiu-se, nesta fase da pesquisa, que o experimento foi realizado e o protótipo *e-Quilt* é uma publicação viável e totalmente acessível pela *Internet*.

REFERÊNCIAS

- AALBERSBERG, Ijsbrand Jan; DUNHAN, Judson; KOERS, Hylke. **Connecting scientific articles with research data:** new directions in online scholarly publishing. 2011. Disponível em: <<http://isds.nict.go.jp/wds-kyoto-2011.org/pdf/IS704.pdf>>. Acesso em: 19 jan. 2014.
- BORGMAN, C. L. **Scholarship in the digital age:** information, infrastructure, and the Internet. Cambridge: The MIT Press, 2010. (E-book).
- BORGMAN, Christine L. **From Gutenberg to the Global Information Infrastructure:** Access to Information in the Networked World. London: MIT Press, 2003. (E-book).
- BORGMAN, Christine L. The conundrum of sharing research data. **Journal of the American Society for Information Science and Technology**, v. 63, n. 6, p. 1059-1078, June 2012. Disponível em: <<http://onlinelibrary.wiley.com.ez45.periodicos.capes.gov.br/doi/10.1002/asi.22634/pdf>>. Acesso em: 17 abr. 2014.
- DANS. Disponível em:<<http://www.dans.knaw.nl/en/content/categorieen/nieuws/enhanced-publications-accessible-narcis>>. Acesso em: 20 maio 2014
- DIAS, Guilherme A. Evolução tecnológica dos periódicos científicos eletrônicos: do paradigma gutenbergiano para um cenário hipermediático. **Revista do Mestrado Profissional Gestão em Organizações Aprendentes**, v.2, n.1, p. 1-2, 2013. Disponível em: <<http://periodicos.ufpb.br/ojs/index.php/mpgoa/article/view/16996/9677>>. Acesso em: 4 maio 2014.
- JANKOWSKI, N. W. Exploring e-Science: an introduction. **Journal of Computer-Mediated Communication**, v. 12, n. 2, p. 549–562, January 2007.
- ISA COMMONS. **Isa-Tab**. Disponível em: <<http://isacommons.org/>>. Acesso em: 10 jul. 2014.
- NATURE PUBLISHING GROUP. Disponível em: <http://www.nature.com/ng/journal/v44/n2/fig_tab/ng.1054_F2.html>. Acesso em: 15 jul. 2014.

Openaire. **What is an Enhanced Publication?** Disponível em:<<https://www.openaire.eu/en/component/content/article/76-highlights/344-a-short-introduction-to-enhanced-publications>>. Acesso em: 6 dez. 2013.

RODRIGUES, Nelson de Almeida. Introdução ao METS: Preservação e Intercâmbio de Objetos Digitais. **Enc. Bibli. R. Eletr. Bibliotecon. Ci. Inf.**, n. 26, p. 1-16, 2008.

SANTOS, Plácida Leopoldina Ventura Amorim da Costa; SIMIONATO, Ana Carolina; ARAKAKI, Felipe Augusto. Definição de metadados para recursos informacionais: apresentação da metodologia BEAM. **Inf. Inf.**, Londrina, v. 19, n. 1, p. 146 – 163, jan./abr. 2014. Disponível em: <[file:///C:/Users/Adriana/Downloads/15251-73984-2-PB%20\(1\).pdf](file:///C:/Users/Adriana/Downloads/15251-73984-2-PB%20(1).pdf)>. Acesso em: 28 jul. 2014.

SALES, Luana Farias; SAYÃO, Luís Fernando; SOUZA, Rosali Fernandez Souza. Publicações ampliadas: um novo modelo de publicação acadêmica para o ambiente de e-Science. In: Encontro Nacional de Pesquisa em Ciência da Informação (ENANCIB), GT 8: Informação e Tecnologia, 14, Florianópolis, 2013. **Anais...** Florianópolis: Ancib, 2013.

TENOPIR, Carol; ALLARD, Suzie; DOUGLASS, Kimberly; AYDINOGLU, Arsev Umur *et al.* **Data Sharing by Scientists: Practices and Perceptions**. **PLoS ONE**, Volume 6, Issue 6, June 2011. Disponível em: <<http://www.plosone.org/article/info%3Adoi%2F10.1371%2Fjournal.pone.0021101>>. Acesso em: 18 jun. 2014.

VERNOOY-GERRITSEN, Marjan (ed.). Emerging standards for enhanced publications and repository technology: survey on technology. **Amsterdam University Press**. Driver and Surf studies, 2009. Disponível em: <<http://dare.uva.nl/cgi/arno/show.cgi?fid=150752>>. Acesso em: 2 dez. 2013.

ACESSIBILIDADE A INFORMAÇÃO: ANÁLISE DO WEBSITE DA FUNDAÇÃO APOIO À PESSOA COM DEFICIÊNCIA

*ACCESSIBILITY TO INFORMATION: ANALYSIS OF THE FOUNDATION OF SUPPORT
TO PERSON WITH DEFICIENCY*

Célia Medeiros Dantas
 Hellosman de Oliveira Silva
 Marckson Roberto Ferreira de Sousa

Resumo: Objetiva realizar análise de acessibilidade digital no *Website* da Fundação de Centro Integrado de Apoio ao Portador de Deficiência (FUNAD) vinculada ao portal do Governo do Estado da Paraíba. Tal análise tem como base os princípios da Arquitetura da Informação, assim como seus sistemas integrantes (sistema de rotulação, sistema de navegação, sistema de busca, sistema de organização), além da utilização da ferramenta de avaliação Hera como forma de verificar a acessibilidade do referido *Website*. Referencia e contextualiza os princípios da Arquitetura da Informação, além das características relativas à acessibilidade, como forma de facilitar o acesso das pessoas com e sem deficiência aos *Websites*. Neste aspecto, é realizada uma avaliação utilizando-se a ferramenta Hera, onde ficam evidenciadas quais as intervenções necessárias para uma melhor compreensão das necessidades informacionais dos usuários e a sociedade em geral, com sugestões de melhorias. Diante da análise, constata-se a importância da construção de *Websites* com itens de acessibilidade.

Palavras-chave: Pessoa com Deficiência. Arquitetura da Informação. Acessibilidade. FUNAD. Interação humano-computador.

Abstract: Aims to perform digital accessibility analysis of the Website of the Integrated Centre Foundation of Support to Person Bearer Deficiency (FUNAD) linked to the Government portal of the State of Paraiba. Such analysis is based on the principles of information architecture, as well as its integral systems (labelling system, navigation system, search system, organization system), in addition to the use of the Hera accessibility tool as a way of checking the accessibility of this Website. References and contextualizes the principles of information architecture, in addition to the characteristics relating to accessibility, in order to facilitate the access of people with and without disabilities to Websites. In this regard, an assessment is carried out using the Hera tool, where they are shown which interventions necessary for a better understanding of user's informational needs and society in general, with improvements suggestions. On the analysis, noted the importance of building Websites with items of digital accessibility in order to provide adequate access to information by all users regardless of their conditions.

Keywords: Person with deficiency. Information architecture. Accessibility. FUNAD. Human-computer interaction.

1 INTRODUÇÃO

A Ciência da Informação é uma área do conhecimento que tem em sua gênese o objetivo de atender as necessidades sociais de informação, independentemente do formato em que estão estocadas e disponíveis. Dessa forma, esta deve servir como base para o desenvolvimento de políticas de inclusão social, apontando para várias discussões e reflexões sobre os benefícios trazidos pelas Tecnologias de Informação e Comunicação.

Neste panorama é que se insere a discussão sobre as formas de inclusão e a acessibilidade à informação para as pessoas com deficiência, criando necessidades impostas pela própria evolução da sociedade, reforçando uma discussão sobre a conjuntura de exclusão deste segmento, num país de desigualdades históricas.

Assim como a Ciência da Informação, outras áreas, tais como Biblioteconomia, Computação, Arquivologia, que possuem como objeto de estudo a informação, tem voltado os olhares para a aplicação de novos conhecimentos, principalmente no que diz respeito à melhor forma de armazenar e recuperar as informações sempre que requisitada pelos usuários.

Nesse sentido, em um ambiente considerado cada vez mais moderno, há uma maior predisposição do homem em assimilar e aplicar as tecnologias em seu dia a dia. Essa realidade também é notável para aquelas pessoas que apresentam deficiências, independente da mesma ser temporária ou permanente. Todavia, em contrapartida percebe-se que as instituições, sejam elas de cunho público ou privado, não permitem que os usuários com deficiência possam transitar sem nenhuma interrupção durante a navegação em seus *Websites* sem sofrer algum empecilho.

Segundo Assmann (2000), a sociedade contemporânea denomina-se de sociedade da informação devido à acentuada presença e do uso intensivo das tecnologias da informação e comunicação. Para o referido autor a sociedade da informação é

a sociedade que está atualmente a constituir-se, na qual são amplamente utilizadas tecnologias de armazenamento e transmissão de dados e informação de baixo custo. Esta generalização da utilização da informação e dos dados é acompanhada inovações por inovações organizacionais, comerciais, sociais e jurídicas que alterarão profundamente o modo de vida tanto no mundo do trabalho como na sociedade em geral (ASSMANN, 2000, p. 8).

Por sua vez, percebe-se atualmente uma crescente valorização da informação e do conhecimento proporcionado e disseminado pelas tecnologias que, quando utilizadas eficientemente, permitem mesmo para aqueles públicos que possuem algum tipo de deficiência uma autonomia para suas buscas informacionais.

Quanto aos tipos de deficiências, o Decreto Federal de nº 5296/2004 em seu artigo 5º afirma que:

- **Deficiência física-** alteração completa ou parcial de um ou mais segmentos do corpo humano, acarretando o comprometimento da função física, apresentando-se sob a forma de paraplegia, paraparesia, monoplegia, monoparesia, tetraplegia, tetraparesia, triplegia, triparemia, hemiplegia, hemiparesia, ostomia, amputação ou ausência de membro,

paralisia cerebral, nanismo, membros com deformidade congênita ou adquirida, exceto as deformidades estéticas e as que não produzam dificuldades para o desempenho de funções;

- **Deficiência auditiva**- perda bilateral, parcial ou total, de quarenta e um decibéis (dB) ou mais, aferida por audiograma nas frequências de 500HZ, 1.000HZ, 2.000Hz e 3.000Hz;

- **Deficiência visual**- cegueira, na qual a acuidade visual é igual ou menor que 0,05 no melhor olho, com a melhor correção óptica; a baixa visão, que significa acuidade visual entre 0,3 e 0,05 no melhor olho, com a melhor correção óptica; os casos nos quais a somatória da medida do campo visual em ambos os olhos for igual ou menor que 60, ou a ocorrência simultânea de quaisquer das condições anteriores;

- **Deficiência mental**- funcionamento intelectual significativamente inferior à média, com manifestação antes dos dezoito anos e limitações associadas a duas ou mais áreas de habilidades adaptativas, tais como: Comunicação; Cuidado pessoal; Habilidades sociais; Utilização dos recursos da comunidade; Saúde e segurança; Habilidades acadêmicas; Lazer; e Trabalho;

- **Deficiência múltipla**- associação de duas ou mais deficiências.

Corroborando com este contexto, a Convenção Internacional dos Direitos da Pessoa com Deficiência, assinada por diversos países e promulgada pelo Congresso Nacional do Brasil com força de norma constitucional em 2009, garante às pessoas com deficiência a liberdade de expressão e de opinião e acesso à informação, respeitando-se o direito de cada pessoa em escolher ou exercer com autonomia o método de comunicação de sua preferência, assegurando-lhe o desenvolvimento de todas as suas capacidades para a vida independente.

De modo geral, as dificuldades e barreiras existentes para o pleno exercício da cidadania por parte das pessoas com deficiência são muitas, principalmente no que diz respeito ao acesso eficiente da informação em meio digital, uma vez que os Websites não se encontram aptos para atender as necessidades informacionais deste segmento da sociedade.

Este estudo avalia, através da ferramenta Hera, o Website da Fundação Centro Integrado de Apoio à Pessoa Portadora de Deficiência – FUNAD, pertencente ao Governo do Estado da Paraíba e de acordo com a Lei Estadual nº 5.208 de Dezembro de 1989 tem como principais atividades:

- O Planejamento e a coordenação, a nível estadual, a reabilitação das pessoas com deficiências;

- Dar assistência às pessoas com deficiência física, intelectual, visual, auditiva e múltipla visando ao desenvolvimento de suas potencialidades;
- Efetivação de convênios, acordos, contratos e ajustes com entidades públicas ou privadas, nacionais, e estrangeiras que objetivem a realização das pessoas com deficiência;
- A formação de pessoal técnico especializado e gestores sobre políticas de inclusão.

Para a análise metodológica do *Website* da FUNAD, foi utilizado o apporte teórico encontrado na área de Arquitetura da Informação, bem como a utilização da ferramenta automática de avaliação Hera para conteúdo web 1.0 que é recomendada por utilizar as diretrizes de acessibilidade internacional WCAG 1.0 (*Web Content Accessibility Guidelines*). Os itens de acessibilidade na web recomendados pelo WCAG 1.0 são os mais discutidos mundialmente, servindo em grande maioria, como padrão para desenvolvimentos de páginas web acessíveis em todo o mundo. A ferramenta de avaliação Hera foi escolhida por facilitar a revisão manual e permitir modificar os resultados automáticos, gerando um relatório final sobre a revisão. Ressalta-se que a ferramenta Hera, segundo sua característica foi desenvolvida para trabalhar exclusivamente com as diretrizes do WCAG 1.0.

2 ELEMENTOS DA ARQUITETURA DA INFORMAÇÃO

A construção adequada de *Websites*⁸² e blogs para as pessoas com deficiência perpassam pelo planejamento de ambientes informacionais inclusivos. Para facilitar a acessibilidade, a sua construção deve seguir os princípios da Arquitetura da Informação (AI). Portanto, a AI surge como base conceitual e tecnológica capaz de apresentar elementos eficazes ao acesso e uso com autonomia e independência de usuários com ou sem deficiência.

Em linhas gerais acredita-se que os Arquitetos da Informação têm como objetivo principal estruturar a informação em meio digital de tal modo, que o próprio *Website* seja autoexplicativo, mesmo para aqueles que nunca o visitaram e, principalmente, que permitam a eficiência quanto à recuperação da informação por parte dos usuários, independentemente de suas condições. Nessa mesma linha de pensamento Mattos (2010, p. 91) revela que o arquiteto da informação tem como função “apresentar a informação por

⁸² *Website, Site* (sítio eletrônico) pode ser compreendido como um conjunto de páginas *web*, isto é, de hipertextos acessíveis geralmente pelo protocolo HTTP na internet. **Dicionário escolar da língua portuguesa/Academia Brasileira de Letras**. 2. ed. São Paulo: Companhia Editora Nacional, 2008, p. 1187.

princípios sistemáticos e estruturados (daí o nome ‘arquitetos’), geralmente por meios visuais, de forma clara e facilmente comprehensível, ou de uma maneira totalmente nova”.

Toub (2000) afirma que a arquitetura da informação é a arte e a ciência de estruturar e organizar ambientes de informação para ajudar as pessoas a satisfazerem suas necessidades de informação de forma efetiva. Nessa perspectiva, percebe-se que a Arquitetura da Informação não pode ter sua imagem vinculada a apenas uma ciência em que necessita de um profissional da área formado para articular as informações nos *Websites*, mas é uma sensibilidade em que o arquiteto da informação constrói uma ponte entre a informação a ser disposta, o contexto em que essas informações estão inseridas e os usuários a quem se destinam essas informações.

Nesse contexto, Morville e Rosenfeld (2006) garantem que a essência da AI consiste em compreender e atender a três dimensões de variáveis: os usuários, as características do conteúdo e a especificidade do contexto de uso. De fato, verificar qual conteúdo é relevante para o usuário e o dispor de forma fácil, relacionando o contexto em que a informação está inserida e na forma como tais informações serão dispostas, tanto no que diz respeito à estética do *Website* como da clareza como se apresentam a informação, são pontos cruciais para o sucesso ou fracasso do mesmo.

Todavia, como constata Sousa (2012) a construção de *Websites* tende a seguir a visão do que os projetistas admitem como ideal e que, por essa razão, não levam em consideração as limitações existentes no público alvo, assim como suas preferências, além da falta de normatização de cores, formas, dentre outras características.

Já com um ponto de vista pragmático, Nielsen (2000) considera que a arquitetura da informação de um *Website* deve ser estruturado para espelhar as tarefas e as visões do espaço de informação ao usuário. De tal modo, o objetivo dos projetos de ambientes na internet deve ser facilitar o desempenho dos usuários em suas tarefas, pois um *Website* mal projetado dificultará a interação do usuário como o sistema informacional, que possui uma grande quantidade de opções e de facilidades. Muitos ambientes informacionais digitais são criados por meio de linguagem *HyperText Markup Language* (HTML)⁸³, sendo primordial seu mapeamento quanto a área e o tipo de uso, tipos de documento (imagens, vídeos, textos, sons), distribuição das informações na página (*frames*), conteúdo expressivo para o público-alvo a que se destina.

Neste sentido, Straioto (2002, p. 20) afirma que:

⁸³ HTML – *Hypertext Markup Language*, linguagem, na qual se baseia grande parte da programação de websites para a Internet.

A arquitetura da informação refere-se ao desenho das informações: como textos, imagens e sons são apresentados na tela do computador, a classificação dessas informações em agrupamentos de acordo com os objetivos do *site* e das necessidades do usuário, bem como a construção de estrutura de navegação e de busca de informações, isto é, os caminhos que o usuário poderá percorrer para chegar até a informação.

Camargo (2004), por sua vez, considera a arquitetura da informação como um dos fatores importantes na construção de qualquer modelo de *Website*, determinando o *design*, a disposição do conteúdo e a estratégia de navegação do usuário. A autora denota que na elaboração de uma interface o desenvolvedor deve se preocupar com o conteúdo que será inserido e como isso será adicionado à página. Tais aspectos podem auxiliar o desenvolvedor e o usuário a organizarem e estruturarem grandes quantidades de informações e envolver o atendimento às necessidades informacionais de seus usuários na interação humano-computador.

Morville e Rosenfeld (2006) delineiam os sistemas independentes que compõem a AI, sendo eles: Sistema de Navegação, no qual aponta as formas de navegação e o percurso pelo ambiente informacional; Sistema de Organização, que determina o agrupamento e a categorização do conteúdo informacional; Sistema de Rotulação, que estabelece as formas de representação, de apresentação da informação, definindo signos iconográficos ou textuais para cada elemento informativo; e Sistemas de Busca que determinam as perguntas que o usuário pode fazer e o conjunto de respostas que este irá obter, apoiados pelas estruturas de representação da informação.

Portanto, planejar ambientes informacionais digitais envolve tanto os sistemas de organização, navegação, rotulação e busca aptos a aplicação dos princípios de acessibilidade para pessoas com deficiência, como o conteúdo dos documentos em diferentes formatos, relacionados a aplicações que possuam recursos e compatibilidade com *hardware* e *software*, que possibilite a interação adequada entre sistema e usuário.

2.1 Acessibilidade à informação em ambientes informacionais digitais

Inicialmente, o termo acessibilidade estava restrito à aplicação ao ambiente construído e assinalava a eliminação de barreiras arquitetônicas. Pouco a pouco, esse sentido compreendeu outros campos do fazer humano; passamos, então, a pensar sobre a acessibilidade nas diversas áreas do convívio social como na educação, no trabalho, lazer, cultura, esportes, informação, entre outras. A ideia da acessibilidade neste estudo é possibilitar que o usuário com deficiência, independentemente de suas condições ou questões adversas, possa acessar a informação com qualidade e eficácia, traduzida no correspondente êxito na realização de suas tarefas.

Em linhas gerais, a Acessibilidade significa “permitir que pessoas com deficiências ou mobilidade reduzida participem de atividades que incluem o uso de produtos, serviços e informação”.⁸⁴

Para Torres, Mazzoni e Alves (2002) a acessibilidade é um processo dinâmico, que se associa ao desenvolvimento da tecnologia e da sociedade em estágios distintos, variando de uma sociedade para outra, conforme a atenção dispensada à diversidade humana e à época em que se encontra.

Assim, pode-se afirmar que, no que se refere à pessoa com deficiência, a garantia e a obtenção das condições de acessibilidade significa o processo de conseguir a equiparação de oportunidades. Portanto, a acessibilidade num contexto geral e especificamente, no que se refere à informação, favorece a participação social das pessoas com deficiência, e expressam políticas inclusivas.

De acordo com o *Website Acessibilidade Brasil*, o conceito de acessibilidade

representa para o nosso usuário não só o direito de acessar a rede de informações, mas também o direito de eliminação de barreiras arquitetônicas, de disponibilidade de comunicação, de acesso físico, de equipamentos e programas adequados, de conteúdo e apresentação da informação em formatos alternativos (ACESSIBILIDADE BRASIL, 2014).

Na verdade, o propósito da acessibilidade é facilitar o direito do cidadão com ou sem deficiência, em obter o acesso à informação de modo a diminuir as dificuldades existentes sejam eles por meios físicos ou digitais.

Já para Conforto e Santarosa (2002, p. 21) a acessibilidade à *Website* é:

[...] como sinônimo de aproximação, um meio de disponibilizar a cada indivíduo interfaces que respeitem suas necessidades e preferências [...]. Muitas vezes as discussões sobre acessibilidade ficam reduzidas às limitações físicas ou sensoriais dos sujeitos com necessidade especiais, mas esses aspectos podem trazer benefícios a um número bem maior de usuários, permitindo que os conhecimentos disponibilizados na Web possam estar acessíveis a uma audiência muito maior, sem com isso, prejudicar suas características gráficas ou funcionais.

Dias (2003) afirma que a acessibilidade na *Website* significa que qualquer pessoa, com qualquer tipo de tecnologia de navegação seja capaz de interagir com qualquer *Website*, e compreenda inteiramente as informações nele apresentadas.

Neste estudo, a acessibilidade está voltada para as condições de uso, com o usuário com ou sem deficiência se apresenta frente às interfaces, como essa troca deve acontecer, e,

⁸⁴ Acesso a Informação, Disponível em: <<http://www.acessoainformacao.gov.br/acessibilidade>>. Acesso em: 20 jul. 2014.

principalmente, como se dará o acesso destas pessoas às informações disponíveis. Ponderase, portanto, que as barreiras que dificultem o acesso à informação e a comunicação estejam diretamente relacionadas à ausência de elementos de acessibilidade, tratamento inadequado das informações e/ou incoerência na construção das interfaces.

3 CONSTRUÇÃO METODOLÓGICA DA PESQUISA

O uso da internet para disseminar informação é cada vez maior nos dias atuais, possibilitando o acesso a uma gama de conteúdos e serviços, de forma rápida e dinâmica. No entanto, a forma como esses conteúdos são disponibilizados afetam diretamente seu acesso.

A acessibilidade na Web tem como objetivo assegurar o acesso à informação a todos os segmentos de usuários, independentemente da deficiência, fazendo com que as tecnologias de informação e comunicação possam atingir de fato toda a sua potencialidade.

O princípio básico desse conceito é projetar ferramentas que sejam flexíveis e possam atender a diferentes necessidades, preferências e situações do usuário, permitindo que pessoas com deficiência possam ser independentes ao usarem a internet, facilitando seu uso e, respectivamente, permitindo que o acesso aos diversos *Websites* e blogs.

Conquanto, é crucial não apenas a conscientização de desenvolvedores de *Websites*, mas, principalmente, dos gestores governamentais, para que exijam acessibilidade nos *Websites* institucionais.

Segundo a *World Wide Web Consortium* (W3C), principal entidade global que trata sobre acessibilidade digital, deve-se considerar algumas dificuldades que poderão ser apresentadas pelas pessoas com deficiência no acesso as tecnologias de informação e comunicação. Dentre elas os usuários podem: não serem capazes de visualizar, ouvir, movimentar, assim como não ter acesso a certas informações com eficiência; apresentar dificuldade quanto a leitura e compreensão dos textos expostos no *Website*; ter dificuldades na utilização de mouse e teclado; possuir tela para leitura de texto de pequena dimensão, assim como uma conexão de internet lenta; não compreender ou expressar fluentemente a língua no qual o documento foi produzido; existir barreiras visuais, sonoros, motoras adversas que impedem ou interfere no acesso as informações (W3C, 2014).

Fundamentados nestes aspectos do W3C, foram desenvolvidas ferramentas para avaliar o nível de acessibilidade digital, sendo essas ferramentas denominadas de avaliadores automáticos de acessibilidade. Essas ferramentas verificam o código HTML e analisam seu conteúdo, verificando se está ou não de acordo com o conjunto de parâmetros pré-estabelecidos; após esta etapa, elas fornecem relatórios com uma série de problemas

encontrados, que devem ser corrigidos para que o *Website* ou Blog possa ser considerado com acessibilidade.

Metodologicamente, esta pesquisa analisou o *Website* da FUNAD (<http://www.funad.pb.gov.br>) que possui conteúdos informacionais voltados para o segmento das pessoas com deficiência no Estado da Paraíba, fazendo um levantamento das características de acessibilidade através da ferramenta automática de avaliação de acessibilidade – Hera (<http://www.sidar.org/hera>). Essa ferramenta possibilita a revisão de acessibilidade das páginas Web de acordo com as recomendações das **Diretrizes de Acessibilidade para o Conteúdo Web 1.0** (WCAG 1.0) que são as seguintes:

1. Fornecer alternativas equivalentes ao conteúdo sonoro e visual;
2. Não confiar só na cor;
3. Usar marcação e folhas de estilo e fazê-lo corretamente;
4. Esclarecer o uso de linguagem natural;
5. Criar tabelas passíveis de transformação em conformidade;
6. Assegurar que as páginas dotadas de novas tecnologias sejam transformadas em conformidade;
7. Assegurar o controle do usuário de alterações de conteúdo sensível ao tempo;
8. Assegurar a acessibilidade direta de interfaces embutidas;
9. Projetar o dispositivo para independência;
10. Utilizar soluções provisórias;
11. Usar tecnologias W3C e suas diretrizes;
12. Fornecer contexto e orientações;
13. Fornecer mecanismos de navegação claros e objetivos;
14. Certificar-se de que os documentos são claros e simples.

Outra característica da ferramenta de validação Hera é fornecer um relatório de qualidade e disponibilizar funcionalidades que permitem identificar claramente a localização dos erros e recomendações no código fonte.

Os erros e avisos identificados nos *Websites* são classificados de acordo com o nível de prioridade para o acesso a informação:

1. prioridades que devem ser satisfeitas inteiramente, pois impossibilitam o acesso à informação;
2. prioridades que deveriam ser satisfeitas, pois causam dificuldades a usuários ou grupo deles e;

3. prioridades que poderiam ser satisfeitas, pois não causar desconforto ou alguma dificuldade a alguns usuários ou grupo deles. Após a análise automática, foram destacados os principais pontos dos relatórios.

Sendo assim, com a avaliação pela ferramenta Hera ela possibilita um conteúdo mais acessível no meio informacional digital. Ressaltamos que esta ferramenta, assim como as outras ferramentas existentes para a avaliação quanto à acessibilidade do *Website* não exclui a avaliação realizada com as pessoas com deficiência.

3.1 Descrição da análise realizada no *Website* da FUNAD através da ferramenta Hera

O *Website* da FUNAD encontra-se associado ao portal do Governo do Estado da Paraíba na área classificada como indiretas. De acordo com análise feita, é possível verificar que o mesmo apresenta um aspecto facilitador para o manuseio e que, por essa razão, os usuários não teriam qualquer contratempo durante suas consultas.

Seguem na sequência, dispostos na FIG. 1, os resultados obtidos pela análise da avaliação automática na ferramenta Hera.

FIGURA 1 – Estado dos pontos de verificação Hera

Estado dos pontos de verificação

Prioridade	Verificar	Bem	Mal	N/A
 P1 HERA WCAG 1.0	11 ♂	--	2 ✗	4 ✓
 P2 HERA WCAG 1.0	19 ♂	1 ✓	8 ✗	1 ✓
 P3 HERA WCAG 1.0	13 ♂	1 ✓	3 ✗	2 ✓

Fonte: <http://www.sidar.org/hera/index.php.pt>. Acesso em: 20 jul. 2014

Os erros estão listados de acordo com a ordem decrescente de suas prioridades, citando as diretrizes e os pontos de verificação aos quais pertencem, bem como uma solução para o erro apresentado. De acordo com a nossa visão, seguem na sequência os principais erros diagnosticados pela ferramenta Hera, assim como prioridades, pontos de verificação e a contextualização da solução.

Diretriz: 1 - Proporcione alternativas para os conteúdos das pessoas com deficiência visual e auditiva.

Prioridade: 1

Ponto de Verificação 1.1: “Forneça um equivalente textual para todo o elemento não textual. Pode ser feito através do atributo “alt”, ou “longdesc” ou no conteúdo do elemento. Isto abrange: imagens, representações gráficas de texto, incluindo símbolos, regiões de mapas de imagem, animações, como é o caso dos GIFs animados, applets e objetos

programados, arte ASCII, painéis/frames, programas interpretáveis, imagens utilizadas em listas como sinalizadores de pontos de enumeração, espaçadores, botões gráficos, sons (reproduzidos com ou sem interação do utilizador), ficheiros de áudio independentes, pistas áudio de vídeo e trechos de vídeo”.

Erro 1 - Imagens: Há 1 imagens sem textos alternativos. Também há 18 imagens que contêm o atributo "alt". Deve verificar se os textos alternativos resultam adequados.

Solução: Cada imagem necessita obrigatoriamente do atributo "alt" com um texto que lhe sirva de legenda, nomeadamente que indique a sua função. Se o atributo "alt" resulta insuficiente, devido à complexidade da imagem, deve-se utilizar o atributo "longdesc" para especificar um endereço URL que contenha uma descrição detalhada da imagem. Nestes casos, é conveniente proporcionar também um link textual que leve ao mesmo arquivo que se indica no "longdesc" (usualmente o texto do link consiste numa letra "D" posicionada do lado direito da imagem).

Erro 2 - Elementos embutidos: Há 1 elementos <embed> e não existe nenhum elemento <noembed> na página.

Solução: Por cada elemento <embed> deve existir também um elemento <noembed> com um conteúdo alternativo que descreva adequadamente a sua ação ou substitua a sua função.

Todavia, com base nas análises da arquitetura da informação e acessibilidade, de acordo com a FIG. 2, é possível constatar algumas barreiras que serão explicitadas com a avaliação feita na ferramenta Hera.

FIGURA 2 – FUNAD – resumo de análise automática do avaliador Hera



Fonte: <http://www.funad.pb.gov.br>. Acesso em: 07 jul. 2014

Diretriz: 3 - Utilize marcadores e folhas de estilo e faça-o apropriadamente.

Prioridade: 2

Ponto de Verificação 3.2: “Crie documentos validando a notação com a gramática formal publicada”.

Erro - CSS: O código das folhas de estilo contém erros.

Solução: HERA utiliza os serviços do revisor de CSS do W3C para verificar a sintaxe das folhas de estilo. Se o resultado for indefinido, utilize o ícone para abrir a página do revisor.

Diretriz: 3 - Utilize marcadores e folhas de estilo e faça-o apropriadamente.

Prioridade: 2

Ponto de Verificação 3.3: “Use folhas de estilo para controlar a disposição dos elementos na página e a forma de apresentá-los”.

Erro 1 - Elementos de apresentação: Utilizam-se 1 elementos HTML para controlar a apresentação.

Solução: Verifique que não se utilizam elementos como ``, `<center>` ou ``, por exemplo, para controlar a apresentação da página.

Erro 2 - Propriedades de apresentação: Utilizam-se 22 atributos HTML para controlar a apresentação.

Solução: Verifique que não se utilizam atributos de apresentação como "color", "bgcolor" ou "face", por exemplo.

Diretriz: 6 - Assegure-se de que as páginas que usam tecnologias emergentes se transformam corretamente.

Prioridade: 2

Ponto de Verificação 6.4: “No caso dos scripts e dos applets, certifique-se que os eventos que o manipulam funcionam independentemente do dispositivo de entrada.”

Erro - Manipuladores de evento: Utilizam-se eventos dependentes do dispositivo e não existem eventos redundantes.

Solução: Os eventos devem poder ativar-se quer com o rato quer com o teclado. Por isso devem existir manipuladores de evento independentes do tipo de dispositivo ("onfocus", "onblur", "onselect") ou redundantes para o tipo de dispositivo (se se indica "Onmousedown" deve indicar-se também "Onkeydown", se "onmouseup" também "onkeyup", se "onclick" também "onkeypress").

Diretriz: 9 - Desenhe tendo em conta diversos dispositivos.

Prioridade: 2

Ponto de Verificação 9.3: “No caso dos scripts, especifique manipuladores de eventos por software em vez de manipuladores de eventos dependentes de dispositivos”.

Erro - Manipuladores de evento: Utilizam-se eventos dependentes do dispositivo.

Solução: Destacam-se os elementos que respondem a eventos dependentes do dispositivo, aqueles que são independentes do dispositivo e os que não podem ser emulados através do teclado. Deveriam utilizar-se sempre manipuladores de evento lógicos (i.e. baseados no software).

Diretriz: 10 - Utilize soluções provisórias.

Prioridade: 2

Ponto de Verificação 10.2: “Até que os agentes do utilizador suportem associações explícitas entre os rótulos e os controlos de formulário, para todos os controlos com rótulos implicitamente associados, certifique-se que os rótulos se encontram apropriadamente posicionados”.

Erro - Etiquetas: Há 4 controlos de formulário que devem levar etiquetas mas apenas existem 3 elementos "label".

Solução: Destacam-se as etiquetas e controlos de formulário que necessitam dessas etiquetas. Verifique que o atributo "for" de label coincide com o "id" do controle de formulário. As etiquetas devem aparecer à esquerda ou na linha imediatamente anterior aos campos de edição e à direita ou na linha imediatamente posterior às casas de verificação ou quadros de seleção.

Diretriz: 11 - Utilize as tecnologias e diretrizes do W3C.

Prioridade: 2

Ponto de Verificação 11.1: “Use tecnologias W3C quando a mesma esteja disponível e seja apropriada para uma tarefa. Utilize as versões mais recentes, desde que suportadas”.

Erro - Tecnologias do W3C: Utilizam-se elementos que não correspondem às tecnologias do W3C.

Solução: Destaca-se a DTD (caso exista) e os elementos que evidenciam o uso de folhas de estilo.

Diretriz: 12 - Proporcione informação contextual e de orientação.

Prioridade: 2

Ponto de Verificação 12.4: “Associe explicitamente os rótulos aos respectivos controles”.

Erro - Etiquetas: Faltam etiquetas para alguns controles.

Solução: Destacam-se os atributos considerados em desuso na versão HTML 4.01. (e logo, também, em XHTML).

Diretriz: 9 - Desenhe tendo em conta diversos dispositivos.

Prioridade: 3

Ponto de Verificação 9.5: “Defina teclas de atalho para links importantes (incluindo os que se encontram nos mapas de imagem client-side), controles de formulário, e grupos de controles de formulários”.

Erro - Atalhos de teclado: Não se proporcionam atalhos de teclado.

Solução: Destacam-se os elementos com o atributo "accesskey".

3.2 Resultados obtidos com a avaliação da ferramenta Hera

Fazendo uma menção sobre a análise e avaliação com a utilização do avaliador Hera, identifica-se que os principais erros de acessibilidade do *Website* da FUNAD giraram em torno de problemas como, a necessidade de fornecer um equivalente textual para cada imagem disposta na página (incluindo representações gráficas do texto, símbolos, gifs animadas, imagens utilizadas como sinalizadores, pontos de enumeração, botões gráficos e espaçadores).

Isso é importante porque para uma pessoa com deficiência acessando o *Website*, fazendo uso de um leitor de tela, sem esse equivalente textual, não será possível identificar do que trata a imagem e esta poderá deixar de receber a informação que deseja.

Um evento observado nesta análise referente ao avaliador Hera foi o de que, ao mostrar os resultados obtidos, seus resultados ele nem sempre expressam uma explicação clara e objetiva para os erros assinalados. Entretanto, o problema principal trata-se do fato avaliador restringir-se a uma verificação da sintaxe do *Website*, delimitando-se simplesmente a analisar as diretrizes do WCAG, apresentando uma lista de não conformidade com os padrões de acessibilidade na web.

Outro problema é que os ambientes informacionais digitais não identificaram o principal idioma utilizado em suas páginas. Em termos dos avisos, destacam-se o fato de que é necessário assegurar que toda informação provida com ‘cor’ possa ser visualizada

também ‘sem cor’, onde é necessário verificar a combinação de cores entre fundo e primeiro plano para evitar cores discrepantes que causem ilusão de ótica ou que tragam problemas para pessoas que sofrem de daltonismo. Nesse sentido, é necessário criar uma sequencia lógica de tabulação para percorrer os itens de formulário, para que possam ser percorridos com o uso do teclado.

Como foi possível verificar, as barreiras informacionais constatado no *Website* da FUNAD são de fácil solução, mas que sua existência dificulta o acesso por parte dos usuários assistidos na própria fundação. Por essa razão, se tal instituição tem como propósito habilitar, assim como reabilitar e profissionalizar as pessoas com deficiência para o mercado de trabalho é de se esperar que seu *Website* permita o acesso sem qualquer barreira de acessibilidade.

4. CONSIDERAÇÕES FINAIS

Essa pesquisa realizada tem como enfoque principal não apenas constatar as barreiras informacionais apresentadas no *Website* da FUNAD para a comunidade, mas apresentá-las aos próprios gestores da fundação com vista em favorecer o desenvolvimento de ambientes informacionais digitais com acessibilidade. Nesse aspecto, no ano de 2013, durante evento realizado na própria fundação denominado “II Seminário Técnico Científico - FUNAD” foi apresentado os primeiros resultados sobre a avaliação com base na literatura da Arquitetura da Informação, onde o corpo diretivo da instituição mostrou-se disposto a executar as modificações para contribuir na disseminação da informação digital para os usuários e toda a sociedade.

Considera-se com esta pesquisa, que o acesso ao ciberespaço pelas pessoas com deficiências se depara em barreiras existentes principalmente na concepção dos Websites. Na sua maioria este artefato de informação é construído exclusivamente por profissionais da área de computação, que priorizam os aspectos técnicos e não seu uso e acesso. Entretanto, os métodos e ferramentas de avaliação que envolve diretamente as pessoas com deficiência são necessários para qualidade e implantação dos sistemas. É importante ressaltar que ambas as avaliações, a automática e com o usuário com deficiência se complementam.

Portanto, o acesso à informação é um processo amplo, que transcorre pela incorporação do uso da acessibilidade na web pelos cidadãos com ou sem deficiência e, não apenas pelas questões de conhecimentos e habilidades necessárias para a utilização dos recursos disponibilizados, contribuindo para sua inserção nos espaços sociais e consequentemente na construção de sua cidadania. Fundamental, será a transformação

destas pessoas em indivíduos independentes, que constroem seus conhecimentos de forma individual, contribuindo para o conhecimento coletivo, a partir da análise e da transformação das informações, com ressignificação do paradigma informacional.

Neste sentido, a acessibilidade à informação caracteriza-se como capaz de possibilitar a inclusão digital de segmentos sociais excluídos da Sociedade da Informação. Acreditamos que os desenvolvedores de *websites*, analisem o contexto de implantação destes ambientes para promover a inclusão dos diversos tipos de usuários, dentre os quais as pessoas com deficiência e suas complexidades, diante dos espaços sociais nos quais estão inseridos.

A inserção social das pessoas com deficiência nas tecnologias de informação e comunicação deve ser efetivada de forma autônoma e independente com condições ampliadas de acesso e uso à internet, assim como dos demais usuários potenciais que podem atender, independente de suas condições físicas, sensoriais e cognitivas.

REFERÊNCIAS

ACESSIBILIDADE BRASIL. O que é acessibilidade. 2014. Disponível em: <<http://www.acessobrasil.org.br>>. Acesso em: 29 jul. 2014.

ASSMANN, H. A metamorfose do aprender na sociedade da informação. **Ciência da informação**, Brasília, v. 29, n. 2, p. 7-15, 2000.

BRASIL. Acesso a Informação. Novo modelo de identidade digital padrão do governo federal atende às principais recomendações de acessibilidade indicadas para web. 2014. Disponível em: <<http://www.acessoainformacao.gov.br/acessibilidade>>. Acesso em: 20 jul. 2014.

BRASIL. Decreto Nº 5.296, de 02 de dezembro de 2004. Disponível em: <http://www.planalto.gov.br/ccivil_03/_ato2004-2006/2004/decreto/d5296.htm>. Acesso em: 23 jul. 2014.

BRASIL. Decreto nº 6.949, de 25 de Agosto de 2009. Promulga a Convenção Internacional sobre os Direitos das Pessoas com Deficiência e seu Protocolo Facultativo, assinados em Nova York, em 30 de março de 2007. Disponível em: <http://www.planalto.gov.br/ccivil_03/_ato2007-2010/2009/decreto/d6949.htm> Acesso em: 13 jan. 2013.

CAMARGO, L. S. A. Arquitetura da Informação para biblioteca digital personalizável. 2004. 142 f. Dissertação (Mestrado em Ciência da Informação) - Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília.

CONFORTO, D. SANTAROSA, L. M. C. Acessibilidade à Web: Internet para todos. **Revista de Informática na Educação:** Teoria, Prática, Rio Grande do Sul, v.5, n. 2, p. 01-21, 2002.

DIAS, C. **Usabilidade na web:** criando portais mais acessíveis. Rio de Janeiro: Alta Books, 2003.

MATTOS, A. N. **Informação é prata, compreensão é ouro:** um guia para todos sobre como produzir e consumir informação na Era da Compreensão. [S.l.: s. n.], 2010.

MORVILLE, P.; ROSENFELD, L. **Information Architecture for the World Wide Web.** O'Reilly Media, 2006.

NIELSEN, J. **Projetando Websites.** Rio de Janeiro: Campus, 2000.

PARAÍBA. **Lei nº 5.208, de dezembro de 1989.** Autoriza o Poder Executivo a instituir a Fundação Centro Integrado de Apoio ao Portador de Deficiência (FUNAD) e dá outras providências. Disponível em:

<<http://www.asdef.com.br/innova/assets/leiestadual/5208criacaofunad.pdf>>. Acesso em: 20 ago. 2014.

STRAIOTO, F. **A Arquitetura da Informação para a World Wide Web:** um estudo exploratório. 2002. 120f. Dissertação (Mestrado em Ciência da Informação) - Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília.

SOUZA, M. R. F. O Acesso a Informações e a Contribuição da Arquitetura da Informação, Usabilidade e Acessibilidade. **Informação & Sociedade**, João Pessoa, v. 22, p. 65-76, 2012.

TORRES, E. F., MAZZONI, A. A. e ALVES, J. B. M. A Acessibilidade à Informação no Espaço Digital. **Ciência da Informação**, Brasília, v. 31, n.3, p.83-91, 2002.

TOUB, S. **Evaluating Information Architecture:** A practical guide to assessing Website organization, 2000.

W3C. **World Wide Web Consortium**, 2014. Disponível em: <<http://www.w3.org>>. Acesso em: 29 jul. 2014.

UMA FERRAMENTA PARA RECUPERAÇÃO DE TAGS DE BLOGS BASEADA EM MICROFORMATOS

A TOOL TO RECOVER TAGS OF BLOGS BASED ON MICROFORMATS

Célio Andrade Santana Júnior

Nilton Heck Santos

Steffane Ramires de Lima

Amanda Maria de Almeida Nunes

Resumo: O objetivo deste trabalho é apresentar a ferramenta Microgisbone que realiza a recuperação da informação de blogs na Internet utilizando como referência o padrão de microformato rel-tag. A ferramenta proposta tem como princípios (i) a adequação aos padrões de Big Data, (ii) utilizar uma arquitetura de serviços escalável para que seja possível utilizar computação em nuvem a partir de Infraestrutura como Serviços (IaaS) e (iii) que esteja disponível como um serviço de informação a comunidade. Para validar o funcionamento da ferramenta e escolher a arquitetura de dados foi realizado um experimento utilizando um cenário real, blogs hospedados pelo wordpress.com, onde as marcações (tags) foram coletadas por um período de 3 dias. Ao término deste experimento, onde foram coletadas cerca de 6,6 milhões de tags, foram desenvolvidos alguns serviços de informação, baseados nas marcações coletada. Foi observado que, de fato, um volume relativamente grande de informação foi recuperado de uma quantidade pequena de blogs e de um tipo de informação que é pequeno (marcações). Foi observado também padrão rel-tag dos microformatos tornam mais simples a identificação e recuperação das marcações nos blogs por máquinas se comparados com os mecanismos formais de web-semântica.

Palavras-chave: Micro Formatos. Blogs. Marcações. Produtos de Informação.

Abstract: This paper aims to present Microgisbone tool that performs the retrieval of information from Internet blogs using the standard rel-tag microformat. The proposed tool has the following principles: (i) the adequacy of the standards of Big Data, (ii) use an architecture for scalable services to be able to use cloud computing from Infrastructure as a Service (IaaS) and (iii) that is available as an information service to the community. To validate the tool efficacy and choose the data architecture we conducted an experiment using a real scenario, blogs hosted by wordpress.com, where the tags were collected for a period of 3 days. At the end of experiment, which were collected about 6.6 million tags, some information services, were developed. It was observed that, in fact, a relatively large volume of information was recovered from a small amount of blogs and a kind of information that is small (tags). It was also observed pattern of rel-tag microformat simplify the identification and retrieval of tags in blogs by machines compared to the formal mechanisms of semantic-web.

Keywords: Microformats. Blogs. Tags. Information Products.

1 INTRODUÇÃO

O rompimento da bolha das empresas “.com”, no Outono de 2001, marcou um ponto de transformação da Internet que não ocorreu apenas influenciou no paradigma tecnológico vigente, mas principalmente, na forma como os negócios eram construídos na *web* (O'REILLY, 2007). Esta ruptura no modo de produzir e consumir conteúdo atraiu a atenção de profissionais e pesquisadores, levando à produção de ferramentas e pesquisas que

promoveram mudanças na Internet e transformaram naquilo em que conhecemos hoje. A percepção de uma evolução da *Web*, com base principalmente na maior interação entre usuários, levou ao surgimento do termo “*Web 2.0*”, que representa a evolução dos produtos e serviços na Internet (O'REILLY, 2007).

Esta nova versão da *Web*, quando comparada à sua predecessora, destaca-se pela compreensão do ambiente digital como uma plataforma, isto pois, enquanto “a *Web 1.0* limitava-se ao acesso e raras contribuição dos usuários” (MURUGESAN, 2007), enquanto a *Web 2.0* apresenta um modelo completamente novo, focado nas relações entre os indivíduos e a sociedade digital. Por conta desta possibilidade de maior participação dos indivíduos, esta também pode ser chamada de “*Web da Sabedoria*”, “*Web Centrada nas Pessoas*”, “*Web Participativa*” ou “*Read/Write Web*” (MURUGESAN, 2007). De certo modo, se a Internet e a *Web 1.0* poderiam ser consideradas “a rede mundial de computadores”, então, a *Web 2.0* pode receber a alcunha de “a rede mundial de pessoas”. (CASSANO, 2008).

O surgimento da *Web 2.0* transformou o perfil de seus usuários que passaram de passivos consumidores de conteúdo para produtores e consumidores de informação (PETERS; STOCK, 2007). Nessa transformação, Berners-Lee (2007) destaca o papel os *blogs*, uma vez que estes representam uma das ferramentas mais relevantes em colocar o usuário no controle de sua própria produção.

O próprio Bernes-Lee criou, em 1992, a página “*What is News*” que é considerado o primeiro *blog*, pelo menos da forma como o conhecemos hoje. Mas, apenas muito depois, em 1999, o termo “*blog*” foi criado por Peter Merholz derivado do termo “*web log*” que se tornou “*we blog*”. A popularização dos *blogs* foi quase instantânea tanto que em 2002, um serviço de *blogs* chamado Blogger.com já apresentava mais de dez milhões de postagens em suas páginas. (FU, 2007).

Linton (2002) afirma que os *blogs* caracterizam-se pela organização cronológica de suas publicações, dentro de uma estrutura onde, quase sempre, há somente uma página. Alexander (2006) destaca a diferença implícita na retórica dos *blogs*, quando comparada à sítios comuns na *Web 1.0*. A retórica mais pessoal, amplamente aplicada em *blogs*, culminou no surgimento de um público alvo específico, impelindo ao surgimento de práticas como o “*Blogrolling*”, que se trata de uma listagem de endereços de outros *blogs* que o autor de um determinado *blog* recomenda e que formam uma espécie de rede de interesse. Outro fenômeno observado foi a popularização da utilização de folksonomias na organização do conteúdo digital, colocando o usuário no papel de arquiteto informacional.

O processo de utilização de folksonomias na indexação de conteúdo digital dentro da *Web 2.0* recebe o nome de etiquetagem, ou *tagging* (LUND *et al.*, 2005; VANDER WAL, 2006) e consiste na atribuição livre de termos para representação de uma publicação. Deste modo, os termos atribuídos pelo autor (folksonomia estreita) ou pelos usuários (folksonomia ampla) em uma publicação, tem a finalidade de representar que assunto está sendo tratado naquele canal. A complexidade desta tarefa está expressa nos números, de acordo com estatísticas do Wordpress (2014), cerca de 40,5 milhões de novas publicações são realizadas todos os meses e mais de 409 milhões de pessoas leem publicações em *blogs* hospedados/construídos por esta ferramenta mensalmente.

O fenômeno da descentralização do processo de indexação trazido pela popularização da etiquetagem de recursos em *blogs*, que antes deveria ser feito, teoricamente, por um profissional da informação e é agora atribuído ao usuário, trouxe um grau considerável desorganização neste tipo de informação. Somado a isso o ambiente amplamente fluido e dinâmico dos *blogs* e grande o volume de informação, e de pessoas, que acessam esses serviços, tornou-se quase impossível recuperar conteúdo relevante nestes canais. Hewitt (2005) afirma que muitos destes *blogs* ainda se encontram inacessíveis para grande parte do público, particularmente para aqueles usuários que utilizam mecanismos de busca.

Esta dificuldade de encontrar conteúdo relevante na Internet a partir de sistemas de recuperação da informação (SRI), não foi percebido apenas nos *blogs*, mas em diversos contextos que começaram a se tornar cotidianos na vida dos usuários de Internet tais como informações pessoais, calendário, conexões sociais entre outros. Neste contexto surgiram os microformatos, criados por Tantek Çelik com o intuito de tornar os itens de dados existentes nas páginas da Internet reconhecíveis. A proposta é que esses elementos sejam passíveis de processamento automatizado por agentes de software e que também sejam diretamente legíveis por seres humanos (MENDEZ *et al.*, 2008). Segundo Wharton (2005), a motivação para a criação dos microformatos foi atender necessidades, primeiramente, dos seres humanos e em segundo plano das máquinas.

A justificativa para este trabalho é que considerando as dificuldades encontradas pelos SRI no tratamento e processamento de informações não-estruturadas em *blogs* na *web*, em especial os meta-dados atribuídos pelos usuários a partir das etiquetas (*tags*), o presente trabalho tem como objetivo apresentar a ferramenta MicroGisbone, elaborada para obtenção automática de etiquetas em *blogs* a partir de marcações que utilizem o padrão de microformatos. A ferramenta será avaliada a partir de um experimento onde será recuperada

estas informações de *blogs* reais hospedados em serviços que suportem os microformatos. A partir dos dados coletados, serão desenvolvidos alguns serviços de informação para posterior uso.

Para uma melhor compreensão do contexto desta pesquisa, o presente trabalho encontra-se dividido em cinco seções e além desta Seção introdutória, a Seção 2 apresenta a fundamentação teórica que contempla os conceitos relacionados a esta pesquisa. A Seção 3 apresenta a ferramenta Microgisborne e seu processo de concepção. A Seção 4 apresenta o experimento. Por fim, na Seção 5 serão apresentadas as considerações finais.

2 LITERATURA PERTINENTE

Antes do início da construção da ferramenta se fez necessária uma revisão da literatura referente ao uso de etiquetas (*tags*) em *blogs* o uso de microformatos nestes ambientes. O resultado desta revisão é apresentado nas subseções a seguir.

2.1 Organização do Conteúdo em Blogs

Murugesan (2007) afirma que os *blogs* se tornaram populares pela sua forma simples e por possibilitarem que pessoas expressem seus pensamentos, ideias, sugestões e comentários na Internet. Rodzvilla (2002) argumenta que os *blogs* conceberam ao indivíduo a chance de satisfazer um desejo natural de se comunicar em sociedade. A consequência deste sucesso foi a explosão informacional que trouxe a tona a necessidade em organizar esta informação pelas empresas de marketing e monitoramento de mídias sociais (TREMAYNE, 2007).

No contexto dos *blogs*, a práticas tradicionais de organização da informação e do conhecimento não são aplicadas. Ao invés disso, são utilizados modelos de classificação natural, descentralizados, independentes de modelos pré-estabelecidos. A proposta dos *blogs* é, dar ao usuário a liberdade de organizar a sua produção de forma pessoal (RUSSEL, 2005). Dar ao usuário o controle sobre o processo organizacional não implica necessariamente em uma organização efetiva, ainda que este indivíduo possua, supostamente, um maior conhecimento semântico sobre a sua publicação (SOUZA; ALVARENGA, 2004).

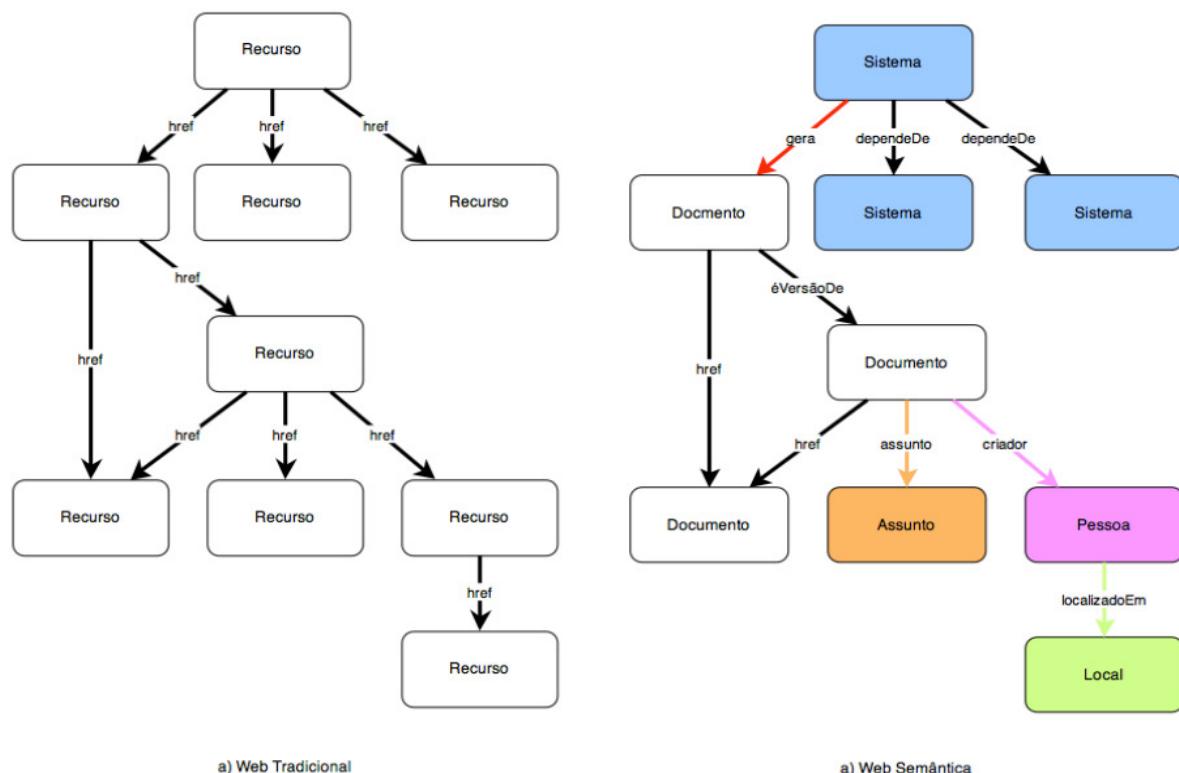
Cenários como este apresentam um grande desafio para que sistemas computacionais realizem o processo de mineração deste ambiente (*web mining*) uma vez que para isto se exige a existência de estabelecimentos semânticos no código fonte das páginas para auxiliar na atribuição de um significado àquilo que foi publicado (BAEZA-YATES; RIBEIRO NETO, 2013). Estes sistemas precisam deste auxílio para realizar a

identificação de conteúdo e uma das abordagens mais utilizadas para isso é a Web Semântica (KIM *et al.*, 2009).

A Web Semântica permite que os dados contidos na Internet passem a ser compreendidos por computadores e esta é baseada na descrição dos recursos da Internet de forma comprehensível pelos computadores. Esta descrição pode ser obtida através da anotação de meta-dados, agregando dados a outros pedaços de dados (OREN *et al.*, 2006).

Deste modo, ao aplicar a Web Semântica, estes *blogs* são atendidos por um modelo mais inteligente de organização que prevê não apenas a organização interna, na forma de categorias ou marcações, mas também abre portas para modelos de extração externos que elevam o nível da organização para fora do ambiente do *blog*, apresentando um possível adjacente para novos modelos de organização, recuperação e representação da informação (WITTEN; FRANK; HALL, 2011). A FIGURA 1 apresenta um paralelo entre um conjunto de documentos organizados através de hiperlinks tradicionais (a) e outro utilizando os microformatos (b). Podemos perceber a dificuldade de prover significado a um elemento no modelo tradicional uma vez que não é possível determinar qualquer tipo de relação entre os recursos relacionados que só se utiliza do recurso de hiperlinks (*href*) que impossibilita a inferência informações.

FIGURA 1 – Comparativo dos modelos de organização Tradicional e de Web Semântica



Fonte: adaptado de Koivunen e Miller (2001)

2.2 Etiquetagem

Tradicionalmente o papel de catalogação e organização da informação em grandes unidades de informação e conhecimento, como as bibliotecas, é atribuição inquestionável de profissionais da informação. Contudo, estes profissionais, na maioria dos casos, lidam com ambiente e recursos padronizados que em nada se parece com o ambiente da Internet. Na rede, incontáveis recursos são produzidos diariamente e estes são criados nos mais diversos formatos (vídeo, texto, áudio, foto, etc.), e são organizados de formas completamente diferentes uns dos outros. Isto é um reflexo do aumento latente do conteúdo disponível, causado pela popularização do acesso e da produção de conteúdo e empregar profissionais de informação para catalogar e organizar o conteúdo da Web seria uma tarefa praticamente impossível (PRESSLEY, 2005).

Na Internet, muitos dos serviços disponibilizados adotaram um modelo de atribuição de palavras-chave que possibilita o usuário a atribuir etiquetas (*tags*) a recursos na *web* sem a necessidade de um vocabulário controlado (MARLOW *et al.*, 2006). Este processo, denominado *Tagging*, propõe que usuários comum tornem-se o ator principal do processo de organização, consiste na associação de palavras-chave a um objeto ou item na Internet com o intuito de torná-los acessíveis de forma menos complicada para o usuário final (KIM *et al.*, 2011).

Marlow *et al.* (2006) afirmam que:

a utilização de tags no processo de organização apresenta potencial para melhorar o resultados de buscas, detectar spams, desenvolver sistemas de reputação e organização pessoal, enquanto possibilita novas possibilidades de comunicação e oportunidade de mineração de dados. Ainda, naquilo que tange sua concepção (MARLOW *et al.*, 2006).

Kim *et al.* (2011) define que um modelo de marcação é composto de um conjunto de conceitos:

- “Objeto: o recurso que está sendo etiquetado. Por exemplo, um livro, uma foto, ou uma publicação de *blog*, etc.;
- *Tag*: a etiqueta associada ao recurso.
- Etiquetador: o agente – normalmente uma pessoa – que cria a ligação entre o objeto e a etiqueta.
- Fonte: o espaço onde o ato de etiquetar foi realizado. Por exemplo: Twitter.
- Polaridade: um voto contra ou à favor a declaração da etiqueta com o objetivo de solucionar problemas com *spams*.

A utilização de *Tagging* pode ser encontrada na web em duas vertentes: (a) através do uso colaborativo, ou amplo, e (b) através do uso restrito, ou estreito. No uso colaborativo, também conhecido como *social tagging*, há a interação social entre os

usuários de um sistema para a atribuição de etiquetas a um determinado recurso. Esta abordagem exige a existência de modelos de classificação que mensurem a relevância de uma etiqueta com base na quantidade de usuários que atribuem a mesma *tag* a um mesmo recurso, obtendo uma espécie de ranking (SINCLAIR; CARDEW-HALL, 2007). Este modelo de *Tagging* tem se tornado popular na Web em diversos serviços como por exemplo os *trend topics* do Twitter. (AMARAL; AQUINO, 2008) .

O conceito de folksonomia ampla de Varder-Wal (2013) é resultado do processo de etiquetagem a partir de um conjunto de palavras-chave que representam como um grupo de indivíduos classifica suas informações dentro de um contexto. Outro modelo possível de ser encontrado na *web* é aquele baseado no uso restrito, onde um único usuário, ou uma comunidade bastante reduzida, é responsável por atribuir etiquetas a um recurso, não havendo interação por parte dos outros usuários.

Guimarães (2008) afirma:

Este modelo não possui a finalidade de ser interativo, mas representar como o autor do produto informacional classifica a sua produção, com a intenção de apresentar ao seu consumidor o conteúdo daquilo que se produziu, uma vez que o autor possui um conhecimento semântico profundo sobre aquilo que foi criado (GUIMARÃES, 2008).

2.3 Microformatos

Segundo Khare e Elik (2006)

Para auxiliar na organização e recuperação da enorme quantidade de conteúdo que apresenta baixa semântica na web, especialmente aqueles advindos de blogs e ambientes de produção colaborativa, foram elaborados os microformatos sendo estes um conjunto de padrões que buscam “facilitar a descrição de pessoas, lugares, eventos e outras formas comuns de elementos semi-estruturados para linguagem humana” (KHARE; ELIK, 2006).

O conceito delimitado por Khare e Elik (2006) é fundamentado na descrição oferecida no próprio sítio da Microformatos: “Feito para humanos primeiro e máquinas em segundo”. Os microformatos são um conjunto simples de formatos abertos de dados construída sobre a premissa de padrões existentes e já amplamente utilizados (Microformats, 2014).

A primeira menção aos microformatos ocorreu durante a conferência *South By Southwest* (SXSW) em 2004 e, desde então, “a utilização de microformatos vem ganhando adeptos rapidamente entre produtores de informação e desenvolvedores de serviços, grandes e pequenos, nos últimos dois ou três anos” (ALLSOP, 2007).

Para a utilização dos microformatos é necessária a existência de recursos semi-estruturados, tais como as linguagens de marcação *hipertexto mark-up language* (HTML) e

extensible mark-up language (XML), e através do uso de recursos já adotados, como o atributo *class*, são inseridos novos descritores contendo os microformatos responsáveis pelo valor semântico do documento (KHARE; ELIK, 2006). Um modelo de aplicação pode ser visualizado na FIGURA 2, onde há a descrição sintática de um evento acrescido do conteúdo semântico (em negrito).

FIGURA 2 – Exemplo de utilização do padrão hCalendar em microformatos para a definição de um evento em calendários

```
<div class="vevent">
  <a class="url" href="http://conferences.oreillynet.com/pub/w/40/program.html">
    http://conferences.oreillynet.com/pub/w/40/program.html
  </a>
  <span class="summary">Web 2.0 Conference</span>;
  <abbr class="dtstart" title="2005-10-05">October 5</abbr>-
  <abbr class="dtend" title="2005-10-07">7</abbr>,
  at the <span class="location">Argent Hotel, San Francisco, CA</span>
</div>
```

Fonte: Os Autores

A grande vantagem dos microformatos para a organização e recuperação da informação na Internet reside, não em uma nova tecnologia, mas no reuso daquelas já utilizadas. Deste modo, facilitando a sua adoção. Neste sentido, Allsop (2007) argumenta à favor da importância dos microformatos para a Web Semântica: “os Microformatos são um conjunto de abordagens para resolver o importante problema da falta de produção de marcações com estruturas semânticas na web de hoje”.

De acordo com Kim *et al.* (2011) a semântica aplicada aos microformatos é menos poderosa quando comparada à capacidade de outros modelos tais como o *Resource Description Framework* (RDF). Mas, Khare e Elik (2006) sugerem que a utilização de um arquivo externo, como no caso do RDF, para atribuição de semântica em ambientes como a Blogosfera pode não funcionar por exigir um conhecimento específico do produtor da informação.

Khare (2006) destaca que existem quatro características que auxiliam na popularização de microformatos: (a) ser redutível, ou seja, ser utilizado somente quando há real necessidade, (b) ser reutilizável, não sendo necessária a definição de cada atributo sempre que este for utilizado, (c) ser reciclável, podendo ser utilizado em diversos recursos tais como *Feeds* sejam estes *Rich Site Summary* (RSS) ou *Atom* e *blogs*, e (d) ser representável e analisável, isto é, os meta-dados estão sempre visíveis, tanto para a máquina

quanto para o ser humano, embora sejam vistos de formas diferentes, proporcionando, por exemplo, análises automáticas do conteúdo.

Hoje, muitas são as classes, também conhecidas como vocabulários, criadas pela comunidade Microformats.org para atribuição de semântica na *web* com base nos microformatos, entre elas a hCalendar, hCard e rel-license, descritas a seguir (Microformats, 2014).

- **hCalendar:** Formato aberto de padronização para a descrição de eventos na web com base no padrão iCal utilizado em diversas aplicações de calendário. Este padrão permite que sistemas de busca na web sejam capazes de identificar um evento (no passado, no presente, ou no futuro) e converte-los de modo a serem automaticamente incorporados a ferramentas de calendário como, por exemplo, o Google Calendar.
- **hCard:** Este formato é utilizado para permitir que sistemas de busca sejam capazes de identificar pessoas ou empresas que estejam listadas em páginas na Internet, auxiliando na localização deste tipo de informação. Um serviço que utiliza este microformato é o “logar com Facebook” onde um usuário pode utilizar as suas informações contidas no Facebook e utilizá-las para se cadastrar em outros serviços.
- **rel-license:** Este padrão tem como objetivo delimitar espaços onde há a indicação do tipo de licença atribuída ao recurso. Se é livre, aberta, proprietária ou mesmo *copyleft*.
- **rel-tag:** Formato responsável por reconhecer as etiquetas em diversos serviços.

Outros vocabulários definidos são: rel-nofollow (comunicação com agentes), VoteLinks (concordância com links), XFN (relacionamentos entre pessoas), XMDP (metadados), XOxo (especificação de HTML), adr (endereços), geo (coordenadas geográficas), hAtom (informações semânticas), hAudio (conteúdo de áudio), hListing (listas abertas e distribuídas), hMedia (imagens, vídeo e áudio), hNews (informações semânticas em notícias), hProduct (informações padronizadas sobre produtos), hRecipe (receitas culinárias), hResume (currículos), hReview (revisões), rel-directory (indica que o destino de um hyperlink é um diretório), rel-enclosure (anexos), rel-home (indica um link para a página principal de um site), rel-payment (mecanismo de pagamento), robotsexclusion (informações para robôs de busca), xFolk (marcações sociais).

3 MICROGIBORNE

Com o objetivo elaborar um produto/serviço de informação que utilizasse a informação contida nas etiquetas criadas pelos usuários em blogs, os autores decidiram utilizar os padrões de microformatos para auxiliar nessa coleta baseado no vocabulário “*rel-tag*”. Para tanto foi montado um experimento em um cenário real. Faz-se importante, ainda, ressaltar que a ausência de ferramentas para este tipo de problemática torna razoável a hipótese de pioneirismo desta pesquisa, principalmente devida sua relevância para a Organização e Recuperação da Informação.

Desta forma, para realização do experimento proposto nesta pesquisa, foi elaborada a ferramenta Microgisborne concebida como uma ferramenta acessível através da Web com o objetivo de processar e recuperar meta-dados, mais especificamente *tags*, de postagens de *blogs* que utilizem a classe de atributos *rel-tag* provenientes da semântica aplicada pelos microformatos. Neste sentido, a aplicação busca apresentar uma forma de se extrair este conhecimento através dos microformatos.

A ferramenta utiliza-se de algoritmos desenvolvidos em linguagem *Hypertext Preprocessor* (PHP), utilizando o framework Slim como base para efetuar a leitura dos microformatos dos documentos na Internet e recuperar para o usuário quais foram as *tags* atribuídas pelo produtor da informação para rotular uma postagem, tornando, assim, possível uma análise posterior do método de organização, ou mesmo a realização de inferências menos óbvias.

Os dados recuperados pela ferramenta são armazenado através do uso de bases de dados. Havia uma dúvida se a base de dados mais adequado seria a MySQL (livre e amplamente utilizada) ou a Google Big Query (livre e que suporta aplicações de *big data*). A intenção desta ferramenta é que ela seja aberta e acessível a outros desenvolvedores e para isso foi disponibilizada uma interface para aplicação de programas (API) para a utilização deste serviço de informação⁸⁵. A arquitetura do sistema utiliza a arquitetura de serviços REST, também adequada para escalonamento de usuários se aplicado em infraestrutura como serviço (IaaS), isso significa que se 1 ou 1.000.000 de usuários utilizarem o sistema de forma simultânea, o produto estará apto a realizar esse escalonamento de forma automática. A FIG. 3 apresenta as relações, métodos, e técnicas utilizadas para elaboração do serviço.

O modelo inicial o MicroGisborne foi delineado após grande dificuldade dos autores em encontrar outra ferramentas que tornassem possível a captura dos meta-dados utilizados

⁸⁵ <http://docs.microgisborne.apiary.io/>

no *blog* Brainsorm9.com.br, objetivando uma análise da organização informacional do *blog*, inicialmente quantitativa. As questões estabelecidas pela pesquisa giravam em torno da utilização de etiquetas no *blog* e eram as seguintes:

- Quantas *tags* os autores do *blog* utilizam em média para descrever seu conteúdo?
- Quais são as *tags* que mais se repetem?
- Os autores reutilizam *tags* ou tendem a criar novas *tags*?
- Qual o total de *tags* utilizadas pelos autores?

FIGURA 3 – Arquitetura do Microgisborne



Fonte: Os Autores

A ferramenta já está disponível no endereço www.microgisborne.com

4 EXPERIMENTO

O cenário escolhido para realização deste experimento foram os *blogs* presentes dentro das categorias “*News*⁸⁶” e “*Music*⁸⁷” do sítio da plataforma Wordpress. Estes blogs que utilizam a plataforma Wordpress foram escolhidos pois o suporte aos microformatos é uma configuração padrão em todos os blogs derivados deste serviço, assim, sendo garantida a consistência dos dados. O experimento foi realizado para avaliar três objetivos (i) se a utilização de microformatos era adequada para a recuperação desse tipo de informação, (ii)

⁸⁶ <http://en.wordpress.com/tag/news/>

⁸⁷ <http://en.wordpress.com/tag/music/>

saber qual base de dados seria mais adequada ao projeto, se o MySQL (relacional) ou Google Big Query (suporte a *big data*) e (iii) verificar se a arquitetura escolhida era escalável para múltiplos acessos.

Após a escolha do escopo a ser analisado, foi colocado o Microgisborne para executar o primeiro passo da recuperação das *tags* que é encontrar o endereço (URL) dos *blogs* a serem varridos. O *Wordpress* disponibiliza um arquivo .xml para cada categoria (*music* e *news*) onde se fez necessário buscar a marcação “*blog-url*” que apresentava a URL específica de um *blog*, e ao percorrer todo o arquivo xml, todos os *blogs* daquela categoria estariam mapeados. Esta primeira etapa finaliza quando todas as fontes de informação forem encontradas e os endereços dos *blogs* estejam armazenados nas base de dados (MySQL e *Google Big Query*).

O segundo passo foi verificar todas as postagens que existem no *blog*. Em boa parte dos serviços de *blogs*, uma nova postagem sempre é criada como uma nova página (nova URL), então era necessário encontrar os endereços destas páginas. Para isso, basta procurar pelo arquivo *sitemap.xml* que contem, como o nome sugere, o mapa do site e todos as suas “subpáginas” a partir do domínio original. Por exemplo, no *blog* “jonwizardnews.wordpress.com” era necessário procurar pelo endereço jonwizardnews.wordpress.com/sitemap.xml. A partir dai, o microgisborne deveria recuperar todo o conteúdo entre as marcações <loc></loc> que contem os endereços da subpágina. Esta segunda etapa finaliza quando todas as subpáginas forem localizadas e cadastradas nas bases de dados vinculadas as páginas principais.

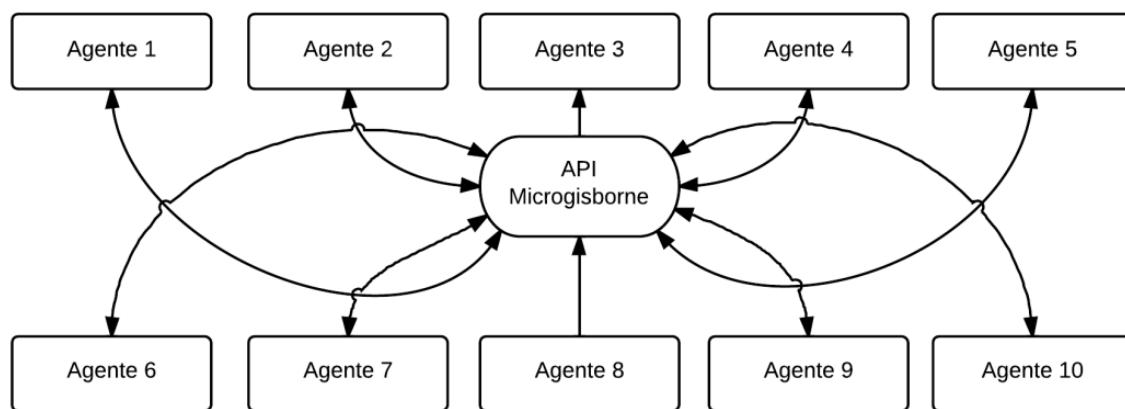
O terceiro passo seria “visitar” cada página e subpágina e ao olhar o código fonte procurar pelas marcações “*rel-tag*” do padrão de microformatos. Estas marcações representam as etiquetas realizadas pelos usuários em uma postagem específica. Então cada subpágina foi varrida pelo microgisbourne e cada *tag* foi vinculada a uma postagem “subpágina”, e cada postagem vinculada a um *blog* (página).

Para este experimento a ferramenta foi configurada a Infraestrutura como serviço (IaaS) da amazon.com e foram criados 10 agentes onde cada uma simulava algum tipo de serviço do microgisborne. Os 10 agentes se alternavam, de forma aleatória, entre os três passos da atividade, para simular momentos em que fossem necessárias indexar mais informações ou simular que 10 usuários simultâneos estão buscando dados na ferramenta e assim o serviço não deveria estar realizando tarefas *back-end*. A FIGURA 4 a seguir apresenta este funcionamento.

Esta estrutura pode ser dimensionada para 1 único agente ou para 10.000 agentes simultâneos. A carga de trabalho dos agentes foram monitorados e estiveram em média, apenas 13% do tempo em estado de espera (*Idle*). Nas primeiras 6 horas, o número de instâncias, ou agentes funcionando, foram variados entre 3 e 30 para simular a escalabilidade do serviço.

Todo os dados processados foram colocados em dois sistemas gerenciadores de banco de dados (SGBD). O SGBD MySQL utilizou a modelagem relacional para a organização lógica dos dados, enquanto que no Google Big Query utilizou o formato JSON para modelar os dados. Em três dias de experimento, foram visitados 2.546 blogs que continham 764.429 páginas e foram coletadas 6.699.643 *tags*. Enquanto no MySQL estes dados ocuparam 418,72 Mb de dados, enquanto no Google Big Query se utilizou cerca de 356,99 Mb. Entretanto, as consultas utilizando o Google Big Query são em média 56% mais rápidas do que no MySQL e a tendência é que com a quantidade de dados aumentada (Big) é que este SGBD seja melhor.

FIGURA 4 – Funcionamento do Microgisborne



Fonte: Os Autores

5 CONSIDERAÇÕES FINAIS

O experimento demonstrou que a ferramenta foi bem sucedida em coletar etiquetas de *blogs* utilizando o vocabulário “*rel-tag*” de microformatos se mostrando uma solução simples para a recuperação de informações em serviços de Internet. A partir dessa possibilidade, novos produtos e serviços de informação, baseados em microformatos podem ser elaborados. Por exemplo, nuvens de *tags* das categorias *News* e *Music*, FIGURAS 5 e 6 respectivamente, foram criadas a partir dos dados coletados pela ferramenta e geradas pela ferramenta D34js.

FIGURA 5 – Nuvens de Tags dos Categorias *News*



Fonte: Os Autores

FIGURA 6 – Nuvens de Tags dos Categoria Music



Fonte: Os Autores

Atualmente a ferramenta já implementa os seguintes serviços:

Tags de um Blog: Ao digitar a URL de qualquer *blog* que tenha suporte aos microformatos ele retorna a lista de *tags*, exemplo na FIGURA 7. A nuvem de *tags* por *blogs* ainda está em construção.

FIGURA 7 – Tags do blog vernasunhas.blogspot.com



Fonte: Os Autores

Lista de Blogs a Partir de uma Tag: Este serviço apresenta uma lista de *blogs* que contenham uma *tag* escolhida pelo usuário. Este ainda não está implementado no serviço *online*, mas a FIG. 8, mostra o resultado deste serviço no experimento.

FIGURA 8 – Lista de Sites a partir de uma Tag

Fonte: Os Autores

Tag Blogrolling: Este serviço apresenta uma lista de *blogs* que apresentam o mesmo conjunto de *tags* que um outro *blog* escolhidos pelo usuário. Por exemplo, um usuário pode entrar com o endereço <http://vernasunhas.blogspot.com.br> e o microgisborne sugere outros *blogs* que possuam o mesmo conjunto, ou o máximo, conjunto de *tags*. Este serviço ainda não está implementado devido.

Finalmente, sugerimos a reflexão de que os maiores desafios levantados quanto a evolução do microgisborne não provêm da tecnologia em si, mas sim da elaboração de fluxos da informação para as *tags* armazenadas nas bases de dados e transformar estes fluxos em produtos e serviços úteis a sociedade.

REFERÊNCIAS

- ALEXANDER, B. Web 2.0: A new wave of innovation for teaching and learning? *Educause review*, v. 41, n. 2, p. 32-44, 2006.
- ALLSOP, J. **Microformats:** Empowering Your Markup for Web 2.0. Nova York: Friends of, 2007. 368 p.
- AMARAL, A.; AQUINO, M. Práticas de folksonomia e social tagging no Last. fm. In: Simpósio Brasileiro de Fatores Humanos e Sistemas Computacionais, 8., 2008, Porto Alegre. *Anais...*, Porto Alegre, Simpósio Brasileiro de Fatores Humanos e Sistemas Computacionais, 2008.
- BAEZA-YATES, R.; RIBEIRO-NETO, B. **Recuperação de Informação:** Conceitos e Tecnologia das Máquinas de Busca. 2. ed. Porto Alegre: Bookman, 2013. 614 p.
- CASSANO, R. **Somos tão Jovens:** Uma breve retrospective da Web. 2011. Disponível em: <<http://www.techtudo.com.br/platb/internet/2011/06/09/somos-tao-jovens-uma-breve-retrospectiva-da-web/>> Acesso em: 13 Ago. 2013.
- FU, B. Trust Management in Online Social Networks. 2007. 90 f. **Dissertação** (Mestrado em Ciência da Computação) - Departamento de Ciências da Computação. Trinity College, University of Dublin, Dublin. 2007.
- GUIMARÃES, C. P. Tags: palavras-chave em blogs. In: Simpósio Hipertexto e Tecnologias na Educação. 2. 2008, Recife. *Anais...* Recife. Simpósio Hipertexto e Tecnologias na Educação, 2008. p. 1-22.
- HEWITT, H. **Blog:** Understanding the information reformation that's changing your world. Nashville: Thomas Nelson Inc., 2005. 256 p.
- KHARE, R; ELIK, T. Microformats: A Pragmatic Path to the Semantic Web. 2012. Disponível em: <<http://commerce.net/wpcontent/uploads/2012/04/CN-TR-06-01.pdf>>. Acesso em: 18 fev. 2014.
- KHARE, R. Microformats: The next (small) thing on the Semantic Web. *Internet Computing*, v. 10, n. 1, p. 68-75, 2006.
- KIM, H.; SCERRI, S.; PASSANT, A.; BRESLIN, J. Integrating Tagging into the Web of Data: Overview and Combination of Existing Tag Ontologies. *Journal Of Internet Technology*, Taiwan, v. 12, n. 4, p. 561-572, 2011.
- KOIVUNEN, M.; MILLER, E. W3c semantic web activity. *Semantic Web Kick-Off in Finland*, p. 27-44, 2001.
- LIPTON, R., **What is a Weblog?**. 2002. Disponível em: <<http://radio.weblogs.com/0107019/stories/2002/02/12/whatIsAWeblog.html>>. Acesso em: 13 Ago. 2013.

LUND, B.; HAMMOND, T; FLACK, M; HANNAY, T. Social bookmarking tools (II). **D-Lib magazine.** v. 11, n. 4, 2005.

MARLOW, C; NAAMAN, D; BOYD, D; DAVIS, M. Tagging paper, taxonomy. In: Conference on Hypertext and Hypermedia. 17. Nova York, 2006. **Anais...** Nova York, ACM, 2006.

MENDEZ, E; LÓPEZ, L. M.; SICHES, A.; BRAVO, A. G. **DCMF: DC & Microformats, a Good Marriage.** In: International Conference on Dublin Core and Meta-Data Applications. Berlin, 2008. **Anais...** Berlin, DC-2008: International Conference on Dublin Core and Metadata Applications, 2008.

MICROFOMARTS. **What is Microformats.** 2014. Disponível em: <www.microformats.org>, Acesso em: 18 fev. 2014.

MURUGESAN, S. Understanding Web 2.0. **It Professional**, USA, v. 9, n. 4, p. 34-41, 2007.

O'REILLY, T. What Is Web 2.0: Design Patterns and Business Models for the Next Generation of Software. **Communications & Strategies**, v. 1, n. 65, p. 17, 2007.

OREN, E.; MÖLLER, K.; SCERRI, S.; HANDSCHUH, S. **What are semantic annotations.** Galway: Digital Enterprise Research Institute, National University of Ireland, 2006. 14 p.

PETERS, I.; STOCK, W. G. Folksonomy and information retrieval. **Journal of American Society for Information Science and Technology**, v. 44, n. 1, p. 1-28, 2007.

PRESSLEY, L. **Folksonomies, tags, and user-created metadata:** A truly emerging technological trend unpublished seminar on emerging technological trends in libraries. Winston-Salem : Reynolds Library - Wake Forest University, 2005. 22 p.

RODZVILLA, J. **We've Got Blog:** How Weblogs are Changing Our Culture. Cambridge: Basic Books, 2002. 176 p.

RUSSELL, T. **Contextual authority tagging:** cognitive authority through folksonomy. 2005. Disponível em: <<http://www.terrellrussell.com/projects/contextualauthoritytagging/conanhtag200505.pdf>>. Acesso em: 1 out. 2013.

SINCLAIR, J.; CARDEW-HALL, M. The folksonomy tag cloud: when is it useful?. **Journal of Information Science**, v. 34, n. 1, p. 15-29, 2008.

SOUZA, R.; ALVARENGA, L. A Web Semântica e suas contribuições para a ciência da informação. **Ciência da Informação**, Brasília, v. 33, n. 1, p 132-141, 2004.

TREMAYNE, Mark. Introduction: Examining the blog-media relationship. **Blogging, citizenship, and the future of media**, New York, p. 3-20, 2007.

VANDER WAL. Explaining and Showing Broad and Narrow Folksonomies. 2005. Disponível em: <<http://www.vanderwal.net/random/entrysel.php?blog=1835>>. Acesso em: 1 Ago. 2013.

WITTEN, H.; FRANK, E.; MARK, A. Hall. **Data Mining:** Practical machine learning tools and techniques. Morgan Kaufmann 2011. 560 p. (The Morgan Kaufmann Series in Data Management Systems – Series)

WHARTON. What's the Next Big Thing on the Web? It May Be a Small, Simple Thing - Microformats. 2005. Disponível em:
<<http://knowledge.wharton.upenn.edu/index.cfm?fa=printArticle&ID=1247>> Acesso em: 1 ago. 2013

WORDPRESS. Stats -Wordpress.com. 2014. Disponível em:
<<http://en.wordpress.com/stats/>>. Acesso em: 18 fev. 2014.

POLÍTICAS DE FORMATOS DE ARQUIVOS PARA OBJETOS DE APRENDIZAGEM: PRESERVAÇÃO DIGITAL NO SABER TECNOLOGIAS EDUCACIONAIS E SOCIAIS

FILE FORMAT POLICY FOR LEARNING OBJECTS: DIGITAL PRESERVATION ON THE SABER EDUCATIONAL AND SOCIAL TECHNOLOGIES

Vildeane da Rocha Borba
 Sandra de Albuquerque Siebra
 Marcos Galindo
 Josiane Lemos Machiavelli
 Cristine Martins Gomes de Gusmão

Resumo: Este artigo tem o objetivo de apresentar as políticas de formatos de arquivos para objetos de aprendizagem implementadas no SABER Tecnologias Educacionais e Sociais, com o propósito de viabilizar o acesso a longo prazo aos recursos educacionais produzidos nos cursos na modalidade a distância. A fundamentação teórica se pauta em conceitos sobre estratégias de preservação digital, políticas de preservação, educação a distância e formatos de arquivos, adotando instrumentos necessários para o controle e a preservação da memória em meio digital e sua permanência continuada. A metodologia utilizada foi a pesquisa qualitativa e o procedimento metodológico utilizado foi o método de estudo de caso. O objeto de estudo foram os objetos de aprendizagem produzidos pelo grupo SABER, devido ao grande quantitativo produzido para atender às demandas de cursos de especialização e capacitação, especialmente, na área de saúde. A política de formatos de arquivos para objetos de aprendizagem abrangeu arquivos em formato textual, imagem, áudio e vídeo, apresentando a padronização necessária levando-se em consideração as vertentes acesso e preservação. Como conclusão, verifica-se que, na prática, é encorajado que os objetos de aprendizagem sejam criados e possam ser reutilizados por outros cursos e iniciativas, quando necessário. Dessa forma, é preciso garantir a sua preservação, para proporcionar o contínuo acesso e uso. Assim, a preservação desses objetos deve ser encarada como algo primordial, considerando os impactos advindos das mudanças tecnológicas.

Palavras-chave: Preservação Digital. Objetos de Aprendizagem. Formato de Arquivo.

Abstract: This article aims to present the policies of file formats for learning objects implemented in SABER Educational and Social Technologies for the purpose of facilitating the long-term access to educational resources produced in courses in distance mode. The theoretical foundation is guided by concepts of digital preservation strategies, preservation policies, distance education and file formats, adopting the necessary instruments for the control and preservation of memory in digital media, to ensure its continued access. The methodology was qualitative research and the methodological procedure used was the case study. The focus of the study were the learning objects produced by the SABER group, due to the large quantity produced to meet the demands of specialized courses and training, especially in the health area. The file formats policy for learning objects included text, images, audio and video files, specifying the standardization required, taking into account the access and preservation aspects. In conclusion, it appears that, in practice, it is encouraged that learning objects are created and can be reused by other programs and initiatives, as needed. Thus, it is necessary to ensure their preservation, to provide continuous access and use. Therefore, the preservation of these objects should be seen as something essential, considering the impacts from technological change.

Keywords: Digital Preservation. Learning Objects. File Format.

1 INTRODUÇÃO

A educação a distância (EAD), segundo Moran (2005), refere-se a um processo de ensino-aprendizagem no qual docentes/tutores e discentes/estudantes estão separados fisicamente, espacial e/ou temporalmente, podendo estar conectados mediante tecnologias, sobretudo as telemáticas, a exemplo da internet.

Apesar de a internet ser o seu meio mais utilizado nos dias atuais, a EAD também pode ocorrer utilizando correio, rádio, televisão, vídeo, CD-ROM, telefone, entre outras tecnologias. Em um país de dimensões continentais como o Brasil, essa modalidade propicia a expansão da missão educadora das Instituições de Ensino Superior (IES), eliminando as barreiras de tempo e espaço, contribuindo para a redução das desigualdades sociais, no sentido de ampliar o raio de atuação das IES, comumente concentradas nas capitais dos estados e proximidades.

A EAD vem se consolidando cada vez mais no cenário nacional, e grandes projetos nessa modalidade vêm sendo financiados pelas esferas federais (especialmente pelos Ministério da Educação e Ministério da Saúde), estaduais e municipais, tais como a Universidade Aberta do Brasil (UAB), o E-TEC Brasil e a Universidade Aberta do Sistema Único de Saúde (UNA-SUS), entre outros. Os financiamentos envolvem a criação de cursos técnicos, de graduação, de especialização, de mestrados e de doutorados, além de cursos de capacitação com curta ou média duração.

Nesse sentido, cada vez mais estão sendo produzidos e desenvolvidos objetos de aprendizagem (doravante OAs) adequados às estratégicas pedagógicas e também às tecnologias utilizadas, como meio de propagação e disseminação do conhecimento. Os OAs podem ser compreendidos como “qualquer recurso digital que possa ser reutilizado para o suporte ao ensino” (WILEY, 2000, p. 3, tradução nossa). Assim, estes recursos educacionais contribuem para o processo de ensino-aprendizagem e sofrem impacto direto no que diz respeito à sua durabilidade, disseminação e preservação, devido à sua natureza digital. Por conseguinte, eles merecem uma atenção especial, principalmente devido à expansão de sua produção, visto que são produzidos tanto nos contextos dos cursos a distância, como também para atender necessidades de cursos presenciais.

Inicialmente, os produtores dos OAs tinham como foco apenas a produção, disponibilização e acesso dos conteúdos digitais, sem a preocupação com a garantia da proteção informacional de valor permanente. Atualmente, desperta-se para a necessidade de determinação de políticas e estratégias de preservação que devem estar lastreadas em um documento dirigido a orientar ações emergenciais e estruturais.

Nesse cenário, o objetivo deste artigo é apresentar as políticas de formatos de arquivos para OAs implementadas no SABER Tecnologias Educacionais e Sociais, com o propósito de viabilizar o acesso a longo prazo aos recursos educacionais produzidos em seus cursos na modalidade a distância.

Espera-se, com esse artigo, contribuir para a discussão sobre a preservação de OAs e colaborar com a conscientização das diversas iniciativas de Educação a Distância espalhadas pelo Brasil da necessidade de se pensar no acesso a longo prazo da memória digital com fins educacionais produzida, muitas vezes, sem critérios ou padrões ou preocupação com o futuro.

2 PRESERVAÇÃO DIGITAL

Uma das características mais importantes da natureza da informação é a perecibilidade dos novos suportes do registro do conhecimento e a rápida maturação ou obsolescência dos instrumentos tecnológicos. Esta condição exige a construção de estratégias especializadas que permitam não somente a garantia da manutenção do acesso aos recursos de informação digital de longo termo, mas também a garantia do contínuo acesso aos conteúdos e funcionalidades dos recursos eletrônicos originais.

A preservação da memória digital diz respeito à aplicação de estratégias, capazes de viabilizar a permanência continuada de informações para uso das futuras gerações, preocupando-se com os impactos das mudanças tecnológicas. Miguel Ferreira destaca a capacidade que a preservação eletrônica possui de: “Garantir que a informação digital permaneça acessível e com qualidades de autenticidade suficientes para que possa ser interpretada no futuro, recorrendo-se a uma plataforma tecnológica diferente da utilizada no momento de sua criação” (FERREIRA, 2006, p. 20).

Hedstrom define a preservação digital como o:

Planejamento, alocação de recursos e aplicação de métodos de preservação e tecnologias necessárias para que a informação digital de valor contínuo permaneça acessível e utilizável por longo prazo, considerando-se neste caso longo prazo o tempo suficiente para preocupar-se com os impactos de mudanças tecnológicas (HEDSTROM, 1997, p. 190, tradução nossa).

A obsolescência tecnológica dos suportes físicos, de hardware e de software, dos formatos dos dados, são pontos nevrálgicos, portanto, para o desenvolvimento de estratégias e de estudos onde se concentra o esforço atual dos especialistas para garantir a perenidade dos documentos digitais.

Arellano apresenta três níveis de preservação: física, lógica e intelectual para os objetos digitais. Diz o autor:

A Preservação física está centrada nos conteúdos armazenados em mídias magnéticas (fitas cassete de áudio e de rolo, fitas VHS e DAT etc.) e discos ópticos (CD-ROMs, WORM, discos ópticos regraváveis). A Preservação lógica procura na tecnologia formatos atualizados para inserção dos dados (correo eletrônico, material de áudio e audiovisual, material em rede etc.), novos software e hardware que mantenham vigentes seus bits, para conservar sua capacidade de leitura. No caso da Preservação intelectual, o foco são os mecanismos que garantem a integridade e autenticidade da informação dos documentos eletrônicos (ARELLANO, 2004, p.17).

A preservação digital é uma etapa basilar, e necessariamente posterior ao desenvolvimento das Tecnologias da Informação e Comunicação e do sistema de redes que criou a nova ambiência para a informação em meio digital. Corresponde, de certo modo, ao fenômeno do desenvolvimento das estratégias e métodos para conservação e preservação de papel e materiais bibliográficos, ocorrido em momento pós-expansão documental, iniciado com o ciclo das grandes guerras mundiais.

Porém, esse problema da preservação do digital tomou dimensão inesperada, notadamente quando a humanidade se deu conta do risco que pairava sobre os acervos digitais ameaçados pela rápida obsolescência de formatos, suportes, hardware e software. Daí se iniciaram os estudos sobre a preservação da informação em meio digital e se motivou a criação de políticas nesse sentido.

3 POLÍTICAS DE PRESERVAÇÃO DIGITAL

A implementação de políticas de preservação digital se faz através do desenvolvimento e da implantação de diretrizes, programas, planos e projetos de preservação. Isso com o objetivo explícito de oferecer subsídios orientadores para um programa de política pública estrutural, dirigido ao atendimento unívoco das demandas de estratégias de gestão, para a problemática da obsolescência do conhecimento suportado em formato digital.

Dessa forma, em primeiro lugar, é necessário selecionar e criar coleções digitais com um valor duradouro. Depois, deve haver uma política de preservação bem definida, que estabeleça as regras e procedimentos, assim como o que deve ser preservado. Essa política deve ser revisada periodicamente, tanto para melhorar os métodos, quanto para redefinir o conjunto de objetos a ser preservado. Aos objetos preservados deve-se atribuir um limite de vida. Alguns serão mais perecíveis que outros, e estas durações devem ser revisadas periodicamente (BIA PLATAS; SÁNCHEZ QUERO, 2002, tradução nossa).

No Brasil, é notável a quase inexistência de políticas de preservação: “Na atualidade, a política para garantir o armazenamento de longo prazo, manutenção, migração

digital e acesso a materiais, quer a nível local ou nacional, não está frequentemente presente tanto no setor privado quanto no setor público” (ERPANET, 2003, p. 3, tradução nossa).

A nível local foi instituída, pelo Conselho Nacional de Arquivos (CONARQ) por meio da Portaria nº. 07, de 21 de agosto de 1995, a Câmara Técnica de Conservação de Documentos, que foi alterada para o nome Câmara Técnica de Preservação de Documentos pela Portaria nº. 81, de 16 de junho de 2008. Essa Câmara tem o objetivo de atender às demandas de normalização e criação de instrumentos metodológicos, técnicos e normativos, visando à preservação dos acervos documentais dos arquivos públicos e privados.

A Carta para Preservação do Patrimônio Arquivístico Digital: Preservar para garantir o acesso (2005), desenvolvida pelo CONARQ, visa, entre suas propostas, à elaboração de estratégias e políticas para salvaguardar documentos em meio digital, considerando, principalmente, a finalidade do CONARQ em definir “[...] a política nacional de arquivos públicos e privados e exercer orientação normativa visando à gestão documental e à proteção especial aos documentos de arquivo, independente do suporte em que a informação está registrada [...].” (CONARQ, 2005, p. 1).

Com essa carta, o CONARQ reafirma o seu compromisso com a aplicação de políticas públicas voltadas para a preservação do patrimônio arquivístico digital, além de convocar os setores públicos e privados, envolvidos com a produção e proteção especial dos documentos em formato digital, a evidarem esforços para garantir sua preservação e acesso contínuo, o que é fundamental para a democratização da informação arquivística em nosso país e a preservação da memória nacional (CONARQ, 2005).

Porém, a falta de consenso quanto ao uso de práticas e métodos mais adequados e a ausência de modelos firmados de preservação digital podem ser algumas das causas ou dificuldades de se desenvolverem políticas de preservação digital nas instituições, que visem à orientação de boas práticas de preservação de materiais digitais e à garantia da autenticidade, da confiabilidade e da acessibilidade de longo prazo. Além disso, a implantação de políticas de preservação digital poderia fornecer às empresas privadas ou do setor público governamental uma visão partilhada e muito mais clara do que é preciso desenvolver na indústria de TI, na prestação de produtos, serviços e ferramentas mais adequados a uma preservação mais eficiente.

Segundo Ferreira (2006), uma política de preservação digital deve definir, para cada tipo de objeto digital, um conjunto de propriedades e ações que preservem suas características, seu ciclo de vida e sua autenticidade, garantindo a perenidade do seu conteúdo. Uma política deve transmitir a própria filosofia de uma organização, que deve

induzir a um entendimento comum dos objetivos, do fato de que cada item da coleção deve ser preservado com o máximo esforço possível (ERPANET, 2003). Uma política de preservação deve ser abrangente, ter um alcance amplo e incluir não só questões técnicas, mas também organizacionais, jurídicas, de recursos humanos, de gestão de direitos, de acesso e de propriedade intelectual.

Alguns estudos sobre a preservação digital têm estabelecido que a imediata implementação de políticas de preservação digital é a forma mais efetiva de garantir o armazenamento e uso dos recursos de informação por longos períodos de tempo. A falta dessas políticas nos projetos de repositórios digitais sugere a carência de conhecimentos técnicos sobre a importância das estratégias de preservação digital existentes (ARELLANO, 2004).

Uma das consequências da crescente utilização da informação digital é que as organizações estão se tornando mais e mais conscientes da necessidade de desenvolver políticas de preservação. A política de preservação digital pode e deve produzir benefícios para as instituições, tais como garantir materiais digitais disponíveis para a atual e futura utilização. Além disso, a formulação de uma política permite tratar de temas difíceis, como a curta duração de vida digital, a pequena capacidade dos materiais digitais, a obsolescência do hardware e software para leitura dos dados e formatos de arquivo e, finalmente, a heterogeneidade estrutural e técnica dos diferentes tipos de materiais digitais.

4 OBJETOS DE APRENDIZAGEM

Os OAs começaram a ser estudados muito recentemente, assim, ainda não há um consenso entre os autores sobre a sua definição. Para este trabalho, será adotada a definição dada por Wiley (2000, p. 3, tradução nossa), que define o OA como “qualquer recurso digital que possa ser reutilizado para o suporte ao ensino”. Assim, os OAs podem ser criados em qualquer mídia ou formato, podendo ser simples, como uma animação, um arquivo de texto, uma imagem ou foto ou uma apresentação de slides, ou complexos, como um software de simulação de situações-problema ou um website, que fazem uso de diversos tipos de recursos.

Um OA deve ter um propósito educacional bem definido, sendo um elemento que estimule a reflexão do estudante, e deve poder ser reutilizável, ou seja, ele não deve ficar restrito a um único contexto (BETTIO; MARTINS, 2004), haja vista que deve poder ser usado a longo prazo, e sua principal vantagem é poder ser reutilizado.

Diversos fatores favorecem o uso de OAs na área educacional, seja como material didático complementar na educação presencial ou como material básico ou complementar

nos cursos a distância. Entre eles, é possível citar (WILEY, 2000; LONGMIRE, 2001; SÁ FILHO; MACHADO, 2004):

- **Flexibilidade** – os OAs são construídos de forma que possam ser reutilizáveis em vários contextos, sem modificações, trazendo uma economia de custos e de tempo de desenvolvimento/adaptação.
- **Interoperabilidade** – os OAs podem ser utilizados em qualquer plataforma de ensino;
- **Customização** – como os OAs são criados para serem independentes, eles podem ser utilizados em vários cursos, e cada instituição educacional poderá utilizar-se dos objetos e arranjá-los da maneira que lhes for mais conveniente. Inclusive, os objetos poderiam ser arranjados para formarem um novo objeto mais complexo a ser aplicado em um novo propósito, em um contexto diferente.

Em geral, a construção do OA é realizada por meio da colaboração de três subequipes: a pedagógica, a tecnológica e a de design; buscando um objetivo em comum, cada uma contribuindo com sua especialidade. E, como a ideia principal na construção do OA é estender a sua utilização para vários projetos e vários contextos, pensar na preservação a longo prazo desses objetos é algo de grande relevância.

Com o advento da Internet e das Tecnologias de Informação e Comunicação (TIC`S) tornou-se possível o desenvolvimento de repositórios digitais, atrelados ao movimento em favor do acesso livre à informação científica. Quebrando, assim, o paradigma tradicional no qual editoras científicas comerciais, detentoras dos direitos autorais patrimoniais, atribuíam altos preços, além de impor barreiras de permissão sobre publicações de resultados de pesquisas. Estas financiadas, muitas vezes, com recursos públicos, limitando a visibilidade e a circulação do conhecimento científico.

É amplamente verificado no contexto da educação a distância também o desenvolvimento de grandes repositórios digitais. Como exemplo, é possível citar o Banco Internacional de Objetos Educacionais do Ministério da Educação⁸⁸, o Acervo de Recursos Educacionais em Saúde (ARES) da Universidade Aberta do SUS⁸⁹, a Rede Interativa Virtual de Educação (RIVED) da Secretaria de Educação a Distância – SEED⁹⁰ entre vários outros. Todos com a finalidade de ampliar as fronteiras de acesso e disseminação de

88 <http://objetoseducacionais2.mec.gov.br>

89 <https://ares.unasus.gov.br/acervo/>

90 <http://rived.mec.gov.br>

informação de OAs e/ou recursos educacionais, levando em consideração uma de suas principais filosofias: a reutilização.

Com relação a repositórios digitais de OAs, Souza e Viera (2012, on line) afirmam que “o seu emprego no segmento educacional prospecta ganhos quali-quantitativos no processo de aprendizagem e vislumbra ao pesquisador, volume de informações crescente e relevante na produção de novos conhecimentos”. Porém, como os repositórios digitais de OAs se enquadram na filosofia do acesso aberto/livre, grande parte dos OAs disponibilizados atendem aos requisitos da vertente acesso, para facilitar o download e a não sobrecarga em relação ao quantitativo de armazenamento no repositório.

Esta realidade favorece a questão da disseminação e ampliação na difusão de OAs. Porém, até que ponto para fins de reutilização de um vídeo ou imagem, por exemplo, que necessite estar em alta qualidade digital ou precise ser convertido para um outro formato, aquele objeto digital disponível no repositório (em baixa qualidade, para facilitar o acesso) irá atender as necessidades específicas para determinado contexto?

É neste sentido, que este estudo se pauta. Na preocupação de como tratar o problema da ausência de conhecimento sobre os instrumentos de controle e de preservação da memória, emprestada ao corpo sutil do meio digital e, em especial, aos OAs. Este trabalho não exclui a importância dos formatos para disseminação e difusão de OAs em meio digital, mas sim pretende sensibilizar quanto a importância da preservação da memória armazenada em meio digital, a fim de proporcionar o acesso a longo prazo destes conteúdos.

5 SABER TECNOLOGIAS EDUCACIONAIS E SOCIAIS

O SABER Tecnologias Educacionais e Sociais é um grupo de pesquisa colegiado formado por docentes, pesquisadores, estudantes e pessoal de apoio técnico de natureza multidisciplinar da Universidade Federal de Pernambuco (UFPE). O SABER atua no desenvolvimento de processos educacionais, ferramentas tecnológicas e apoio acadêmico/administrativo na execução de cursos na modalidade a distância.

No desenvolvimento de processos educacionais, o SABER se pauta no planejamento pedagógico e instrucional para construção de OAs para cursos de educação a distância e semipresenciais, suportados por ferramentas tecnológicas. Na produção de soluções tecnológicas, desenvolve aplicações para a web, mobile e desktop, dando apoio na gestão acadêmica, desenvolvimento de OA e de ambientes virtuais de aprendizagem. No suporte acadêmico/administrativo, atua na estruturação e treinamento de equipes para dar apoio aos estudantes, tutores e professores que fazem parte dos cursos promovidos pelo SABER.

Devido ao seu contexto de atuação, no SABER são desenvolvidos vários OAs. Em especial, objetos no formato de texto, áudio, vídeo e imagem são desenvolvidos como parte do processo de construção de recursos educacionais, visando atender às necessidades de ensino e aprendizagem, complementados com a aplicação correta de recursos interativos, usabilidade adequada e foco no público, colaborando para uma formação mais efetiva.

Diante do exposto, no grupo SABER, sentiu-se a necessidade de determinar formatos de arquivos para os OAs, sendo levadas em consideração as vertentes acesso e preservação. Essa determinação dos formatos é considerada um primeiro passo em direção ao planejamento de políticas de preservação digital a serem adotadas pelo grupo, com o propósito de possibilitar a perpetuação dos diversos OAs criados para as gerações.

6 METODOLOGIA

Essa pesquisa quanto aos objetivos tem um caráter qualitativo, pois existe “[...] uma relação dinâmica, particular, contextual e temporal entre o pesquisador e o objeto de estudo”, enfatizando a capacidade de conhecê-lo, compreendê-lo e adaptá-lo à realidade pesquisada. (MICHEL, 2009, p. 36).

Quanto aos meios, pauta-se em um estudo exploratório ou pesquisa bibliográfica, buscando o levantamento teórico sobre o tema, familiarizando-se melhor com o problema para o aprimoramento de ideias e descoberta de intuições. (MICHEL, 2009).

O método utilizado foi o estudo de caso, pois “trata-se de uma técnica [...] que se caracteriza por ser o estudo de uma unidade [...] com o objetivo de compreendê-la em seus próprios termos, ou seja, no seu próprio contexto.” (MICHEL, 2009, p. 53).

Nesse sentido, o foco do estudo recaiu sobre os OAs desenvolvidos pelo grupo SABER Tecnologias Educacionais e Sociais, com o propósito de entender os fatos e fenômenos que se configuravam no desenvolvimento de formatos de arquivos dos objetos desenvolvidos. A proposta do estudo foi determinar uma política para padronização e regulamentação dos formatos a serem utilizados, primando por formatos específicos para preservação a longo prazo.

7 FORMATOS DE ARQUIVOS

O formato de arquivo a ser utilizado para preservar ou prover o acesso a objetos digitais costuma ser bastante debatido, pois questões sobre qual formato utilizar, principalmente para preservar o arquivo digital para futuras gerações, são as mais preocupantes.

Formatos de arquivo podem ser definidos como informações codificadas que só podem ser processadas e compreensíveis por combinações específicas de hardware e software, e sua acessibilidade é vulnerável na atualidade, pois evoluem rapidamente em ambiente tecnológico. Isso traz preocupação para todos os responsáveis pela gestão e acesso aos registros eletrônicos em escalas de tempo relativamente curta (BROWN, 2008).

Rimkus *et al.* (2014) afirma que os formatos de arquivo abertos são mais indicados do que os fechados ou chamados formatos proprietários, porque a maneira como eles codificam o conteúdo é transparente. A adoção de um formato de arquivo proprietário por uma ampla comunidade de criadores de conteúdo, divulgadores e usuários, é muitas vezes considerada um indicador confiável de longevidade do formato. Nesse sentido, os autores apresentam que não existe, de fato, nenhuma fórmula infalível para decisões de política de formato de arquivo.

Hodge e Frangakis (2004) enfatizam que o melhor formato para preservação de longo prazo continua elusivo, talvez porque não há nenhuma resposta única para a pergunta, afirmindo, porém, que a maioria dos especialistas concorda que o melhor formato para preservação é o que é menos proprietário, pois carrega aspectos significativos do original.

O que é consenso é que, como reafirmam Rog e Van Wijk (2008, p.1, tradução nossa), “a escolha de um formato de arquivo específico no momento da criação, ou mais tarde no ciclo de vida de um objeto digital influencia o acesso a longo prazo ao objeto digital.”

Um trabalho nessa área que merece destaque foi o realizado pelo Arquivo do Estado de Illinois, localizado no meio oeste dos Estados Unidos, que desenvolveu um guia para as agências governamentais sobre a adoção de formatos de arquivos sustentáveis para registros eletrônicos. A sustentabilidade refere-se à acessibilidade contínua ao longo do tempo, e, nesse sentido, esses formatos devem abranger aos requisitos de formatos abertos ou não proprietários, adoção e utilização generalizada, formatos autodescritivos e arquivos não criptografados (WHITE, 2013).

8 FORMATOS DE ARQUIVOS PARA OBJETOS DE APRENDIZAGEM

Com base nas diretrizes e políticas do Florida Digital Archive (FLVC, 2013), do The U.S. National Archives and Records Administration (NARA, 2010), do Illinois State Archives (WHITE, 2013) e da Library of Congress (LIBRARY OF CONGRESS, 2014), foram definidos os formatos de arquivos para os OAs implementados no SABER Tecnologias Educacionais e Sociais, atendendo as vertentes de preservação e acesso de objetos digitais. Essa definição objetiva estabelecer boas práticas de preservação digital,

com relevante impacto futuro para o conhecimento produzido nos cursos desenvolvidos pelo SABER.

8.1 Estratégia de normalização

A Estratégia de Normalização visa simplificar o processo de preservação através da redução do número de formatos distintos que se encontram no repositório de objetos digitais (FERREIRA, 2006). Os Formatos desenvolvidos para os OAs se pautam em texto, imagem, áudio e vídeo. Nesse sentido, foi determinada a utilização de extensões específicas para acesso e preservação de cada um desses formatos, conforme será apresentado nas subseções a seguir.

8.1.1 Texto

Como extensões de preservação para texto, foram definidos o PDF/A (Portable Document Format/Archives) nos formatos PDF/A-1, PDF/A-2 e o PDF/A-3, no caso de arquivos de texto e o XML (Extensible Markup Language) no caso de texto para Web (Quadro 1). (FLVC, 2013; LIBRARY OF CONGRESS, 2014; NARA, 2010; WHITE, 2013).

QUADRO 1- Padronização SABER - Formato de Arquivos Textuais - Preservação

POLÍTICA DE FORMATO DE ARQUIVO (PRESERVAÇÃO)			
TIPO	FORMATO	BASE	RECOMENDAÇÃO
Textual	Portable Document Format Archives (PDF/A)	ISO 19005-1:2005	Recomenda-se que todos os OAs em formato textual devem ser preservados no formato PDF/A.
	PDF/A-1	ISO 19005-2:2011	
	PDF/A-2	ISO 19005-3:2012	
	PDF/A-3		
Textual	XML (Extensible Markup Language)	World Wide Web Consortium (W3C)	Recomenda-se que todos os cursos desenvolvidos na web, utilizem a linguagem XML.

Fonte: Os Autores, 2014.

Os formatos PDF/A-1, PDF/A-2 e o PDF/A-3 são advindos da versão 1.4, 1.7 e 1.7 do Adobe PDF respectivamente e foram adequados para a preservação a longo prazo de documentos. Esses formatos se tornaram uma norma ISO (19005-1:2005), (19005-2:2011), (19005-3:2012), respectivamente, desenvolvidos por um grupo de trabalho com representantes do governo, da indústria e do meio acadêmico com o apoio ativo da Adobe Systems Incorporated. Assim, essas são as extensões de arquivos mais difundidas para preservação a longo prazo de textos. (FLVC, 2013; LIBRARY OF CONGRESS, 2014; NARA, 2010; WHITE, 2013)

O XML (Extensible Markup Language) também é bastante utilizado, pois permite a preservação dos metadados associados e é adequado para informações em websites. É mantido e desenvolvido pela World Wide Web Consortium (W3C), com o código aberto e adoção universal por ser autodescritivo, requerendo associação com esquemas apropriados. (FLVC, 2013; LIBRARY OF CONGRESS, 2014; NARA, 2010; WHITE, 2013)

Para fins de acesso (Quadro 2), foram definidos os formatos PDF (Portable Document Format) e ODF (Open Document Format), cujas extensões de arquivo variam dependendo do tipo específico de arquivo: .odt (processamento de texto), .ods (planilhas) e .odp (apresentações). (FLVC, 2013; LIBRARY OF CONGRESS, 2014; NARA, 2010; WHITE, 2013)

QUADRO 2 - Padronização SABER - Formato de Arquivos Textuais - Acesso

POLÍTICA DE FORMATO DE ARQUIVO (ACESSO)			
TIPO	FORMATO	BASE	RECOMENDAÇÃO
Textual	Portable Document Format (PDF)	Adobe Systems Incorporated	Recomenda-se que todos os OAs em formato textual devem ser disponibilizados no formato PDF.
Textual	ODF (Open Document Format)	ISO/IEC 26300:2006	Recomenda-se que todos os OAs em formato textual devem ser disponibilizados no formato PDF ou ODT/ODS/ODP, extensão específica para processamento de texto, planilhas e apresentações, respectivamente.

Fonte: Os Autores, 2014.

Para fins de acesso, o PDF é um formato mundialmente utilizado, com o seu código-fonte aberto e bem aplicado para textos. O formato ODF é um formato aberto desenvolvido pela Sun Microsystems e que se tornou um padrão ISO (ISO/IEC 26300:2006).

8.1.2 Imagem

Como formato de preservação para arquivos imagéticos (Quadro 3), foi definido o formato TIFF (Tagged Image File Format), que é amplamente utilizado para fins de preservação. O TIFF deve estar sem a utilização de compressão, e a resolução mínima utilizada deve ser de 300dpi (dots per inch).

QUADRO 3 - Padronização SABER - Formato de Arquivos Imagéticos - Preservação

POLÍTICA DE FORMATO DE ARQUIVO (PRESERVAÇÃO)			
TIPO	FORMATO	BASE	RECOMENDAÇÃO
Imagen	Tagged Image File Format (TIFF) TIFF, Revision 6.0 and earlier TIFF Uncompressed File with Exif Metadata	Adobe Systems Incorporated	Recomenda-se que todos os OAs nativos digitais ou gerados a partir de escâneres ou máquinas fotográficas devem possuir a resolução mínima de 300dpi e ser preservados na extensão TIFF.

Fonte: Os Autores, 2014.

Para fins de acesso, o formato escolhido (Quadro 4) foi o JPEG (Joint Photographic Experts Group). Esse é um dos formatos mais conhecidos e utilizados na internet sendo uma extensão de arquivo que possui características de perda na qualidade de imagem, porém, sua utilização para visualização, acesso e disseminação na web são recomendadas. Sua utilização atrelado aos dados EXIF (Exchangeable image file format for Digital Still Camera), possibilitam informações técnicas da imagem embutidas na forma de metadados, contribuindo para o registro de dados e informações. (FLVC, 2013; LIBRARY OF CONGRESS, 2014; NARA, 2010; WHITE, 2013)

QUADRO 4 - Padronização SABER - Formato de Arquivos Imagéticos - Acesso

POLÍTICA DE FORMATO DE ARQUIVO (ACESSO)			
TIPO	FORMATO	BASE	RECOMENDAÇÃO
Imagen	JPEG file with Exif metadata	ISO/IEC 10918 ISO/IEC 14495	Recomenda-se que todos os OAs sejam disponibilizados para fins de disseminação com o formato JPEG com dados EXIF.

Fonte: Os Autores, 2014.

8.1.3 Áudio

Para arquivos de áudio, o formato para fins de preservação adotado (Quadro 5) foi o BWF ou BWAVE (Broadcast WAVE Format), que é um formato desenvolvido pela European Broadcasting Union. Ele é uma variante do formato Microsoft Wave, porém suas especificações são abertas, autodescritivas, descompactadas e compatíveis com qualquer software que aceite a extensão .wav. (FLVC, 2013; LIBRARY OF CONGRESS, 2014; NARA, 2010; WHITE, 2013)

QUADRO 5 - Padronização SABER - Formato de Arquivos Áudio - Preservação

POLÍTICA DE FORMATO DE ARQUIVO (PRESERVAÇÃO)			
TIPO	FORMATO	BASE	RECOMENDAÇÃO
Áudio	Broadcast WAVE file, version 1, with LPCM encoded audio.	EBU Tech 3285 - Specification of the Broadcast Wave Format (BWF) - Version 1 - second edition (2001)	Recomenda-se que todos OAs em formato de áudio, principalmente os advindos do Projeto Sintonia Saúde, sejam gerados para fins de preservação no formato BWF.
	Broadcast WAVE file, version 2, with LPCM encoded audio.		

Fonte: Os Autores, 2014.

Para acesso, o formato escolhido (Quadro 6) foi MP3 (MPEG Layer III Audio Encoding), que é um formato de arquivo bastante universal e divulgado para registro sonoro em meio digital. (FLVC, 2013; LIBRARY OF CONGRESS, 2014; NARA, 2010; WHITE, 2013)

QUADRO 6 - Padronização SABER - Formato de Arquivos Áudio - Acesso

POLÍTICA DE FORMATO DE ARQUIVO (ACESSO)			
TIPO	FORMATO	BASE	RECOMENDAÇÃO
Áudio	MP3 (MPEG Layer III audio encoding)	MPEG-1: 11172-3 MPEG- 2: 13818-3	Recomenda-se que todos OAs em formato de áudio, principalmente os advindos do Projeto Sintonia Saúde, sejam gerados para fins de acesso no formato mp3.

Fonte: Os Autores, 2014.

8.1.4 Vídeo

Os formatos de preservação para vídeo escolhidos (Quadro 7) foram o MPEG-4 (Motion Picture Experts Group) salvo com compactação sem perdas de qualidade e o Motion JPEG 2000 (Joint Photographic Experts Group), usando a compactação sem perdas, nas extensões .mj2 e .mjp2. (FLVC, 2013; LIBRARY OF CONGRESS, 2014; NARA, 2010; WHITE, 2013)

QUADRO 7 - Padronização SABER - Formato de Arquivos Vídeo - Preservação

POLÍTICA DE FORMATO DE ARQUIVO (PRESERVAÇÃO)			
TIPO	FORMATO	BASE	RECOMENDAÇÃO
Vídeo	MP4 File Format MPEG-4 file format, version 2	ISO/IEC 14496- 14:2003	Recomenda-se que todos OAs em formato de vídeo, principalmente os advindos da apresentação de disciplinas, vídeos complementares, entre outros, sejam gerados para fins de preservação no formato MP4 sem compactação.
Vídeo	Motion JPEG 2000	ISO/IEC 15444- 3:2004	Recomenda-se que todos OAs em formato de vídeo, principalmente os advindos da apresentação de disciplinas, vídeos complementares, entre outros, sejam gerados para fins de preservação no formato .mjp2 ou .mjp2.

Fonte: Os Autores, 2014.

O MPEG-4 foi desenvolvido pela Motion Picture Experts Group e é um formato de padrão aberto, que oferece suporte à incorporação de metadados descritivos, para ajudar a apoiar o acesso futuro. Já o Motion JPEG 2000 é um formato padrão aberto internacional baseado na ISO/IEC 15444-3:2004. (FLVC, 2013; LIBRARY OF CONGRESS, 2014; NARA, 2010; WHITE, 2013)

Como formato para acesso, foi escolhido (Quadro 8) o formato MPEG-4 (Motion Picture Experts Group), porém salvo com compactação para possibilitar o acesso ao conteúdo do vídeo, com tamanho menor, sem perdas visuais de qualidade. (LIBRARY OF CONGRESS, 2014)

QUADRO 8 - Padronização SABER - Formato de Arquivos Vídeo - Acesso

POLÍTICA DE FORMATO DE ARQUIVO			
TIPO	FORMATO	BASE	RECOMENDAÇÃO
Vídeo	MPEG-4	ISO/IEC 14496- 2:2004	Recomenda-se que todos os OAs em formato de vídeo, sejam disseminados para fins de acesso no formato MP4 com compactação.

Fonte: Os Autores, 2014.

9 CONSIDERAÇÕES FINAIS

A implementação de formatos de arquivos para OAs com o propósito de determinar políticas específicas, que facilitem a preservação digital no SABER Tecnologias Educacionais e Sociais é apenas uma pequena iniciativa diante de um ciclo, que deverá ter continuidade, para viabilizar o acesso a longo prazo da memória científica do SABER Tecnologias Educacionais e Sociais da UFPE.

Outras vertentes que já se encontram em andamento, também visando ao armazenamento e à preservação dos objetos são: 1) a definição de diretrizes que resguardem a propriedade intelectual dos autores e contribuidores responsáveis pela criação dos OAs, levando em consideração a legislação brasileira vigente sobre direitos autorais e propriedade intelectual. 2) a criação de políticas e ações de guarda, armazenamento e backup dos conteúdos digitais, levando em conta que tanto os ambientes virtuais do grupo SABER quanto os ambientes de desenvolvimento de cursos e OAs encontram-se, hoje, hospedados na nuvem da empresa Amazon (<http://aws.amazon.com/pt/>).

Os autores agradecem ao Ministério da Saúde e à Universidade Aberta do Sistema Único de Saúde (UNA-SUS) pelo apoio necessário à execução dos trabalhos apresentados nesse artigo.

REFERÊNCIAS

- ARELLANO, M. A. Preservação de documentos digitais. **Ciência da Informação**, Brasília, v. 33, n. 2, p. 15-27, maio/ago. 2004. Disponível em: <<http://eprints.rclis.org/archive/00002940/01/Ci%5B1%5D.Inf-2004-343.pdf>>. Acesso em: 10 maio. 2014.
- BIA PLATAS, A.; SÁNCHEZ QUERO, M. **Desarrollo de una política de preservación digital**: tecnología, planificación y perseverancia. Cervante Virtual, Alicante, Espanã, p. 1-10, 2002. Disponível em: <<http://www.cervantesvirtual.com/research/articles/JBIDI02a.pdf>>. Acesso em: 26 maio 2014.
- BETTIO, R. W. de; MARTINS, A. **Objetos de aprendizado**: um novo modelo direcionado ao ensino a distância. 17 dez. 2004. Universia Brasil. Disponível em: <<http://noticias.universia.com.br/destaque/noticia/2004/12/17/493047/bjetos-aprendizado-um-novo-modelo-direcionado-ao-ensino-distncia.pdf>>. Acesso em: 10 jun. 2014.
- BROWN, A. **Digital preservation guidance note 1**: selecting file formats for long-term preservation. The National Archives, Reino Unido, p. 1-10, 2008. Disponível em: <<https://www.nationalarchives.gov.uk/documents/selecting-file-formats.pdf>>. Acesso em: 22 jun. 2014.
- CONSELHO NACIONAL DE ARQUIVOS (CONARQ). **Carta para preservação do patrimônio arquivístico digital**: preservar para garantir o acesso. Rio de Janeiro: Conarq, 2005. Disponível em: <<http://www.conarq.arquivonacional.gov.br/media/carta.pdf>>. Acesso em: 4 jun. 2014.
- ERPANET. **Digital preservation policy tool**. Glasgow: ERPANET, 2003. Disponível em: <<http://www.erpanet.org/guidance/docs/ERPANETPolicyTool.pdf>>. Acesso em: 10 jul. 2014.
- FERREIRA, M. **Introdução à preservação digital**: conceitos, estratégias e actuais consensos. Guimarães: Escola de Engenharia da Universidade do Minho, 2006. 85p. Disponível em: <<https://repositorium.sdum.uminho.pt/bitstream/1822/5820/1/livro.pdf>>. Acesso em: 10 jun. 2014.

FLVC. Florida Virtual Campus. **Recommended data formats for preservation purposes in the Florida Digital Archive**. Flórida: FLVC, nov. 2013. Disponível em: <http://fclaweb.fcla.edu/uploads/recFormats_2.pdf>. Acesso em: 2 jun. 2014.

HEDSTROM, M. Digital preservation: a time bomb for digital libraries. **Language Resources and Evaluation**, v. 31, n. 3, maio 1997. Disponível em: <<http://www.springerlink.com/content/h73v57h6587k4l7n/fulltext.pdf>>. Acesso em: 7 abr. 2008.

HODGE, G.; FRANGAKIS, E. **Digital preservation and permanent access to scientific information**: the state of the practice. Oak Ridge: CENDI, 2004. Disponível em: <http://cendi.dtic.mil/publications/04-3dig_preserv.pdf>. Acesso em: 24 jun. 2014.

LIBRARY OF CONGRESS. **Sustainability of digital formats**: planning for Library of Congress collections: format descriptions. Washington: The Library of Congress, 2014. Disponível em: <<http://www.digitalpreservation.gov/formats/fdd/descriptions.shtml>>. Acesso em: 12 jun. 2014.

LONGMIRE, W. **A primer on learning objects**. Virginia: American Society for Training & Development, 2001.

MICHEL, M. H. **Metodologia e pesquisa em ciências sociais**. 2. ed. São Paulo: Atlas, 2009.

MORAN, J. M. **As múltiplas formas do aprender**. São Paulo, Jul. 2005. Entrevista concedida a Revista Atividades & experiências. Disponível em: <<http://ucbweb.castelobranco.br/webcaf/arquivos/23855/6910/positivo.pdf>>. Acesso em: 24 jun. 2014.

NARA. The U.S. National Archives and Records Administration. **Preservation**: digital file types. College Park: NARA, 2010. Disponível em: <<http://www.archives.gov/preservation/products/definitions/filetypes.html>>. Acesso em: 10 jun. 2014.

RIMKUS, K. *et al.* Digital preservation file format policies of ARL member libraries: an analysis. **D-Lib Magazine**, v. 20, n. 3/4, mar./abr. 2014. Disponível em: <<http://www.dlib.org/dlib/march14/rimkus/03rimkus.print.html>>. Acesso em: 26 jun. 2014.

ROG, J.; VAN WIJK, C. **Evaluating file formats for longterm preservation**. Den Haag: National Library of the Netherlands, 2008. p. 1-11. Disponível em: <http://www.kb.nl/sites/default/files/docs/KB_file_format_evaluation_method_27022008.pdf>. Acesso em: 25 jun. 2014.

SÁ FILHO, C. S. e; MACHADO, E. de C. **O computador como agente transformador da educação e o papel do objeto de aprendizagem**. 17 dez. 2004. Universia Brasil. Disponível em: <<http://noticias.universia.com.br/destaque/noticia/2004/12/17/493049/computador-como-agente-transformador-da-educao-e-papel-do-objeto-aprendizagem.pdf>>. Acesso em: 10 jun. 2014.

SOUZA, J. de; VIERA, F. G. A Tecnologia Open Archives Initiative, Object Reuse and Exchange: histórico e perspectivas. **DataGramZero**, Rio de Janeiro, v. 13, n. 2, abr. 2012. Disponível em: <http://www.dgz.org.br/abr12/Art_04.htm>. Acesso em: 15 set. 2014.

WHITE, J. **Sustainable file formats for electronic records:** a guide for government agencies. Springfield: Illinois State Archives, 2013. Disponível em: <http://www.cyberdriveillinois.com/departments/archives/records_management/sustainable_formats.html>. Acesso em: 25 jun. 2014.

WILEY, D. Connecting learning objects to instructional design theory: a definition, a metaphor, and a taxonomy. In: WILEY, D. (Ed.). **The instructional use of learning objects:** on line version. Bloomington: Association for Instructional Technology, 2000. Disponível em: <<http://reusability.org/read/chapters/wiley.doc>>. Acesso em: 20 jun. 2014.

A INTERAÇÃO DO USUÁRIO COM CATÁLOGOS BIBLIOGRÁFICOS ON-LINE DE BIBLIOTECAS UNIVERSITÁRIAS⁹¹

USER INTERACTION WITH ONLINE BIBLIOGRAPHIC CATALOG FROM UNIVERSITY LIBRARIES

Flavia Maria Bastos

Silvana Aparecida Borsetti Gregorio Vidotti

Resumo: Este estudo⁹² tem como objetivo identificar as formas de interação do usuário com o catálogo bibliográfico on-line, no processo de busca da informação, de modo a encontrar recursos bibliográficos. Para o desenvolvimento desta pesquisa foi escolhida a metodologia da Teoria Fundamentada (*Grounded Theory*), visando identificar padrões durante o processo de busca de informação realizado pelo usuário assim como perceber sob que condições esses padrões são estabelecidos, quando se aplicam os dados num referido contexto. Para tanto, os dados coletados e analisados compreendem o período de 2008 a 2012, possibilitando a identificação das categorias centrais de usuários por meio da realização de diversos refinamentos desses dados alcançados com uma análise comparativa constante até chegar à saturação deles dados. Os dados foram agrupados em dimensões de opções de busca, *staff*, bases de dados, áreas de conhecimento, e tempo (ano, mês e semestre), que identificam diversos padrões nas formas de interação das categorias centrais de usuário com o catálogo bibliográfico *on-line*. Os resultados são apresentados por meio de características gerais dos padrões evidenciados pelos dados e também pelas categorias centrais de usuário. Desta forma, os dados demonstram o perfil e a preferência do usuário por uma interação rápida com o sistema, que se caracteriza pelo anonimato, em função da maioria dos usuários não se identificar ao realizar suas pesquisas e ao utilizar a opção de busca default do sistema. Evidenciam o comportamento atual do usuário, que deve ser absorvido pelos gestores de catálogos bibliográficos durante suas ações futuras, norteando-lhes as práticas biblioteconómicas com foco no usuário.

Palavras-chave: Catálogos bibliográficos. Teoria Fundamentada. OPAC – *Online Public Access Catalog*.

Abstract: The objective of this research is to identify the ways user interact with Online Public Access Catalogs in his process of information search to find bibliographic resources. For the development of this research the chosen methodology was the Grounded Theory, in order to identify patterns during the information search, as well as the conditions under which these patterns are established when the data is applied in a context. To achieve this, the collected and analyzed data cover the period from 2008 to 2012, allowing the identification of core categories of users through various refinements of the data with a constant comparative analysis until reaching data saturation. Thus, the data were gathered into dimensions of search options, staff, databases, knowledge areas, and time (year, month and semester), allowing the identification of different forms of interaction patterns from the core categories of users with the online bibliographic catalog. Results are presented by

⁹¹ Texto derivado: BASTOS, F. M. **A interação do usuário com catálogos bibliográficos on-line** : investigação a partir da Teoria Fundamentada. 2013. 257f. Tese (Doutorado em Ciência da Informação) – Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília, 2013.

⁹² Financiamentos da pesquisa: CNPq PQ – Processo: 308443/2010-1; CNPq CHSSA – Processo: 407149/2012-0; CNPq Universal – Processo: 486147/2011-8.

general characteristics of the highlighted data and also by the core categories of users. The data demonstrate the profile and user preferences for a quick interaction with the system, characterized by anonymity, since most users do not identify themselves when conducting their research and use the default search option to the system. This way, the patterns of interaction demonstrate the actual user search behavior, which must be absorbed and understood by managers of bibliographic catalogs during their future actions, in order to guide Librarianship practices with focus on the end user.

Keywords: Bibliographic Catalog. Grounded Theory. OPAC - Online Public Access Catalog.

1 INTRODUÇÃO

O presente estudo apresenta como tema a interação do usuário com o catálogo bibliográfico on-line durante o processo de busca de informações.

Os catálogos bibliográficos, inseridos no contexto histórico das bibliotecas em diversos momentos de sua evolução, no que tange ao desenvolvimento e ao uso de tecnologias, são alvos constantes de estudos (VILLÉN-RUEDA, et al., 2007). Visam não somente mensurar a eficácia do seu uso, como também revelar os problemas ocorridos durante o processo de busca de informações.

Para tanto, considerando esses estudos, bem como a variedade de metodologias aplicadas para avaliar o uso do catálogo on-line, questionamos como os usuários interagem com os catálogos on-line de uma rede de bibliotecas universitárias, de modo a encontrar recursos bibliográficos.

Assim, a partir desse questionamento estabelecemos como objetivo identificar a forma de interação do usuário com o catálogo on-line de uma rede de bibliotecas universitárias, por meio de um processo de busca de informação, de modo a encontrar recursos bibliográficos.

Com este objetivo nos colocamos diante das diferentes formas de olhar, entender e avaliar o usuário durante sua interação com o catálogo bibliográfico on-line, analisando, mediante o seu comportamento os dados produzidos e armazenados internamente no sistema.

Para o desenvolvimento deste trabalho foi aplicada a Teoria Fundamentada em Dados (Grounded Theory), que neste caso não apresenta como objetivo coletar e analisar dados para comprovar uma hipótese, mas sim produzir um conhecimento sobre as formas de interação em catálogos automatizados. Desta forma, não se tem a fixação de uma teoria a priori, mas sim a construção de uma reflexão teórica, tendo como subsídios os dados coletados e as análises que geraram a necessidade de outras coletas de dados. Essa reflexão foi estabelecida por um processo dinâmico que possibilitou a construção sistematizada de

evidências sobre a realidade da interação dos usuários com os Online Public Access Catalogs (OPAC), visando à identificação de padrões, bem como as condições que evidenciam um determinado comportamento dos dados dentro de um contexto.

Os procedimentos metodológicos foram aplicados aos dados coletados no catálogo bibliográfico on-line da UNESP (Catálogo Bibliográfico Athena) referente ao período de 2008 a 2012, que possibilitaram a identificação das categorias centrais de usuários por meio dos diversos refinamentos dos dados produzidos e alcançados com análise comparativa constante até chegar à saturação dos dados. Assim, os dados foram agrupados nas seguintes dimensões: opções de busca, equipe da biblioteca (staff), bases de dados, áreas de conhecimento e tempo (ano, mês e semestre), que possibilitaram identificar diversos padrões nas formas de interação das categorias centrais de usuário com o catálogo bibliográfico on-line.

O estudo foi estruturado em quatro seções, incluindo a introdução. A segunda seção apresenta a revisão de literatura na área voltada para a interação do usuário, a metodologia a partir do seu conceito, características e procedimentos que envolvem a Teoria Fundamentada em Dados (Grounded Theory). Na seção seguinte, são explicitados e discutidos os resultados obtidos e em seguida tem-se a conclusão.

Justificamos este estudo por considerarmos de extrema importância a descoberta de modelos que representem a forma como os usuários utilizam os produtos e serviços oferecidos pelas bibliotecas, visando ao aumento da interação entre o usuário e o sistema, a partir da integração e retroalimentação dessas informações para os ambientes informacionais.

As ações que ocorrem durante a prática biblioteconômica podem ser utilizadas em benefício da gestão do próprio sistema informacional, bem como para o fortalecimento da área, com base em subsídios teóricos e práticos que apresentem formas de processar os dados gerados a partir da interação dos usuários.

2 REVISÃO DE LITERATURA

Na revisão de literatura apresentamos o posicionamento de alguns autores sobre a influência da tecnologia na evolução das bibliotecas e consequentemente dos catálogos bibliográficos, assim como sobre o reflexo da área da Ciência da Informação na interação entre o usuário e o sistema, quando busca informações em uma biblioteca, por meio dos catálogos automatizados.

Apesar das grandes mudanças tecnológicas, absorvidas pela estrutura social e proporcionadas pelo avanço das Tecnologias de Informação e Comunicação (TIC), a

biblioteca, como agente mediador na produção e nas formas estratégicas para a distribuição da informação, necessita considerar os efeitos da implantação e do uso das TIC a partir de uma reflexão constante do rumo que os ambientes informacionais, tais como os sistemas de bibliotecas, estão tomando, principalmente sem observar a interação desses ambientes com o próprio usuário.

Com a Internet, a interação do usuário com o sistema de biblioteca passa a ser mais intensa, do processo de acesso à recuperação e ao uso da informação, com várias opções de acesso a diversas fontes de informação. A partir dessas mudanças, notamos modificações no comportamento do usuário quando produz conteúdo, quando navega nas redes digitais e atua nos ambientes informacionais através de práticas colaborativas.

Nesse contexto, as bibliotecas universitárias mantêm suas funções ligadas às atividades voltadas para o ensino, pesquisa e extensão, ofertando serviços com foco no atendimento das necessidades de informação dos vários grupos de usuários, de forma a acompanhar seu constante desenvolvimento acadêmico-científico.

Os serviços gerados pelas bibliotecas universitárias, intensamente influenciados pelas tecnologias, passam a ter as suas rotinas racionalizadas a partir dos sistemas de automação, que agilizam os processos de recuperação e a transferência da informação, tornando-as acessíveis para toda comunidade científica.

Como parte do contexto histórico das bibliotecas, os catálogos bibliográficos estiveram presentes em diversos momentos de sua evolução e começaram como simples inventários ou relação do conteúdo de determinada coleção (SHERA, 1969).

Os catálogos, considerados um dos principais meios de interação do usuário para recuperação das informações armazenadas pelos sistemas de bibliotecas, com os avanços na tecnologia, passam a oferecer aos usuários consultas aos catálogos das bibliotecas via Web. Estes avanços proporcionaram a expansão dos catálogos automatizados presentes na maior parte das bibliotecas contemporâneas, denominados Catálogos On-line de Acesso Público (Online Public Access Catalogs – OPAC), que passaram a ser considerados um dos módulos integrados do sistema de gerenciamento de bibliotecas, com suas funções, objetivos e composição.

No entanto, a partir da influência das TIC em ambientes informacionais, como são as bibliotecas, aumenta a necessidade de se ampliar o olhar sobre as novas formas de interação do usuário com a biblioteca, em especial com o catálogo, de forma que os usuários passem a ter autonomia na busca e na recuperação da informação desejada.

Podemos considerar que as bibliotecas também enfrentam o desafio de assumir condutas e práticas profissionais em consonância com uma perspectiva teórica da área, no que tange a incorporação das tecnologias de comunicação e informação.

Como ciência interdisciplinar, Saracevic (1997) afirma que a Ciência da Informação está relacionada com as Tecnologias de Informação e Comunicação, tendo uma forte dimensão social e humana. Nesse sentido, o autor ressalta o interesse no diálogo com a recuperação de informação multimídia e multilíngue, as bibliotecas virtuais, as buscas na internet e os estudos de interação.

Na visão de Le Coadic (2004), o uso de um produto de informação ou de um sistema de informação está relacionado a um efeito que satisfaça a uma necessidade de informação, compreendendo a utilização, o uso e o consumo, considerando, assim, que a função mais importante do produto ou do sistema é a maneira como a informação modifica a realização das atividades dos usuários.

Le Coadic (2004) menciona um enfoque tradicional da Ciência da Informação quando esta considera que o usuário chega a um sistema de informação com uma necessidade de informação mais ou menos bem especificada; no entanto, o autor ressalta que para o sistema oferecer ao usuário um melhor desempenho, a área deveria considerar uma análise de suas necessidades de informação, uma vez que o ato de consultar o catálogo pressupõe que este irá ajudá-lo a sanar suas necessidades informacionais.

Nesse sentido, reconhecemos a importância do conhecimento sobre as circunstâncias que levam um usuário a dar início a um processo de busca de informações, para compreendermos o efetivo uso dos diferentes sistemas, serviços e produtos a partir da própria atuação do usuário em ambientes informacionais. Além disso, podemos considerar que o ato de consultar um catálogo pressupõe que este irá ajudar o usuário a sanar suas necessidades informacionais, pois a disponibilização de um sistema de informação só se justifica a partir do momento em que ele é aceito pelo usuário, estabelecendo-se assim, uma troca mútua entre ambos.

Le Coadic (2004) indaga o tipo de diálogo existente num processo de interação informacional, afirmindo que o componente central de todo sistema de informação é a interação entre o usuário e o sistema, diretamente ou por intermédio de um terceiro, ocasionando um diálogo entre dois participantes, sejam eles pessoa-pessoa ou pessoa-máquina.

Lancaster (1996), ao refletir sobre os estudos referentes ao uso do catálogo divide-os em duas categorias principais: os que se destinam a determinar a proporção dos usuários da

biblioteca que fazem uso do catálogo, mencionando as características dos usuários e as dos não usuários, e os estudos dos usuários que “sabidamente” usam o catálogo, sendo possível descobrir como o usam e com qual finalidade, entre outras informações.

Lancaster (1996) também considera que a maioria dos estudos realizados se concentra nos usuários que sabem utilizar o catálogo, pois nesses estudos os usuários são observados durante o uso, proporcionando condições de reunir dados sobre padrões de uso sem que eles saibam que suas ações estão sendo observadas.

Outra observação importante de Lancaster (1996, p.127) sobre os estudos de uso dos catálogos diz respeito à forma como as buscas são realizadas; assim, uma busca pode ser para:

- 1) Determinar se a biblioteca possui ou não determinado livro ou outro item sobre autor ou título, ou ambos;
- 2) Identificar itens que a biblioteca possui e que tratam de determinado assunto – busca por assunto (LANCASTER, 1996, p.49).

O autor (LANCASTER, 1996) considera que a avaliação da busca por um item conhecido é muito mais fácil do que com as buscas por assunto, pois na primeira o usuário encontra o que procura ou não encontra; já na busca por assunto, essa condição binária não se aplica, uma vez que, na avaliação das buscas por assunto, para se considerar uma busca bem-sucedida o usuário deverá ser capaz de obter uma coincidência entre seus termos de busca e os que são usados nos catálogos.

Nesse contexto, Santos (2010) também expõe que o processo de busca de informações envolve a identificação e a avaliação do usuário em relação às representações dos recursos informacionais existentes nos ambientes informacionais que podem influenciar sua decisão no momento do acesso.

Desta forma, investigamos a interação do usuário durante seu processo de busca de informação no catálogo de bibliotecas da Unesp (Catálogo Bibliográfico Athena), extraíndo os elementos que indicam a interação do usuário via interface do catálogo.

A Rede de Bibliotecas da Unesp deu início ao seu processo de automação em 1994, com a nítida preocupação, nesse percurso, não apenas com a construção do catálogo, bem como com a qualidade de seus registros bibliográficos, o que se constata pelas decisões tomadas ao longo desses anos, como a adoção de um padrão de formato internacional para as informações bibliográficas: - a abrangência da base no que tange ao acesso a todo acervo da universidade por toda comunidade usuária; - a produção do registro bibliográfico visando à competitividade e à qualidade, tendo o registro bibliográfico como elemento formador de uma base de dados com alcance e visibilidade nacional e internacional; - a

uniformidade das entradas visando oferecer ao catalogador / indexador subsídios para a adoção de termos uniformes na descrição dos documentos; - e pontos de acesso ao documento que permitem a recuperação de um registro bibliográfico por qualquer dos elementos que o compõem, e que são possibilidades presentes até os dias de hoje.

Esse fato demonstra como as convenções estabelecidas pelas comunidades de prática, compostas pelos bibliotecários da Rede Unesp, permaneceram ao longo desse período, acompanhando os desenvolvimentos da área, influenciados pelos avanços tecnológicos.

Os dados utilizados nesta pesquisa foram extraídos do catálogo bibliográfico on-line da Universidade Estadual Paulista UNESP, composta atualmente por 34 bibliotecas distribuídas no Estado de São Paulo, e que disponibiliza seu catálogo bibliográfico on-line desde 1998 à comunidade acadêmica.

Conforme os dados do Anuário Estatístico da UNESP de 2013, o acervo bibliográfico está composto por 1.350.030 livros; 28.227 títulos de periódicos; 22.080 CDs, vídeos e DVDs; 65.005 teses; e 150.488 outros materiais.

Ainda conforme o Anuário (2013), a Universidade apresentava o total de 3.625 docentes; 35.485 alunos de graduação; e 11.804 alunos de pós-graduação, totalizando 50.914 usuários potenciais do catálogo bibliográfico on-line, que realizaram 1.542.441 empréstimos e 583.9984 consultas.

Após a contextualização do ambiente informacional da Rede de Bibliotecas da Unesp, apresentamos os princípios teóricos da Teoria Fundamentada em Dados que nortearam sistematicamente os dados extraídos, produzidos a partir do uso do catálogo.

3 METODOLOGIA

Para o desenvolvimento deste estudo foi escolhida a Teoria Fundamentada em Dados que tem sua origem na linha de pensamento de dois sociólogos: Glaser e Strauss (1967), que consideram a análise comparativa como um aspecto central dessa teoria, também citada como um método comparativo constante.

De acordo com Gasque (2008), a Teoria Fundamentada em Dados pode ser considerada uma metodologia de natureza exploratória, orientada para os dados, pois estes são coletados sem a preocupação de testar uma teoria está relacionada ao caráter exploratório e indutivo da pesquisa, com a possibilidade de gerar teorias a partir de relações conceituais.

A teoria caminha essencialmente a partir da interpretação dos dados analisados por meio das etapas de codificação, que incluem as operações de constante comparação dos

fenômenos, casos, conceitos, além de questões dirigidas ao texto que direcionam o pesquisador para o desenvolvimento de teorias mediante um processo de abstração.

Assim, os dados coletados necessitam ser criteriosamente categorizados, classificados para que se obtenha uma análise correta dos elementos e de suas articulações. Essa categorização é fundamental, pois a reflexão sobre ela fornece categorias analíticas que contribuirão para a construção da pesquisa.

Inicialmente, o método exige uma descrição exaustiva dos processos estudados e dos resultados obtidos e, em seguida, uma análise interpretativa, que inclui a codificação dos dados obtidos, observando-se a regularidade apresentada, buscando assim o significado desses dados em um processo de conceitualização, a partir da descrição dos fenômenos observados.

A Teoria Fundamentada em Dados já foi adotada em alguns estudos da área da Ciência da Informação derivados da década de 1980, em Sheffield, com destaque para os trabalhos desenvolvidos por Wilson (1981), Brown (1990), Soto (1992), Ellis (1993), Powell (1999), e Mansourian (2006), que contribuíram para os estudos sobre o comportamento de busca de informação, conforme menciona Sélden (2005, p.120) e Tan (2010, p.94).

A Ciência da Informação, como área do conhecimento que se estabelece em um contexto social contemporâneo, demanda métodos que possam ser aplicados no desenvolvimento de estudos voltados para os comportamentos de busca, de recuperação, de necessidade e de uso de informações presentes nas relações sociais nas quais os indivíduos estão inseridos.

No contexto da Teoria Fundamentada em Dados, a análise dos dados é sistematizada a partir da codificação, onde os dados coletados são fragmentados, conceituados e reunidos de uma nova maneira. A codificação é composta por três etapas importantes: a codificação aberta (open coding); a codificação axial (axial coding); e a codificação seletiva (selective coding).

A codificação aberta tem por objetivo expressar os dados e os fenômenos na forma de conceitos. Para isto, os dados são fragmentados e classificados em expressões para serem relacionados às anotações e aos códigos.

Os dados utilizados nesta pesquisa foram extraídos do histórico dos eventos de buscas realizados pelos usuários e que são registrados na tabela de dados que juntamente com as demais tabelas de configuração do software, é armazenada no banco de dados Oracle, responsável por armazenar e recuperar os dados utilizados pelo sistema.

Assim, a partir da identificação dos eventos registrados na tabela de dados extraímos e mapeamos o histórico das interações de um usuário relacionadas a pesquisas, navegação e solicitações DSI porém, para atender aos objetivos desta pesquisa coletou-se o histórico das interações do usuário na busca por informações em registros bibliográficos.

Os eventos da tabela de dados correspondem às ações realizadas pelos usuários na busca (por palavras) da interface OPAC WEB. Esta interface oferece aos usuários dois tipos de buscas, sendo elas: por lista e por palavras. Este estudo concentra-se nas interações dos usuários que utilizaram a busca por palavra, em suas opções de campos disponíveis, para realizarem suas consultas aos catálogos.

Os dados utilizados nesta pesquisa correspondem ao período de 01 de janeiro de 2008 a 31 de dezembro de 2012. Nesta coleta optou-se pela geração de um programa para leitura e extração dos dados, ao invés de um comando de consulta – select – no banco de dados, em função de uma série de motivos, sendo o principal a repetição do comando de consulta para a retirada dos dados, que acarretaria uma possível lentidão no sistema, em função do volume dos dados existentes.

Para a leitura e o processamento de cada linha dos registros armazenados na tabela foram desenvolvidos programas na linguagem própria dos bancos de dados relacionais para essencialmente, realizar a contagem dos registros utilizados neste estudo.

Os dados extraídos do catálogo OPAC WEB foram analisados de acordo com as categorias de usuários estabelecidas por cada biblioteca, o que contribuiu para a elaboração do critério de reagrupamento das categorias iniciais coletadas, possibilitando-nos identificar as categorias centrais de usuário, conforme indicações metodológicas da Teoria Fundamentada em Dados. O critério adotado como base para o reagrupamento das categorias iniciais em centrais consistiu em agrupar categorias que dizem respeito aos níveis de atuação na universidade: aluno de graduação, aluno de pós-graduação e docente. No entanto, foi necessário criar o grupo “outras categorias”, onde inserimos todas aquelas que poderiam ser classificadas em uma ou mais das categorias adotadas para o estudo, bem como as que não faziam parte de nenhuma das categorias anteriores.

Na próxima seção apresentamos a análise dos dados, e os resultados obtidos após a aplicação dos princípios teóricos da Teoria Fundamentada em Dados que nortearam sistematização dos dados extraídos do catálogo.

3.1 Análise dos dados e resultados da pesquisa

Os dados utilizados neste estudo foram autorizados pela Coordenadoria Geral de Bibliotecas (CGB) que gerencia o catálogo bibliográfico on-line da Unesp.

Conforme a sistemática apresentada na seção anterior, no período de 01 de janeiro de 2008 a 31 de dezembro de 2012 foram coletados os dados de 11.950.812 pesquisas realizadas no tipo de busca por palavras realizadas no catálogo, sendo que 8.141.292 pesquisas foram realizadas dentro da Rede Unesp, correspondendo a 68% dos acessos, e 32% dos acessos foram realizados da Rede externa a Unesp, conforme TABELA 1.

TABELA 1 - Total de acessos por IP

<hr/> Acessos por IP <hr/>		
Rede Unesp	8.141.292	68%
Rede Externa	3.809.520	32%
TOTAL	11.950.812	100%

Fonte: adaptado de BASTOS (2013, p.137)

Os dados evidenciaram que do total de 11.950.812 consultas realizadas no catálogo, 11.018.340 correspondem a 92% de usuários não identificados no sistema, sendo que apenas 932.472 usuários se identificaram no sistema informando seu número de identificação e senha, conforme TABELA 2.

TABELA 2 - Categoria de usuários identificados e não identificados

IDENTIFICADOS	932.472	8%
NÃO IDENTIFICADOS	11.018.340	92%
TOTAL	11.950.812	100%

Fonte: adaptado de BASTOS (2013, p.139)

A partir da coleta de dados no sistema identificamos as categoriais iniciais de usuário que fazem parte da política de empréstimo de cada biblioteca local, sendo que, conforme a necessidade, são criadas novas categorias. Num primeiro momento foi necessário listar cada uma das categorias para posteriormente estabelecer uma definição, que contribuiu para a elaboração do critério de reagrupamento das categorias iniciais coletadas nos possibilitou identificar as categorias centrais de usuário, conforme indicações metodológicas da Teoria Fundamentada em Dados.

Desta forma, as categorias centrais identificadas e reagrupadas com base nos níveis de atuação na universidade foram: o aluno de graduação, o aluno de pós-graduação e o docente. Foi necessário criar o grupo “outras categorias” para inserirmos uma ou mais

categorias adotadas para o estudo, bem como as que não faziam parte de nenhuma das categorias anteriores, conforme TABELA 3.

TABELA 3 - Percentual do total de buscas realizadas pelas categorias centrais

TOTAL GERAL	11.950.812	%
Aluno de graduação	797.509	6,67%
Aluno de pós-graduação	77.289	0,65%
Docente	11.335	0,09%
Não identificado	10.383.866	86,89%
Outras categorias	680.813	5,70%

Fonte: adaptado de BASTOS (2013, p.158)

Para uma análise geral da preferência nas opções de busca pelas categorias de usuários, apresentamos uma compilação dos dados coletados, conforme TABELA 4.

TABELA 4 - Total de buscas (por palavras) por todos os campos, palavras de título, palavras de autor e palavras de assunto, realizadas por todas as categorias de usuários

Total de buscas por palavras realizadas entre janeiro de 2008 e dezembro de 2012		
Campo de busca	Número	%
Todos os campos	6.731.664	59,57%
Palavras de título	1.974.415	17,47%
Palavras de autor	1.950.354	17,26%
Palavras de assunto	643.170	5,69%
TOTAL	11.299.603	100,00%

Fonte: adaptado de BASTOS (2013, p.164)

Dando continuidade à análise geral sobre as preferências dos usuários em relação às opções de busca, apresentamos um quadro geral com a compilação dos dados coletados, relacionando as categorias centrais de usuário e todas as opções de busca consideradas neste estudo, conforme QUADRO 1.

QUADRO 1 - Total de buscas por palavras em todos os campos, título, autor e assunto e por categorias centrais de usuários

Total de buscas (por palavras) realizadas entre janeiro de 2008 e dezembro de 2012

Campo de busca	Categoria de usuário			
	Não identificado	A - Aluno de graduação	B - Aluno de Pós-graduação	C - Docente
Todos os campos	6.160.735	504.701	43.944	5.329
Palavras de título	1.823.927	127.562	13.624	2.436
Palavras de autor	1.792.156	134.372	16.122	3.129
Palavras de assunto	607.048	30.874	3.599	441
TOTAL	10.383.866	797.509	77.289	11.335

Fonte: adaptado de BASTOS (2013, p.152)

Com base nos dados do QUADRO 1 analisamos a média das pesquisas realizadas por categorias de usuários na busca em todos os campos, palavras de título, autor e assunto, conforme apresentamos na TABELA 5.

TABELA 5 - Média das pesquisas realizadas por categorias de usuários na busca em todos os campos, palavras de título, autor e assunto

	Todos os campos	Palavra de título	Palavra de autor	Palavra de assunto
Não identificado	59,33%	17,57%	17,26%	5,85%
Aluno de graduação	63,28%	16,00%	16,85%	3,87%
Aluno de pós-graduação	56,86%	17,63%	20,86%	4,66%
Docente	47,01%	21,49%	27,60%	3,89%
Média	56,62%	18,17%	20,64%	4,57%

Fonte: adaptado de BASTOS (2013, p.159)

Assim, considerando a média de usuários percebemos que a maioria das categorias centrais utiliza a opção de busca configurada como padrão na interface, ou seja, a busca por todos os campos, enquanto que os demais usuários selecionam uma opção de busca diferente do padrão utilizado, conforme TABELA 6.

TABELA 6 - Escolha da opção de busca

Opção de busca padrão	56,62%
Escolha da opção de busca	43,38%
100,00%	

Fonte: adaptado de BASTOS (2013, p.161)

Desta forma, com o intuito de compreendermos a interação do usuário com o catálogo, durante seu processo de busca de informações, identificamos os atributos de relacionamentos que poderiam ser considerados na extração de dados para formar as dimensões: tempo, bases de dados, área/curso, e opções de busca, visando oferecer subsídios para identificarmos o processo de interação das categorias centrais de usuários com o catálogo, durante seu processo de busca de informações.

Assim, para cada dimensão foram realizados diversos refinamentos, apresentados com dados sintetizados a seguir:

Dimensão opção de busca – foram identificados os campos de busca com maior incidência de uso, e como esses campos fazem parte da configuração padrão do sistema, concentrarmos neles nossa análise, descartando outras opções de busca. Um demonstrativo do montante das pesquisas realizadas podem ser visualizadas na TABELA 7.

TABELA 7 - Total geral – busca (por palavras) nos campos: todos os campos, palavras de título, autor, assunto e outros campos.

Período:	01/01/2008 a 31/12/2012	Total GERAL	%
Todos os campos	6.731.664	56,33%	
Palavra de título	1.974.415	16,52%	
Palavra de autor	1.950.354	16,32%	
Palavra de assunto	643.170	5,38%	
Outros campos	651.209	5,45%	
TOTAL GERAL		11.950.812	100,00%

Fonte: adaptado de BASTOS (2013, p.162)

Dimensão *staff* – para identificação dos dados para esta dimensão mapeamos os IPs (*Internet Protocol*) das máquinas utilizadas pelos bibliotecários pertencentes às equipes de trabalho das bibliotecas da Unesp e, com isso, foi possível separar o total de pesquisas realizadas pela equipe do total de pesquisas realizadas pelos usuários, conforme TAB. 8.

TABELA 8 - Total de pesquisas realizadas através de equipamentos das bibliotecas da Unesp

	Pesquisas	%
NÃO STAFF	11.151.969	92,81%
STAFF	798.843	7,19%
TOTAL	11.950.812	100%

Fonte: adaptado de BASTOS (2013, p.171)

As equipes de trabalho estão divididas na Seção Técnica de Referência, Atendimento ao Usuário e Documentação (STRAUD) e a Seção Técnica de Aquisição e

Tratamento da Informação (STATI). No quadro 2, detectamos que a preferência, tanto dos bibliotecários de ambas as seções como do usuário não identificado, era pelo campo – todos os campos, sendo seguido pelo campo – palavras de título, sendo a última opção de escolha o campo – palavras de assunto.

QUADRO 2 - Campo pesquisado por seção

	Total	Todos os campos	Palavra de autor	Palavra de título	Palavra de assunto
Não identificado	11.151.969	6.322.655	1.834.092	1.802.788	601.442
STRAUD	526.094	302.875	65.008	103.851	22.879
STATI	258.929	99.818	49.644	63.352	18.150
STATI-STRAUD	13.820	6.316	1.610	4.424	699
TOTAL	11.950.812	6.731.664	1.950.354	1.974.415	643.170

Fonte: adaptado de BASTOS (2013, p.173)

Dimensão bases – para esta análise foram consideradas as trinta e cinco bases lógicas, que possibilitam a recuperação do acervo individual de cada base e o catálogo coletivo, conhecido pela sigla UEP. Os dados analisados apresentados demonstram que os usuários preferem utilizar as bases locais ao catálogo coletivo, sendo que nas bases locais a busca se dá por um item conhecido enquanto que na UEP, por reunir o acervo de todas as bases, a busca é mais exploratória, conforme os dados da TABELA 9.

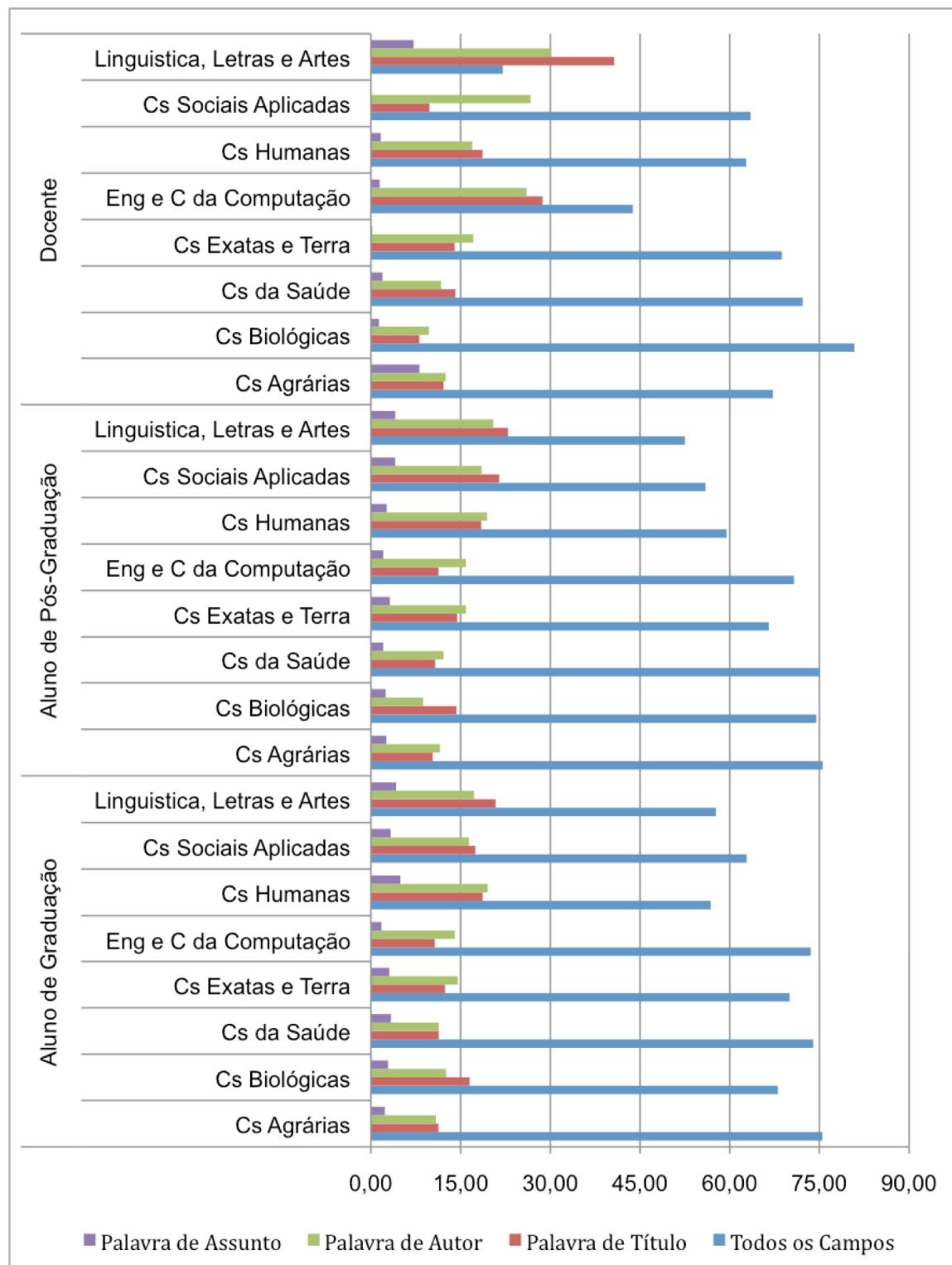
TABELA 9 - Todos os campos, palavras de título, autor e assunto – Catálogo Athena e Média das Bases das Bibliotecas

	TOTAL	Todos os campos	%	Palavra de título	%	Palavra de autor	%	Palavra de assunto	%
UEP01 - Catálogo Athena Média das Bibliotecas (%)	4.046.205	2.770.262	68,57	601.702	14,87	531.975	13,15	142.266	3,52

Fonte: adaptado de BASTOS (2013, p.177)

Dimensão áreas de conhecimento – para análise dos dados referentes às áreas de conhecimento foram consideradas as informações relacionadas aos cursos de aproximadamente 260 mil usuários, pois as informações sobre os cursos nem sempre são inseridas pelas equipes de trabalho no cadastro do usuário. Os cursos existentes na Unesp foram agrupados seguindo a tabela de áreas do Conselho Nacional de Pesquisa (CNPq). A análise dos dados com o percentual de utilização pelos usuários e campos de busca podem ser visualizados na FIG. 1.

FIGURA 3 - Percentual de utilização de usuários e campos de busca por área de conhecimento



Fonte: adaptado de BASTOS (2013, p.189)

Dimensão tempo – diante da amplitude do período, de janeiro de 2008 a dezembro de 2012, procuramos identificar a forma de interação do usuário com o catálogo on-line considerando a dimensão temporal, sendo que para isso extraímos os dados relativos a ano,

semestre e mês. Esses dados foram exaustivamente refinados e para que se tenha essa dimensão, apresentamos somente a síntese realizada com os dados relativos ao ano, conforme QUADRO 3.

QUADRO 3 - Total de buscas (por palavras) por todos os campos, palavras de título, autor e assunto – por categoria de usuários – 2008 a 2012

		Todos os campos	Palavras de título	Palavras de autor	Palavras de assunto
2008	Não identificado	1.054.101	351.625	365.191	147.727
	A – Aluno de graduação	100.106	28.408	31.405	9.378
	B – Aluno de pós-graduação	8.864	3.322	4.091	1.126
	C – Docente	835	325	503	60
2009	Não identificado	1.300.399	400.445	415.096	157.899
	A – Aluno de graduação	110.069	30.628	32.024	7.945
	B – Aluno de pós-graduação	8.870	2.709	3.366	928
	C – Docente	848	504	557	137
2010	Não identificado	1.312.744	376.928	368.547	115.267
	A – Aluno de graduação	109.108	26.019	28.011	5.426
	B – Aluno de pós-graduação	8.821	2.769	3.218	483
	C – Docente	1.040	413	688	68
2011	Não identificado	1.324.926	364.091	338.762	104.751
	A – Aluno de graduação	104.160	23.340	23.678	4.609
	B – Aluno de pós-graduação	9.574	2.475	3.145	573
	C – Docente	1.381	540	735	136
2012	Não identificado	1.168.565	330.838	304.560	81.404
	A – Aluno de graduação	81.258	19.167	19.254	3.516
	B – Aluno de pós-graduação	7.815	2.349	2.302	489
	C – Docente	1.225	654	722	40

Fonte: adaptado de BASTOS (2013, p.190)

Com o estabelecimento das categorias a serem analisadas neste estudo realizamos a etapa de codificação aberta que, conforme a Teoria Fundamentada em Dados, comprehende o agrupamento das categorias. E em função da grande quantidade de dados relacionados às ações dos usuários percebemos que era possível realizar agrupamentos das categorias de usuários com os padrões identificados a partir dos dados da interação do usuário com o catálogo OPAC.

4 CONCLUSÕES

Neste estudo apresentamos as formas de interação do usuário com os catálogos bibliográficos on-line de uma rede de bibliotecas universitárias, durante o processo de busca de informação.

Para identificarmos as formas de interação aplicamos as etapas da Teoria Fundamentada em Dados, a partir dos dados coletados no catálogo bibliográfico on-line da Unesp, e realizamos diversos refinamentos que possibilitaram a identificação de categorias centrais de usuários e de seus padrões, utilizados durante o processo de busca por informação.

Dessa forma, a preferência da maioria dos usuários ao interagir com o catálogo durante seu processo de busca demonstra que eles optam por uma busca rápida, adotando a opção de busca padrão do catálogo e de forma anônima, ou seja, sem se identificarem no sistema. Os dados evidenciam que apesar dos recursos disponíveis nos catálogos bibliográficos, em função dos avanços tecnológicos, os usuários não utilizam os serviços oferecidos.

Outra evidência observada nos dados é que os usuários, ao realizarem suas opções de busca no catálogo preferem realizá-las por um item conhecido, utilizando a busca por autor ou por título, deixando em última opção a busca por assunto, que possibilitaria a exploração da coleção sobre um determinado assunto.

A identificação de um perfil de usuário, estabelecido a partir da evidência de dados coletados das suas próprias ações, oferece subsídios para os gestores de bibliotecas universitárias fundamentarem seus planejamentos por meio de ações futuras adequadas à real necessidade informacional da comunidade usuária. Sendo assim, tanto é possível realizar ajustes junto à interface, como antecipar tendências relacionadas à disponibilização de novos serviços online.

Durante o levantamento dos dados também observamos a possibilidade e desenvolvimento de pesquisas futuras complementares a essa, considerando quantidade e qualidade dos dados armazenados no banco de dados que gerencia o sistema.

Assim, os padrões de interação do usuário com o catálogo bibliográfico on-line da Unesp, que evidenciam o comportamento atual do usuário, devem ser absorvidos pelos gestores de catálogos bibliográficos durante suas ações futuras para nortear as práticas biblioteconómicas com foco no usuário.

REFERÊNCIAS

BASTOS, F. M. A interação do usuário com catálogos bibliográficos on-line : investigação a partir da Teoria Fundamentada. 2013. 257f. Tese (Doutorado em Ciência da

Informação) – Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília, 2013.

CONSELHO NACIONAL DE DESENVOLVIMENTO TECNOLÓGICO. Tabela das Áreas do Conhecimento. Brasília, CNPq, 2012. Disponível em:
<http://www.cnpq.br/documents/10157/186158/TabeladeAreasdoConhecimento.pdf>.
 Acesso em: 20 out. 2013.

GASQUE, K. C. G. D. **O pensamento reflexivo na busca e no uso da informação na comunicação científica.** 2008. 240f. Tese (Doutorado em Ciência da Informação) - Faculdade de Economia, Administração, Contabilidade e Ciência da Informação, Universidade de Brasília, Brasília, 2008.

GLASER, B. G.; STRAUSS, A. L. **The discovery of grounded theory.** Chicago: Aldine, 1967.

LANCASTER, F. W. **Avaliação de serviços de bibliotecas.** Tradução de Antonio Agenor Briquet de Lemos. Brasília: Briquet de Lemos, 1996. 356 p.

LE COADIC, Y-F. **A ciência da informação.** Brasília: Briquet de Lemos, 2004.

ONLINE PUBLIC ACCESS CATALOGUE. Disponível em:
http://www.librarynet.com.my/pls/angkasa/opac3.search_form?pc_product=OPAC.
 Acesso em: 20 out. 2013.

SANTOS, P. L. V. A. da C. **Catalogação revisitada:** sua história e contemporaneidade. 2010. 168f. Tese (Livre Docência) – Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília, 2010.

SARACEVIC, T. Interdisciplinary nature of information Science. **Ciência da Informação,** Brasília, v. 24, n. 1, p. 36-41, 1997.

SÉLDEN, L. On the Grounded Theory – with some malice. **Journal of Documentation,** Bingley, v. 61, n. 1, p. 114-129, 2005.

SHERA, J. H.; EGAN, M. E. **Catálogo sistemático:** princípios básicos e utilização. Brasília: Ed.UnB, 1969.

TAN, J. Grounded theory in practice: issues and discussion for new qualitative researchers. **Journal of Documentation,** Bingley, v. 66, n. 1, p. 93-112, 2010.

UNIVERSIDADE ESTADUAL PAULISTA. Anuário estatístico 2013. São Paulo, 2013. 311 p.

VILLÉN-RUEDA, L. et al. The use of OPAC in a large academic library: a transactional log analysis study of subject searching. **The Journal of Academic Librarianship,** v. 33, n. 3, p. 327-337, 2007.

REDES SOCIAIS EM BIBLIOTECAS: TRABALHOS APRESENTADOS NO XVII SNBU

SOCIAL NETWORKS IN LIBRARIES: PAPERS PRESENTED AT THE XVII SNBU

Maira Nani França

Angela Maria Grossi de Carvalho

Resumo: As redes sociais têm um papel importante na sociedade contemporânea. Além da inter-relação e interatividade proposta por essas mídias sociais, é importante observar como elas podem ser exploradas pelos diversos atores sociais, tais como as bibliotecas. Para atender as demandas dos usuários, as bibliotecas universitárias passaram a perceber a crescente necessidade de utilizar recursos tecnológicos, incluindo nesse contexto o uso das redes sociais. Se, por um lado, a tecnologia auxilia em processos mais dinâmicos, interativos e participativos, por outro, revela a fragilidade de muitas bibliotecas universitárias e cria resistências, mau uso e falta de habilidade na operação dos recursos. Este artigo analisa os trabalhos publicados nos anais eletrônicos do XVII Seminário Nacional de Bibliotecas Universitárias em sua última edição, em 2012, visando identificar como as bibliotecas universitárias têm utilizado as redes sociais como canal de comunicação direta com seus usuários a fim de contribuir para melhoria da qualidade das atividades relacionadas ao ensino, à pesquisa e à extensão. Na oportunidade foram utilizados como categorias de análise os autores mais citados/temática central; a definição de redes sociais e suas características; os tipos de redes sociais utilizados pelas bibliotecas; a importância do uso dessas plataformas nas bibliotecas universitárias; e a competência informacional do bibliotecário nesse novo cenário tecnológico. Espera-se que este trabalho possa contribuir com as discussões sobre o uso das redes sociais pelas bibliotecas e fornecer subsídios para a capacitação dos profissionais da informação de modo a que possam atender as demandas dessa nova configuração da Sociedade da Informação.

Palavras-chave: Rede social. Biblioteca universitária. Tecnologias da Informação e da Comunicação.

Abstract: Social networks have an important role in contemporary society. Besides the interrelatedness and interactivity proposed by these social media, it is important to note how they can be exploited by various social actors, such as libraries. To meet the demands of users, university libraries began to realize the growing need of using technological resources, including the use of social networks. If on one hand the technology helps in making a more dynamic, interactive and participative processes, on the other hand, it reveals the fragility of many university libraries and generates resistance, misuse and lack of skill in the operation of the resources. This article reviews the papers published in the electronic proceedings of the last edition of the XVII National Seminar of University Libraries, in 2012, to identify how academic libraries have been using social networks as a direct communication channel with their users to improve the quality of teaching, research and extension related activities. On that occasion, the analyzed categories were: the most cited authors/central theme; the definition of social networks and their characteristics; types of social networks used by libraries; the importance of using these platforms by the university libraries and librarian literacy in this new technological landscape. We hope that this work will contribute to discussions on the use of social networks by libraries and encourage the training of information professionals so they can meet the demands of this new configuration of the Information Society.

Keywords: Social network. University library. Information and Communication Technologies

1 INTRODUÇÃO

As mudanças ocorridas com a introdução da internet na chamada Sociedade da Informação transformaram não só o modo de consumir, como também de produzir e processar a informação. Com a ampliação do acesso e da velocidade da conexão, a mobilidade e a convergência, a evolução das tecnologias de informação e comunicação (TIC) foi responsável por impulsionar, a partir da década de 1960, o acesso ao conhecimento. “A mediação digital remodela certas atividades cognitivas fundamentais que envolvem a linguagem, a sensibilidade, o conhecimento e a imaginação inventiva” (LÉVY, 1998, p. 17).

A partir de então emerge um novo tipo de sociedade: a Sociedade em Rede. Apesar de a estrutura social ser a responsável por formar a tecnologia, em função das suas necessidades, valores e/ou interesses, esta “é condição necessária, mas não suficiente para a emergência de uma nova forma de organização social baseada em redes, ou seja, na difusão de redes em todos os aspectos da actividade na base das redes de comunicação digital” (CASTELLS, 2005, p. 17).

A concepção de sociedade em rede surge num momento de expansão social, na exploração de novos territórios, na utilização de bens e serviços, levando em consideração um novo paradigma desenhado com a evolução da sociedade industrial para a pós-industrial. No entanto, a organização social em forma de rede não é recente e traz em seu bojo benefícios e malefícios. Se, por um lado, estimula a organização flexível e adaptável, por outro lado, em muitos momentos não foi capaz de ampliar e organizar recursos necessários de modo a abranger a dimensão e a complexidade de organização necessária para tal feito. Com a possibilidade de aproximação cada vez mais presente, o uso das redes se tornou um meio importante para o desenvolvimento social. Podendo ser vista como uma “coluna vertebral da sociedade em rede”, as redes de comunicação digital eram as infraestruturas necessárias para que a sociedade industrial fosse construída, transcendendo fronteiras, materializando a aldeia global, ou redes globais de capital, bens, serviços, comunicação, informação, ciência e tecnologia (CASTELLS, 2005, p. 18).

Com base nessa concepção, podemos compreender que a sociedade em rede auxilia no processamento e na distribuição de informação por meio dos ativos acumulados nos nós dessas redes. Além disso, a concepção de sociedade em rede faz com que se manifeste a transformação da sociabilidade, por um lado, com a desaceleração da interação

face a face e o crescimento no isolamento das pessoas diante dos computadores, e, por outro lado, com o uso social e político das ferramentas digitais, que faz com que a sociedade em rede seja vista como uma sociedade hipersocial, em que pessoas integram tecnologias às suas vidas, em atividades cotidianas, transformando a realidade virtual em virtualidade real, que impacta diretamente na sociabilidade. Mudança essa que não pode ser exclusivamente consequência da Internet, mas que de acordo com Castells (2005, p. 23) é suportada pela própria lógica das redes de comunicação, sendo a emergência do individualismo em rede, já que “as novas tecnologias de comunicação adaptam-se perfeitamente na forma de construir sociabilidades em redes de comunicação auto-selectivas, ligadas ou desligadas dependendo das necessidades ou disposições de cada indivíduo”.

Nessa perspectiva, com a melhoria dos dispositivos e a evolução da Web (1.0, 2.0 e assim sucessivamente), temos a segunda geração (2.0), em que as redes sociais passam a ser abordadas e vistas como parte integrante do ambiente tecnológico.

Em 2004, durante uma sessão de *brainstorming* na *Web 2.0 Conference*, Tim O'Reilly apresentou um conceito que revolucionaria a internet. Apresentou a Web 2.0 como a mudança da internet como plataforma e descreveu suas regras de sucesso, entre as quais se destaca como a mais importante o desenvolvimento de aplicativos que aproveitem os efeitos das redes com base na inteligência coletiva para se tornarem melhores do que são (O'REILLY, 2006). Pode também ser entendida como uma plataforma dinâmica, interativa, que possibilita atualização constante de *softwares*, tendo como competência-chave a arquitetura da participação, com a proposta de reorganização dos dados de múltiplas fontes (CATARINO; CARVALHO; ZANINELLI, 2012; SOUZA; SANTOS; SILVA, 2012). A tecnologia Web 2.0 está diretamente relacionada ao princípio e à evolução das redes sociais, objeto deste estudo.

Redes como Orkut, Facebook, Twitter, blogs, entre outros que se utilizam da interatividade como instrumento principal, emergem. As redes sociais podem ser definidas como

[...] un conjunto bien delimitado de actores-individuos, grupos, organizaciones, comunidades, sociedades globales, etc., vinculados unos a otros a través de una relación o un conjunto de relaciones sociales. [...] El rasgo más característico de las redes sociales consiste en que requieren conceptos, definiciones y procesos en que las unidades sociales aparecen vinculadas unas a otras a través de diversas relaciones (S. Wasserman y K. Faust, 1994:6) (LOZARES, 1996, p. 108).

Dois elementos são essenciais para uma rede social: os *atores* (pessoas, instituições ou grupos; os nós da rede) e suas *conexões* (interações ou laços sociais). “Uma rede, assim, é uma metáfora para observar os padrões de conexão de um grupo social, a partir das

conexões estabelecidas entre os diversos atores” (RECUERO, 2009, p. 24). O foco está na estrutura social, sendo importante salientar que não há um isolamento entre os atores e suas conexões. Ou seja, o princípio é a interatividade, a inter-relação que se estabelece com as plataformas digitais.

Vistas como “metáforas estruturais”, as redes são constituídas em forma de agrupamentos sociais, possuindo topologias, estruturas. “Essas topologias são relacionadas às estruturas das redes sociais, ou seja, à estrutura construída através dos laços sociais estabelecidos pelos atores” (RECUERO, 2009, p. 56).

As redes sociais são dinâmicas e estão em constante transformação, influenciadas pelas interações estabelecidas entre os atores e suas conexões. De acordo com Kaufman (2010, p. 51), as principais funções de uma rede social são: “1) Conectar os indivíduos todo o tempo e em todo lugar; 2) Disponibilizar conteúdo multiplataforma; 3) Compartilhar informações, decisões, conteúdos, etc. 4) Customizar tudo que o usuário quiser”.

Dados recentes do Centro de Estudos sobre as Tecnologias da Informação e da Comunicação (CETIC), TIC Domicílios e Usuários 2013⁹³, divulgada em 26 de junho de 2014, mostram como as redes sociais são utilizadas no país. Com 85,9 milhões de usuários com acesso à internet, a maioria deles, 77%, utiliza redes sociais, como o *Facebook*, o *Orkut* ou o *Google +*, como a principal atividade na internet. Portanto, o uso de redes sociais é imperativo na internet, pelo menos é o que mostram as pesquisas nacionais, liderando e deixando para trás atividades como jogos, música e comércio eletrônico.

A comunicação em rede tem sido explorada como instrumento de ativação de movimentos sociais e culturais como a luta dos direitos humanos, feministas, ambientalistas, etc. [...] Através dessa complexidade de funções, percebe-se que as redes sociais virtuais são canais de grande fluxo na circulação de informação, vínculos, valores e discursos sociais, que vêm ampliando, delimitando e mesclando territórios. Entre desconfiados e entusiásticos, o fato é que as redes sociais virtuais são convites para se repensar as relações em tempos pós-modernos (AFONSO, 2009, p. 29).

Inserida no ambiente universitário, a função da biblioteca é promover o acesso à informação por meio da estrutura física e de acervo impresso e virtual e da disponibilização de serviços que acompanhem as tendências tecnológicas, contribuindo diretamente com as atividades de ensino, pesquisa, extensão e inovação. Uma rede social é um canal de comunicação que possibilita à biblioteca universitária disponibilizar informação independente do lugar em que seu usuário esteja (SANTOS *et al.*, 2012).

⁹³ Disponível em: <http://cetic.br/usuarios/tic/2013/C5.html>.

Buscando acompanhar as evoluções das TIC e, consequentemente, do perfil dos usuários – público jovem, conectado à internet 24 horas por dia –, os gestores das bibliotecas têm percebido a necessidade de adaptar os produtos e serviços oferecidos, propondo novas atividades a seus usuários.

Este trabalho propõe-se a identificar, por meio da revisão de literatura, como as bibliotecas universitárias brasileiras têm utilizado as redes sociais, bem como os fundamentos e princípios que norteiam esse tema, colaborando para que outras bibliotecas possam usar essas ferramentas de modo consciente, satisfazendo as demandas informacionais e comunicacionais de seus usuários.

2 METODOLOGIA

A revisão de literatura deste trabalho ocorreu por intermédio de pesquisa bibliográfica nos anais eletrônicos do XVII Seminário Nacional de Bibliotecas Universitárias (SNBU)⁹⁴.

O SNBU é um fórum de discussão e debate nacional entre profissionais da informação a respeito de questões referentes à Biblioteconomia, subárea da Ciência da Informação⁹⁵, com ênfase nas bibliotecas, criado em 1978, cujas edições acontecem a cada dois anos⁹⁶.

Ao pesquisar no mecanismo de busca dos anais eletrônicos do evento⁹⁷ com o descritor *redes sociais*, objeto deste estudo, foi possível identificar sete trabalhos⁹⁸ a respeito do tema em questão. Como categoria de análise foram definidos: embasamento teórico (temática/autores mais citados), definição de redes sociais e suas funções, tipos de

⁹⁴ Realizado de 16 a 21 de setembro de 2012, em Gramado, RS, com o tema: “A biblioteca universitária como laboratório na Sociedade da Informação. Esse evento foi dividido em 4 eixos temáticos: I. Construção e comunicação da informação; II. Organização, preservação e acesso à informação; III. Recuperação, disseminação e uso da informação; IV. Gestão de biblioteca universitária, sendo o subtema *Divulgação de produtos e serviços (páginas, blogs e redes sociais)*, objeto de estudo, subordinado ao Eixo III.

⁹⁵ Cf. Tabela de áreas do conhecimento do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) disponível em <http://www.cnpq.br/documents/10157/186158/TabeladeAreasdonoConhecimento.pdf>.

⁹⁶ Anos pares. Neste ano, o XVIII SNBU, com o tema: “Bibliotecas Universitárias e o acesso público à informação: articulando leis, tecnologias, práticas e gestão”, acontecerá em Belo Horizonte, de 16 a 21 de novembro.

⁹⁷ Anais disponíveis em: <http://portal.febab.org.br/anais/issue/view/4>.

⁹⁸ Bezerra *et al.* (2012), Catarino, Carvalho e Zaninelli (2012), Nascimento *et al.* (2012), Santos *et al.* (2012), Shirayama e Marquiori (2012), Souza; Santos e Silva (2012) e, Teixeira e Souza (2012).

redes mais utilizadas pelas bibliotecas, a forma como estão sendo utilizados esses recursos e o perfil do profissional da informação nessa era da informação.

3 RESULTADOS

Nesta seção serão apresentados os autores mais citados por cada pesquisador como referencial teórico para embasar sua pesquisa, com a respectiva temática central; a definição de redes e suas características; os tipos de redes sociais utilizados pelas bibliotecas; a importância do uso dessas plataformas nas bibliotecas universitárias brasileiras; e a competência da informação do bibliotecário nesse novo cenário tecnológico.

O QUADRO 1 apresenta o objetivo de cada trabalho analisado e a temática que embasa cada estudo com os respectivos autores que discorrem sobre o tema.

Dentre os conceitos apresentados no material analisado, foram identificados: Redes sociais (6); Redes (4); Redes sociais em bibliotecas (3); Fontes de informação e Web 2.0 (2); Perfil profissional (2); Inteligência competitiva; Estratégia organizacional; TIC/WEB (evolução); Biblioteca 2.0; Ciência da Informação; Clipping; Rede do conhecimento; Biblioteca universitária (1) (Quadro 1).

QUADRO 1 – Apresentação dos trabalhos analisados

TRABALHO	OBJETIVO	TEMÁTICA	AUTORES
1) Bezerra <i>et al.</i> (2012)	abordar o tema das redes sociais como fonte de informação, redes essas que têm como finalidade melhorias na prestação de serviços destinados aos usuários reais e potenciais	Redes	Castells (1999), Watts (2004) apud Rodrigues e Tomaél (2008) e, Tomaél (2008)
		Redes sociais	Gonçalves (2008) e, Silva e Bacalgnini (2009)
2) Nascimento <i>et al.</i> (2012)	mostrar como as redes sociais podem contribuir para a geração de inteligência competitiva nas organizações	Inteligência competitiva Fontes de informação Estratégia organizacional Redes sociais	Chiavenato e Sapiro (2003) Balestrin (2011) Molina e Aguilar (2004) Martinho (2003) e, Alcará <i>et al.</i> (2006)
3) Catarino, Carvalho e Zaninelli (2012)	publicar orientações para a manutenção de conteúdos das Redes Sociais	TIC/WEB (evolução) Web 2.0 Biblioteca 2.0 Redes Redes sociais Redes sociais em bibliotecas	Ramos (1996) e, Watanabe, Paletta e Yamashita (2010) O'Reilly (2005) Maness (2007), Conti e Pinto (2010) e, Gonçalves, Conceição e Luchetti (2010) Castells (2009) Tomaél; Marteleto (2006), Giustini e Wright (2009), Recuero (2009), Aguiar e Silva (2010), Brito (2010), Campos (2012) e, Wikipédia (2012) Gomes, Prudêncio e Santos (2010) e, Pena; Piñol (2010)
4) Souza, Santos e Silva (2012)	descrever a evolução de uma atividade de disseminação da informação – o clipping, desenvolvida pela biblioteca universitária CTC/F da UERJ, desde o seu estágio impresso até a utilização das ferramentas <i>Blogger</i> e <i>Facebook</i> , disponíveis na web	Ciência da Informação Web 2.0 Perfil profissional Clipping Redes sociais Redes sociais em bibliotecas	Bush (1995) apud Saracevi (1996) Blattmann e Silva (2007) Rodrigues (2009) Ferreira (1986), Pftzenreuter (2010)e, ABEMO (2012) Barros (2007), Eiras (2007) e, Recuero apud Aguiar e Silva (2010) Maness apud Aguiar e Silva (2010)

	propor um estudo para implantação de redes sociais que veiculem informações e estabeleçam um canal efetivo de comunicação com o usuário, criando vínculos e divulgando os serviços oferecidos pela biblioteca	Redes sociais em bibliotecas	Aguiar e Silva (2010)
5) Shirayama e Marquiori (2012)		Redes	Alba (1982), Castells (1999), Minarelli (2001), Latour (2007), e Schmitt (2011)
6) Teixeira e Souza (2012)	estudar as redes de conhecimento e os espaços de interação entre os diversos segmentos da sociedade no âmbito das ciências e seu relacionamento com as fontes de informação no compartilhamento do conhecimento	Redes sociais	Martelete (2001), Fontes e Eichner (2004), Freeman (2004), Velázquez e Aguilar (2005), Lozares (2007) e, Meira (2009)
		Fontes de informação	Martin Veja (1995), Jyang Yl (2000), Bernardo, Nobre e Jatene (2004) e, Silva e Castro (2008)
		Rede do conhecimento	Tomaél (2008) e, Johnson (2011)
		Perfil profissional	Delors (2003)
		Biblioteca universitária	Cunha (2010)
7) Santos <i>et al.</i> (2012)	relatar a experiência de uma biblioteca universitária com redes, abordando habilidades desenvolvidas, diretrizes de comunicação e política de governança como estratégia para a edição descentralizada dos conteúdos. Apresentar o desenvolvimento do Twitter, do Facebook, do YouTube, do Flickr, do Blog e seu processo de revisão	Redes	Martelete (2001) e, Tomaél (2005)
		Redes sociais	Recuero (2005, 2007), Tripathi e Kumar (2007), Margaix-Arnal (2008), Viberti (2009), Dickson e Holley (2010) e, Dourado (2010)

Fonte: As autoras.

Com base no QUADRO 2 é possível identificar os autores que discutem cada conceito relacionado ao objeto deste estudo.

QUADRO 2 – Referencial teórico (autor) por conceito

CONCEITOS	AUTORES
Redes sociais	Marteleto (2001), Martinho (2003), Fontes e Eichner (2004), Freeman (2004), Recuero (2005, 2007, 2009), Velázquez e Aguilar (2005), Alcará <i>et al.</i> (2006), Tomaél e Marteleteo (2006), Barros (2007), Eiras (2007), Lozares (2007), Tripathi e Kumar (2007), Gonçalves (2008), Margaix-Arnal (2008), Giustini e Wright (2009), Meira (2009), Silva e Bacalgini (2009), Viberti (2009), Aguiar e Silva (2010), Brito (2010), Dickson e Holley (2010), Dourado (2010), Recuero apud Aguiar e Silva (2010), Campos (2012) e, Wikipédia (2012)
Redes	Alba (1982), Castells (1999, 2009), Marteleteo (2001), Minarelli (2001), Watts (2004) apud Rodrigues e Tomaél (2008), Tomaél (2005, 2008), Latour (2007), e, Schmitt (2011)
Redes sociais em bibliotecas	Aguiar e Silva (2010), Gomes, Prudêncio e Santos (2010), Maness apud Aguiar e Silva (2010), Pena; Piñol (2010)
Fontes de informação	Martin Veja (1995), Jyang YI (2000), Bernardo, Nobre e Jatene (2004), Silva e Castro (2008) e, Balestrin (2011)
Web 2.0	O'Reilly (2005) e, Blattmann e Silva (2007)
Inteligência competitiva	Chiavenato e Sapiro (2003)
Estratégia organizacional	Molina e Aguilar (2004)
TIC/WEB (evolução)	Ramos (1996) e, Watanabe, Paletta e Yamashita (2010)
Biblioteca 2.0	Maness (2007), Conti e Pinto (2010) e, Gonçalves, Conceição e Luchetti (2010)
Ciência da Informação	Bush (1995) apud Saracevi (1996)
Perfil profissional	Rodrigues (2009)
Clipping	Ferreira (1986), Pftzenreuter (2010) e, ABEMO (2012)
Rede do conhecimento	Johnson (2011) e Tomaél (2008)
Perfil profissional	Delors (2003)
Biblioteca universitária	Cunha (2010)

Fonte: As autoras.

Os autores mais citados para a temática **Redes Sociais** foram Recuero, Wikipedia, Aguiar e Silva, e Dickson e Holley; para **Redes**, Castells e Tomaél; e para **Redes sociais em bibliotecas**, Aguiar e Silva (2010). Os autores das demais temáticas não foram destacados, considerando que não foi identificada mais de uma citação do mesmo autor para elas.

As redes sociais são plataformas que facilitam a conectividade (SOUZA; SANTOS; SILVA, 2012), a interatividade e o compartilhamento de informação, conhecimento e experiências em torno de interesses comuns (SOUZA; SANTOS; SILVA, 2012; TEIXEIRA; SOUZA, 2012) com abrangência praticamente ilimitada (CATARINO; CARVALHO; ZANINELLI, 2012).

Para Nascimento *et al.* (2012), esses recursos são ferramentas de competitividade que fortalecem a cultura informacional. Os autores destacam sua importância por ser um novo espaço de relacionamento entre pessoas (cliente x organização) que permite a recuperação e a disseminação de informações variadas (de entretenimento, científicas, laborais etc.).

Os ambientes virtuais possibilitam o trabalho cooperativo, o rápido acesso às informações (SOUZA; SANTOS; SILVA, 2012), o apoio social e a ascensão na carreira (TEIXEIRA; SOUZA, 2012).

Para Bezerra *et al.* (2012), Catarino, Carvalho e Zaninelli (2012), e Teixeira e Souza (2012), a colaboração mútua dos indivíduos que compõem determinada rede favorece a produção do conhecimento tácito – crescimento individual –, que com as parcerias (compartilhamento) reduz a incerteza e promove benefícios recíprocos por intermédio da construção do conhecimento coletivo, que conduz a um crescimento mútuo.

Catarino, Carvalho e Zaninelli (2012), Santos *et al.* (2012) e Souza, Santos e Silva (2012) compartilham suas experiências quanto à criação, ao uso, à manutenção e ao gerenciamento das seguintes redes sociais em suas bibliotecas: Facebook (mensagens, links, vídeos, fotos etc.), Twitter (notícias curtas e rápidas), Flickr (álbums de fotos), Youtube (vídeos), Delicious (*sites* favoritos), Slideshare (apresentações de slides, *e-books* etc.) e Blog (registro cronológico de informações).

Para Santos *et al.* (2012) e Souza, Santos e Silva (2012), dentre todas as ferramentas acima citadas, o blog é a que oferece maior facilidade de criação e administração, por possuir uma interface amigável, popular e dinâmica que permite a produção de conteúdo de maneira simples, sem exigir conhecimento técnico aprofundado para seu gerenciamento.

Souza, Santos e Silva (2012) apresentam também sua experiência de clipping digital baseado no webjornalismo, que com a explosão bibliográfica passou a ser uma atividade essencial. Esse recurso funciona como disseminação seletiva da informação (DSI)⁹⁹, oferecido por muitas bibliotecas com o nome de *Rich Site Sumary* (RSS), que permite maior visibilidade da informação e colabora na formação de opinião dos usuários.

Para que uma biblioteca exista, é essencial que haja usuários com demandas informacionais para o desenvolvimento de suas atividades. O mesmo acontece nas redes sociais, que dependem de pessoas interconectadas para o seu funcionamento. Bezerra *et al.*

⁹⁹ Serviço personalizado, também conhecido como serviço de alerta, que informa e atualiza o usuário quando um novo documento de seu interesse fica disponível no sistema, ou seja, que permite ao pesquisador o monitoramento dos temas que estão sendo publicados em sua área de interesse.

(2012), Catarino, Carvalho e Zaninelli (2012), e Souza, Santos e Silva (2012) destacam que o objetivo principal das redes sociais em bibliotecas é a satisfação dos usuários reais e potenciais.

As mídias sociais possibilitam à biblioteca traçar um perfil de interesse, comportamentos e tendências (SOUZA; SANTOS; SILVA, 2012); identificar novas demandas (BEZERRA *et al.*, 2012; NASCIMENTO *et al.* 2012; SOUZA; SANTOS; SILVA, 2012); mapear o comportamento do usuário (horário de acesso, o que acessa, *posts* curtidos), permitindo reforçar os conteúdos de maior interesse (SANTOS *et al.*, 2012), ou seja, é uma excelente ferramenta para a realização de estudo de usuários (SHIRAYAMA; MARQUIORI, 2012).

O conceito de Biblioteca 2.0 é apresentado por Catarino, Carvalho e Zaninelli (2012) como aquela que utiliza os recursos colaborativos da Web 2.0, em que o usuário se torna agente ativo no processo de criação e troca de informação, reforçando que na Web 2.0 a ênfase deve ser dada aos usuários e não às tecnologias. Nesse ambiente colaborativo os usuários também participam do processo de construção do conteúdo, ou seja, produzem, consomem e compartilham informações e conhecimentos (SANTOS *et al.*, 2012). É uma oportunidade que aproxima a biblioteca de seus usuários, que possibilita que o profissional da informação vá ao encontro da comunidade acadêmica (CATARINO; CARVALHO; ZANINELLI, 2012; SANTOS, 2012), estabelecendo um contato direto e criando laços (relacionamento) de maneira interativa com aquele que necessita da informação (BEZERRA *et al.*, 2012; SOUZA; SANTOS; SILVA, 2012; SHIRAYAMA; MARQUIORI, 2012). Enfim, é uma ferramenta aliada na gestão da informação, mediante a disponibilização dos diversos perfis de usuários da biblioteca (SOUZA; SANTOS; SILVA, 2012).

Nas plataformas digitais é possível à biblioteca disponibilizar material informacional nos mais diversos formatos: **textos** (palestras, manuais de serviço, trabalhos apresentados em eventos, artigos, arquivos escaneados, mensagens¹⁰⁰ etc.) (SHIRAYAMA; MARQUIORI, 2012; SANTOS *et al.*, 2012; SOUZA; SANTOS; SILVA, 2012); **vídeos** (institucionais, formação e capacitação de usuário etc.); **áudios** (*podcast*¹⁰¹, músicas, palestras e entrevistas gravadas etc.) (SANTOS *et al.*, 2012; SOUZA; SANTOS; SILVA, 2012; SHIRAYAMA; MARQUIORI, 2012) e **imagens** (fotos) (CATARINO; CARVALHO; ZANINELLI, 2012), cuja função varia de atividades de entretenimento às científicas, tornando-se um espaço para

¹⁰⁰ *Posts* e comentários.

¹⁰¹ Arquivos de áudio transmitidos pela internet.

registro e resgate da memória e da identidade cultural da unidade informacional e da instituição.

No material analisado foi possível identificar uma série de produtos e serviços que as bibliotecas universitárias podem disponibilizar para seus usuários nas redes sociais, a destacar: a) notícias de interesse da comunidade acadêmica sobre a biblioteca (funcionamento, novas aquisições etc.) e a instituição (CATARINO; CARVALHO; ZANINELLI, 2012; SANTOS *et al.*, 2012; SHIRAYAMA; MARQUIORI, 2012; SOUZA; SANTOS; SILVA, 2012); b) informações de cunho administrativo (missão, histórico, dados estatísticos etc.) (CATARINO; CARVALHO; ZANINELLI, 2012; SHIRAYAMA; MARQUIORI, 2012); c) oportunidades e estágios (SOUZA; SANTOS; SILVA, 2012); d) eventos (exposições, palestras, treinamentos¹⁰², cursos, congressos) promovidos pela biblioteca e pela instituição e indicações de leituras especializadas (CATARINO; CARVALHO; ZANINELLI, 2012; SANTOS *et al.*, 2012; SHIRAYAMA; MARQUIORI, 2012); e) fontes de pesquisa (bibliotecas digitais, bases de dados e periódicos eletrônicos, repositórios institucionais, portais de pesquisa, catálogos eletrônicos de outras bibliotecas etc.), trabalhos relacionados às bibliotecas universitárias e/ou à área da Ciência da Informação) (SANTOS *et al.*, 2012; SHIRAYAMA; MARQUIORI, 2012); e f) outras redes sociais da biblioteca. Como exemplo, Souza, Santos e Silva (2012) citam o Centro de Tecnologia e Ciência (CTC/F) da Universidade do Estado do Rio de Janeiro (UERJ), que identificou aumento significativo no acesso ao seu blog após divulgação no Facebook da instituição.

As redes sociais de bibliotecas são ferramentas de informação e interação (BEZERRA *et al.*, 2012; SANTOS *et al.*, 2012; SOUZA; SANTOS; SILVA, 2012) que intensificam o processo de comunicação entre bibliotecários e usuários (CATARINO; CARVALHO; ZANINELLI, 2012; SANTOS *et al.*, 2012; SHIRAYAMA; MARQUIORI, 2012) de forma síncrona (mensagens instantâneas) e assíncrona (wiki e blog) (SOUZA; SANTOS; SILVA, 2012). Dentre suas diversas vantagens foram identificadas no material analisado: a) redução de distâncias, uma vez que possibilita acesso instantâneo à informação a qualquer hora e lugar (SHIRAYAMA; MARQUIORI, 2012; SOUZA; SANTOS; SILVA, 2012); b) democratização do uso da informação (SOUZA; SANTOS; SILVA, 2012); c) promoção de novos produtos e serviços pelo marketing (BEZERRA *et al.*, 2012; SANTOS *et al.*, 2012); d) baixo custo de

¹⁰² Disponibilização de tutoriais com procedimentos e funções da biblioteca, preferencialmente utilizando uma ferramenta interativa (web 2.0), como o Prezi, por exemplo.

investimento; e) alcance de um grande número de usuários em um curto espaço de tempo; f) enriquecimento do processo dos serviços de alerta (DSI) (SOUZA; SANTOS; SILVA, 2012); g) acesso e apropriação da informação (CATARINO; CARVALHO; ZANINELLI, 2012); h) ampliação de espaços dinâmicos e abertos para debate e discussão (CATARINO; CARVALHO; ZANINELLI, 2012; SANTOS *et al.*, 2012; SOUZA; SANTOS; SILVA, 2012); e i) promoção da instituição e da biblioteca, possibilitando maior visibilidade (BEZERRA *et al.*, 2012; SHIRAYAMA; MARQUIORI, 2012; SOUZA; SANTOS; SILVA, 2012).

Souza, Santos e Silva (2012) e Teixeira e Souza (2012) afirmam que uma rede social utilizada como recurso pedagógico é um ambiente mais atrativo e estimulante que a sala de aula, pois possibilita a discussão no interior da comunidade acadêmica a respeito dos mais diversos temas, por intermédio da criação de grupos de pesquisa, de projetos, de estudos de alunos de uma disciplina, profissionais, entre outros, com objetivo de construir, disseminar e compartilhar conhecimentos.

Catarino, Carvalho e Zaninelli (2012) e Santos *et al.* (2012) discorrem sobre a importância de padronizar a comunicação *on-line* dos conteúdos por meio da elaboração coletiva de políticas de governança, bem como de divulgar as regras de uso, a saber: respeito aos direitos de propriedade intelectual, questões éticas, comunicação responsável e divulgação de notícias verídicas. Essas diretrizes permitem a coerência das ações de manutenção dos canais de comunicação da biblioteca pela equipe designada para esse fim, visando conquistar a credibilidade e a confiança do usuário por meio da divulgação de informações relevantes e atualizadas, de modo interativo, dando oportunidade ao usuário de se sentir como parte integrante da instituição e/ou unidade informacional.

As transformações advindas dos avanços tecnológicos hoje disponíveis, sobretudo no ambiente universitário, que impactaram no pensar e agir das bibliotecas provocam uma reflexão a respeito das competências informacionais do bibliotecário. Nesse contexto, são requeridas novas competências do perfil do bibliotecário. Essas características essenciais foram identificadas no material analisado, a saber: a) estar atento às novas demandas informacionais de seus usuários (SANTOS *et al.*, 2012; SHIRAYAMA; MARQUIORI, 2012); b) promover o atendimento ao usuário de maneira rápida e efetiva, e estimular sua autonomia (SHIRAYAMA; MARQUIORI, 2012, TEIXEIRA; SOUZA, 2012); c) conhecer tendências tecnológicas e acompanhar mudanças (CATARINO; CARVALHO; ZANINELLI, 2012); d) adequar os serviços ao mundo virtual, principalmente no que se refere à disseminação da informação; e) ser constantemente curioso para buscar e testar novas

tecnologias, visando facilitar o acesso à informação, extrair seus benefícios e ampliar o seu uso (CATARINO; CARVALHO; ZANINELLI, 2012; SOUZA; SANTOS; SILVA, 2012; TEIXEIRA; SOUZA, 2012); f) conhecer e utilizar os recursos da Web 2.0, a fim de fornecer respostas mais específicas às demandas informacionais dos usuários; g) estar preparado para atuar tanto em ambientes físicos quanto em ambientes virtuais (SOUZA; SANTOS; SILVA, 2012); e h) estimular o rigor intelectual (TEIXEIRA; SOUZA, 2012), ou seja, participar de projetos de capacitação continuada (leituras, participação em eventos, cursos, apresentação de trabalhos etc.) referentes às redes sociais, isto é, preparar-se para dialogar com a nova geração do conhecimento, da informação e da tecnologia – os usuários reais das bibliotecas universitárias de hoje. Para Souza, Santos e Silva (2012), prazer e curiosidade são os grandes aliados dos bibliotecários nesta nova era tecnológica informacional.

4 CONSIDERAÇÕES PARCIAIS

A biblioteca universitária é um espaço muito dinâmico, em constante adequação dos serviços e das formas de comunicação com seus usuários, e tem entre seus grandes desafios acompanhar o desenvolvimento tecnológico. Entretanto, ainda é pequeno o número de bibliotecas que utiliza as tecnologias e disponibiliza em seus *sites* links de suas redes sociais (CATARINO; CARVALHO; ZANINELLI, 2012; SOUZA, SANTOS; SILVA, 2012).

O *site* de uma biblioteca é um canal de comunicação direta com a comunidade acadêmica que permite maior visibilidade desta, bem como a prestação de serviços, além de receber elogios, sugestões, críticas, comentários, questionamentos e dúvidas dos usuários. Catarino, Carvalho e Zaninelli (2012) alertam quanto à importância de que as redes sociais sejam divulgadas na página da biblioteca como uma forma de promover e medir a qualidade dos serviços prestados (NASCIMENTO *et al.*, 2012) e complementam que, mesmo que a instituição na qual a biblioteca esteja inserida disponibilize redes sociais, é essencial que a unidade informacional tenha seu próprio espaço de interação com a comunidade acadêmica, mais específico, ou seja, menos generalista.

Por serem um recurso dinâmico, é fundamental que as redes sociais da biblioteca sejam constantemente atualizadas. “O trabalho de atualização nas redes sociais exige manutenção, cuidado, profissionalismo e continuidade” (SANTOS *et al.*, 2012, p. 5). Souza, Santos e Silva (2012) destacam ainda a importância da constante avaliação do uso dessas ferramentas, de modo que elas sejam usuais e constantes.

Nos trabalhos analisados foram identificados dois produtos que podem ser úteis para os bibliotecários que pretendem implantar e gerenciar as redes sociais de suas unidades

informacionais: *A política para manutenção de conteúdos nas redes sociais e no portal do Sistema de Bibliotecas da UEL* (CATARINO; CARVALHO; ZANINELLI, 2012) e *Política de governança da Biblioteca Central da Faculdade de Medicina da USP* (SANTOS *et al.*, 2012, p. 5)¹⁰³.

Para Teixeira e Souza (2012), a biblioteca está perdendo a sua supremacia enquanto principal fonte de busca em razão do impacto da tecnologia digital. Para conhecer as tendências tecnológicas e acompanhar as mudanças, os bibliotecários precisam buscar informações relevantes por meio de atualização constante. Para Campello (2003), a biblioteca amplia seu papel de atuação na universidade por meio da competência informacional dos profissionais da informação.

Souza, Santos e Silva (2012, p. 10) sinalizam como desafio aos bibliotecários a necessidade de conhecer e utilizar a Web 2.0 a fim de aplicar os conceitos inteligentes na “organização dos conteúdos, combinação dos dados com o propósito de dar respostas mais definidas às questões de pesquisas” por meio da Web semântica. Essas habilidades específicas ligadas ao uso da informação eletrônica tornam o profissional indispensável para a instituição nas atividades de gestão dos serviços informacionais.

Espera-se que este trabalho possa contribuir com as discussões sobre o uso das redes sociais pelas bibliotecas universitárias e fornecer subsídios para a capacitação dos profissionais da informação, de modo a que possam atender às demandas dessa nova configuração de mundo interligado pelo ambiente virtual.

REFERÊNCIAS

- AFONSO, Alexandre Soares. **Uma análise da utilização das redes sociais em ambientes corporativos**. 2009. 163f. Dissertação (Mestrado em Tecnologias da Inteligência e Design Digital) – Pontifícia Universidade Católica, São Paulo, 2009.
- BEZERRA, Midinai Gomes *et al.* Redes sociais como fonte de informação: uma proposta atual. In: SEMINÁRIO NACIONAL DE BIBLIOTECAS UNIVERSITÁRIAS, 17., 2012, Gramado. **Anais...** Porto Alegre: UFRGS, 2012. 4QAR. Disponível em: <<http://www.snbu2012.com.br/anais/pdf/4QAR.pdf>>. Acesso em: 1 jul. 2014.
- CAMPELLO, Bernadete. O movimento da competência informacional: uma perspectiva para o letramento informacional. **Ciência da Informação**, Brasília, DF, v. 32, n. 3, 2003.
- CASTELLS, Manuel. A Sociedade em rede: do conhecimento à política. In: CASTELLS, Manuel; CARDOSO, Gustavo (Org.). **A sociedade em rede: do conhecimento à ação política**. [Belém]: Imprensa Nacional, Casa da Moeda, 2005. p. 17-30.

¹⁰³ Em 2009 foi instituído o Grupo de Inovação para a aplicação das ferramentas Web 2.0.

CATARINO, Maria Elisabete; CARVALHO, Márcia Marques da Silva; ZANINELLI, Neide Maria Jardinette. Política para a manutenção de conteúdos nas redes sociais e no portal do sistema de bibliotecas da UEL. In: SEMINÁRIO NACIONAL DE BIBLIOTECAS UNIVERSITÁRIAS, 17., 2012, Gramado. *Anais...* Porto Alegre: UFRGS, 2012. 4QXW. Disponível em: <<http://www.snbu2012.com.br/anais/pdf/4QXW.pdf>>. Acesso em: 5 jul. 2014.

KAUFMAN, Dora. **O processo de tomada de decisão no ciberespaço:** o papel das redes sociais no jogo das escolhas individuais. 2010. 139 f. Dissertação (Mestrado em Comunicação e Semiótica, Signo e Significação nas Mídias) – Pontifícia Universidade Católica, São Paulo, 2010.

LÉVY, Pierre. **A máquina universo.** Porto Alegre: ArtMed, 1998.

LOZARES, Carlos. La teoría de las redes sociales. **Papers**, [s.n.], v. 48, p. 103-126, 1996.

NASCIMENTO, Bruna Laís Campos do *et al.* Processo de geração de inteligência competitiva: como as redes sociais podem contribuir? In: SEMINÁRIO NACIONAL DE BIBLIOTECAS UNIVERSITÁRIAS, 17., 2012, Gramado. *Anais...* Porto Alegre: UFRGS, 2012. 4QWF. Disponível em: <<http://www.snbu2012.com.br/anais/pdf/4QWF.pdf>>. Acesso em: 3 jul. 2014.

O'REILLY, Tim. Introduction. In: O'REILLY MEDIA INC. **Web 2.0: Principles and Best Practices.** [S.l], Fall 2006.

RECUERO, Raquel. **Redes sociais na internet.** Porto Alegre: Sulina, 2009. (Coleção Cibercultura).

SANTOS, Cibele Araujo Camargo Marques dos *et al.* As redes sociais como ferramenta de marketing e comunicação com o usuário. In: SEMINÁRIO NACIONAL DE BIBLIOTECAS UNIVERSITÁRIAS, 17., 2012, Gramado. *Anais...* Porto Alegre: UFRGS, 2012. 4QVW. Disponível em: <<http://www.snbu2012.com.br/anais/pdf/4QVW.pdf>>. Acesso em: 13 jul. 2014.

SHIRAYAMA, Cristiane de Melo; MARQUIORI, Mariana. Proposta de implantação de redes sociais na biblioteca da escola de Filosofia, Letras e Ciências Humanas da Universidade Federal de São Paulo. In: SEMINÁRIO NACIONAL DE BIBLIOTECAS UNIVERSITÁRIAS, 17., 2012, Gramado. *Anais...* Porto Alegre: UFRGS, 2012. 4REX. Disponível em: <<http://www.snbu2012.com.br/anais/busca/>>. Acesso em: 9 jul. 2014.

SOUZA, Ester Aparecida Lima de; SANTOS, Marcos Antonio Vianna dos; SILVA, Zelia Maria Pereira da. O uso das ferramentas da Web 2.0 na disseminação da informação: do clipping às redes sociais. In: SEMINÁRIO NACIONAL DE BIBLIOTECAS UNIVERSITÁRIAS, 17., 2012, Gramado. *Anais...* Porto Alegre: UFRGS, 2012. 4QXZ. Disponível em: <<http://www.snbu2012.com.br/anais/pdf/4QXZ.pdf>>. Acesso em: 7 jul. 2014.

TEIXEIRA, Maria do Rocio Fontoura; SOUZA, Diogo Onofre Gomes de. Fontes de informação em Ciências: um estudo de uso a partir da metodologia de Análise de Redes Sociais. In: SEMINÁRIO NACIONAL DE BIBLIOTECAS UNIVERSITÁRIAS, 17., 2012, Gramado. *Anais...* Porto Alegre: UFRGS, 2012. 4RFV. Disponível em: <<http://www.snbu2012.com.br/anais/pdf/4RFV.pdf>>. Acesso em: 11 jul. 2014.

METADADOS ARQUIVÍSTICOS: CONSIDERAÇÕES SOBRE CONCEITOS, TIPOS E INSTRUMENTOS¹⁰⁴

ARCHIVAL METADATA: CONSIDERATIONS ABOUT CONCEPTS, TYPES AND INSTRUMENTS

Rachel Cristina Vesu Alves

Plácida Leopoldina Ventura Amorim da Costa Santos

Resumo: Os metadados são elementos fundamentais a serem determinados no domínio arquivístico para garantir a identificação, representação, acesso e preservação dos documentos. Deste modo, o problema que norteou este trabalho consiste em: como estabelecer os metadados arquivísticos? A hipótese é de que os metadados arquivísticos são estabelecidos com base nas teorias, nos princípios e nas metodologias arquivísticas para o tratamento documental, em especial, nas entidades do domínio arquivístico identificadas nos modelos conceituais do domínio e nos processos de gestão arquivística de documentos. O objetivo consiste em abordar os principais aspectos sobre os metadados arquivísticos: seu conceito no domínio, seus principais tipos e os instrumentos para seu estabelecimento padronizado. A metodologia utilizada na pesquisa consistiu em uma combinação da abordagem exploratória e descritiva com pesquisa bibliográfica realizada com a literatura disponível sobre o tema. Como resultados destaca-se o mapeamento dos principais tipos de metadados arquivísticos analisados pelo esquema de metadados da norma ISO 23081-1:2006 e pelo esquema de metadados do Modelo de Requisitos e-ARQ Brasil. Destaca-se também os instrumentos necessários para o estabelecimento padronizado de metadados no domínio arquivístico. Conclui-se que os metadados arquivísticos são estabelecidos com base nas entidades, atributos e relações do domínio arquivístico.

Palavras-chave: Domínio arquivístico. Metadados arquivísticos. Tipos de metadados arquivísticos.

Abstract: The metadata are essential elements to be determined in archival domain to ensure identification, representation, access, and preservation of documents. Therefore, the problem that guided this work is: how to establish archival metadata? The hypothesis is that the archival metadata are established based on the theories, principles and archival methodologies for documentary treatment, especially entities in the archival domain, identified in the conceptual models of the domain and processes of records management. The goal is to approach the main aspects on archival metadata: concept in domain, types and instruments for your standardized establishment. The methodology used in research is combination of exploratory and descriptive approach with bibliographic survey for analyze the literature of subject. The results showed the mapping of main types of archival metadata analyzed by metadata schema of standard ISO 23081-1:2006 and metadata schema of Requirements Model e-ARQ Brasil. Also noteworthy is necessary instruments for the establishment of standardized metadata in the archival domain. We conclude that archival metadata are established based on the entities, attributes and relationships of the archival domain.

¹⁰⁴ Parte dos resultados da pesquisa “Mapeamento de metadados no domínio arquivístico: aspectos conceituais para implementação de um esquema de metadados” realizada no Programa de Pós-Doutorado da Universidade Estadual Paulista “Júlio de Mesquita Filho” (UNESP), Faculdade de Filosofia e Ciências, campus de Marília, Departamento de Ciência da Informação. Pesquisa financiada pelo CNPq.

Keywords: Archival domain. Archival metadata. Types of archival metadata.

1 INTRODUÇÃO

Tornar os recursos acessíveis e disponíveis nos ambientes informacionais digitais tornou-se um desafio para as áreas envolvidas com a representação da informação. Assim, as áreas de Ciência da Informação, Biblioteconomia, Arquivologia e Museologia passaram a vivenciar significativas mudanças em relação à redefinição das atividades de produção, armazenamento, tratamento, busca, localização, acesso, recuperação e uso das informações, com o intuito de promover um melhor acesso, busca e recuperação das informações na Web.

Neste contexto, os metadados são considerados elementos essenciais para promover a representação, o acesso, a recuperação e a preservação dos recursos nesses ambientes. Contudo, é impossível estabelecer uma unicidade na construção de metadados para serem utilizados por todas as áreas que estabelecem interdisciplinaridade com a Ciência da Informação, devido às características próprias de cada domínio¹⁰⁵. Tampouco, ser utilizado metadados de esquemas amplos e gerais, pois esses não refletem a especificidade necessária em determinados domínios.

Assim, o estabelecimento de metadados ocorre em domínios específicos que estabelecem interdisciplinaridade com a Ciência da Informação, como por exemplo, o domínio de bibliotecas, arquivos, museus e outras comunidades de informação que compõem o denominado Universo Bibliográfico (IFLA, 2009). E se configura como um dos desafios contemporâneos a ser determinado nesses domínios, pois proporcionam o acesso, a busca, a recuperação e a preservação das informações com maior eficiência e especificidade nos ambientes digitais.

De acordo com a IFLA (2009) o domínio arquivístico pode ser definido como o domínio das coleções de arquivos, compondo o denominado Universo Bibliográfico. Assim, o domínio arquivístico pode ser melhor entendido como o domínio da Arquivologia ou da Arquivística que, por sua vez, pode ser definida como a: “Disciplina que estuda as funções do arquivo e os princípios e técnicas a serem observados na produção, organização, guarda, preservação e utilização dos arquivos. Também chamada arquivística” (ARQUIVO NACIONAL, 2005, p. 37).

A Arquivística contemporânea vem passando por rupturas e mudanças de paradigmas que, em parte, deve-se ao desenvolvimento e aplicação das Tecnologias de Informação e

¹⁰⁵ O domínio é definido como “[...] âmbito ou os assuntos abarcados em uma ciência, área, disciplina e comunidade do conhecimento.” (ALVES, 2010, p. 17).

Comunicação – TIC – no fazer arquivístico (FONSECA, 2005; RONDINELLI, 2005). De modo geral, essas mudanças podem ser percebidas com o surgimento de novos tipos de documentos (os documentos digitais) e novos ambientes (sistemas informatizados) fatores que, consequentemente, exigem uma reformulação dos processos de gestão arquivística de documentos e dos instrumentos utilizados para o tratamento documental, visando o acesso a longo prazo dos documentos em ambientes digitais.

Neste trabalho serão abordados alguns aspectos sobre o estabelecimento de metadados no domínio arquivístico, que são parte dos resultados decorrente de uma pesquisa no Programa de Pós-Doutorado da Universidade Estadual Paulista (Unesp), Júlio de Mesquita Filho, vinculado ao Departamento de Ciência da Informação da Faculdade de Filosofia e Ciências, campus de Marília.

Os metadados são elementos importantes para garantir a identificação, a representação, o acesso e preservação dos documentos arquivísticos em sistemas informatizados. Deste modo, o problema que norteou este trabalho consiste em: como estabelecer os metadados arquivísticos? A hipótese é de que os metadados arquivísticos são estabelecidos com base nas teorias, nos princípios e nas metodologias arquivísticas para o tratamento documental, em especial, nas entidades do domínio arquivísticos identificadas nos modelos conceituais do domínio e nos processos de gestão arquivística de documentos.

Sendo assim, o objetivo consiste em abordar os seguintes aspectos sobre os metadados arquivísticos: seu conceito no domínio, seus principais tipos e os instrumentos para seu estabelecimento padronizado. De modo mais específico, a proposta deste artigo é abordar os metadados estabelecidos em dois instrumentos arquivísticos: a norma internacional ISO 23081-1:2006 (*Information and documentation – Records management processes – Metadata for records – Part 1: Principles*) e, em âmbito nacional, os metadados do esquema definido no Modelo de Requisitos e-ARQ Brasil. A norma ISO 23081-1:2006 contempla os tipos de metadados necessários para a implementação de outras normas, a ISO 15489-1:2001 (*Information and documentation - Records management - Part 1: General*), que trata da gestão de documentos, e; a ISO/TR 15489-2:2001 (*Information and documentation - Records management - Part 2: Guidelines*), que trata das diretrizes para a gestão de documentos. Já o Modelo de Requisitos e-ARQ Brasil, estabelecido com base em normas e orientações de referência nacionais e internacionais, determina requisitos mínimos para os Sistemas Informatizados de Gestão Arquivística de Documentos (SIGAD), incluindo a definição de um esquema de metadados. Ambos os instrumentos apresentam um esquema de metadados conceitual, que diferem entre si por serem baseados em diferentes modelos conceituais do

domínio arquivístico, portanto, constituem-se como foco relevante para uma análise de suas características.

A metodologia utilizada nesta pesquisa consistiu em uma combinação da abordagem exploratória e descritiva com pesquisa bibliográfica realizada com a literatura disponível sobre o tema. O levantamento bibliográfico foi realizado em fontes bibliográficas (primárias e secundárias) da literatura científica impressa e digital, levando em consideração os seguintes assuntos: informação e tecnologia, metadados, aspectos tecnológicos para estabelecimento de metadados, aspectos representacionais para estabelecimento de metadados, gestão arquivística de documentos, sistemas informatizados de gestão arquivística de documentos, diplomática, tipologia documental, metodologia de identificação arquivística, metadados arquivísticos, padrões de metadados em arquivologia, entre outros.

Como resultados destaca-se o mapeamento dos principais tipos de metadados arquivísticos analisados no esquema de metadados da norma ISO 23081-1:2006 e no esquema de metadados do Modelo de Requisitos e-ARQ Brasil. Destaca-se também outros instrumentos necessários para o estabelecimento padronizado de metadados no domínio arquivístico.

2 A GESTÃO ARQUIVÍSTICA DE DOCUMENTOS

De acordo com Rodrigues (2003), o conceito de gestão de documentos teve sua origem após a Segunda Guerra Mundial com a explosão documental no âmbito das administrações públicas e a necessidade de organização dessas massas documentais acumuladas. Na base da gestão de documentos está a Teoria das Três Idades, que também surgiu nesta época trazendo uma nova concepção para os arquivos. Trata-se do “[...] princípio pelo qual os documentos passam por fases estabelecidas de acordo com sua vigência administrativa e frequência de consulta [...]” (RODRIGUES, 2003, p. 35). A primeira idade corresponde aos arquivos correntes, com documentos de caráter funcional e administrativo que estão tramitando ou estão arquivados, consultados com maior frequência. A segunda idade refere-se aos arquivos intermediários, que contempla documentos que não são mais de uso corrente, mas que mantém algum interesse administrativo, por isso, aguardam o cumprimento do prazo estabelecido na tabela de temporalidade para serem eliminados ou recolhidos para um arquivo permanente. A terceira idade corresponde aos arquivos permanentes, que guardam e preservam os documentos de modo definitivo, devido ao seu valor histórico, probatório ou informativo (CONSELHO NACIONAL DE ARQUIVOS, 2011).

O ciclo vital dos documentos arquivísticos acontece com base nessa teoria que revolucionou a Arquivística de modo conceitual e prático, pois instaurou um modelo sistêmico de organização de arquivos que passaram a ser tratados por preceitos arquivísticos (a gestão de documentos) controlando os documentos desde o momento de sua produção até a destinação final (eliminação criteriosa ou guarda permanente) (CONSELHO NACIONAL DE ARQUIVOS, 2011; RODRIGUES, 2003).

Em meados da década de 90 do século XX, o domínio arquivístico passa a vivenciar significativas mudanças ocasionadas pelo desenvolvimento e aplicação das TIC, principalmente no que se refere à produção de documentos arquivísticos em ambientes eletrônicos e digitais, e seu gerenciamento em sistemas informatizados. Para garantir que esses documentos pudessem ser confiáveis, autênticos e preservados ao longo do tempo era fundamental incorporar conceitos arquivísticos em seu gerenciamento. Assim, a gestão de documentos passou a ser requerida e denominada como gestão arquivística de documentos para ressaltar o caráter arquivístico no processo de gestão e diferenciá-la de outros tipos de gerenciamento (CONSELHO NACIONAL DE ARQUIVOS, 2011).

De acordo com o Conselho Nacional de Arquivos (2011, p. 10) a gestão arquivística de documentos pode ser definida como o

Conjunto de procedimentos e operações técnicas referentes à produção, tramitação, uso, avaliação e arquivamento dos documentos em fase corrente e intermediária, visando sua eliminação ou recolhimento para guarda permanente.

O estabelecimento da gestão arquivística de documentos depende da definição de alguns requisitos iniciais: a definição de uma política arquivística do órgão ou instituição; a designação de responsabilidades para execução do programa de gestão (definição das pessoas envolvidas na implantação da gestão arquivística de documentos: profissionais, usuários, gestores etc); o planejamento do programa de gestão (com a análise institucional, estabelecimento de diretrizes para a instituição e sistema, elaboração dos instrumentos e manuais, definição do ambiente tecnológico, definição dos metadados para o ciclo de vida dos documentos etc) e a implantação do programa de gestão (com a execução e acompanhamento do projeto). Os processos e operações técnicas do sistema de gestão ocorrerão com base na definição desses requisitos, que também serão determinados de acordo com as características próprias de cada órgão ou instituição.

A gestão arquivística de documentos comprehende seis processos, também denominados como procedimentos e operações técnicas a serem realizadas pelo sistema de gestão de documentos digitais e convencionais: captura; avaliação, temporalidade e

destinação; pesquisa, localização e apresentação dos documentos; segurança; armazenamento e preservação. A seguir abordam-se os principais aspectos de cada um desses processos:

1) CAPTURA: é a declaração e incorporação do documento no sistema de gestão arquivística por meio das seguintes ações: registro, classificação, indexação, atribuição de restrição de acesso e arquivamento, com o objetivo de identificar um documento como sendo arquivístico e demonstrar sua relação orgânica. Nesse processo são incluídos metadados de vários tipos e níveis de detalhamento, de acordo com as necessidades do órgão ou instituição;

2) AVALIAÇÃO, TEMPORALIDADE E DESTINAÇÃO: a avaliação é uma atividade de análise dos documentos arquivísticos para o estabelecimento dos prazos de guarda e destinação, considerando seus valores primários e secundários. Os prazos de guarda e as ações de destinação devem ser formalizados em uma tabela de temporalidade e destinação da instituição. A destinação dos documentos é efetivada após a seleção e ocorre por meio das seguintes atividades: transferência (passagem dos documentos do arquivo corrente para o intermediário, onde aguardarão o cumprimento dos prazos de guarda e destinação final); recolhimento (entrada dos documentos nos arquivos permanentes); e eliminação (dos documentos que não apresentam valor permanente conforme a tabela de temporalidade e destinação). Para os documentos digitais, os procedimentos de transferência e recolhimento exigem, entre outras questões, a indicação dos metadados atribuídos ao documento e informações que garantam sua autenticidade;

3) PESQUISA, LOCALIZAÇÃO E APRESENTAÇÃO DOS DOCUMENTOS: são itens de recuperação e acesso aos documentos em um sistema informatizado. A pesquisa é feita utilizando-se instrumentos de busca como guias, inventários, catálogos, repertórios e índices; a apresentação consiste em exibir os documentos por meio de dispositivos como monitores e impressoras; e a recuperação é realizada por meio do registro do documento no sistema e de seus metadados associados;

4) SEGURANÇA: refere-se às questões que garantem a integridade dos documentos como, por exemplo, o controle de acesso aos documentos para os usuários do sistema (restrições, autorizações e registro de uso dos documentos); as trilhas de auditoria que registram as ocorrências de acesso, intervenções e tentativas de intervenções nos documentos; as cópias de segurança (cópias periódicas das informações) que são realizadas para posterior restauração

do sistema devido à quebra de segurança, degradação do suporte etc; a segurança sobre a infraestrutura de instalações do acervo digital (temperatura do ambiente para os equipamentos tecnológicos, equipamentos contra incêndio, instalações adequadas de para-raios); entre outras questões;

5) ARMAZENAMENTO: essa questão deve permear todo o ciclo de vida dos documentos, tanto convencionais como os digitais, para garantir a autenticidade e o acesso aos mesmos. É importante considerar as seguintes questões sobre o armazenamento: o volume e a estimativa de crescimento dos documentos (áreas de depósito; tipos e quantidades de estantes; capacidade dos dispositivos de armazenamento); o armazenamento das cópias de segurança dos documentos digitais; o controle dos documentos quanto às características físicas do suporte e do ambiente (contaminação do documento no ambiente, temperatura, umidade, troca de suporte, medidas para conservação do ambiente); a frequência de uso dos documentos (para oferecer uma forma de armazenamento mais rápida aos usuários); o custo relativo das opções de armazenamento (devido ao alto custo com o armazenamento, as opções de terceirização devem ser avaliadas quanto às garantias legais de custódia, restrições de acesso e capacidade tecnológica, por exemplo);

6) PRESERVAÇÃO: para os documentos convencionais a preservação está relacionada com a estabilidade e manutenção do suporte. Para os documentos digitais a preservação incide sobre a manutenção do acesso, que pode implicar em questões como: mudança de suporte, de formato e atualização do ambiente tecnológico; preservação da obsolescência tecnológica e danos físicos ao suporte, por meio de procedimentos de migração e conversão; uso de técnicas de emulação, encapsulamento, preservação tecnológica, formatos abertos, entre outras questões. Um fator importante nesse processo é a inclusão de metadados de preservação, para documentar as alterações nos documentos e preservá-los ao longo do tempo nos sistemas (CONSELHO NACIONAL DE ARQUIVOS, 2011).

Portanto, os metadados desempenham um papel fundamental nos sistemas informatizados, pois representam, contextualizam os documentos e registram as características adquiridas em cada processo de gestão arquivística, evitando que essas informações sejam perdidas ao longo do ciclo de vida nos sistemas.

A Arquivística contemporânea vem se preocupando com o estabelecimento adequado da gestão arquivística de documentos em sistemas automatizados, em especial em relação aos

seguintes aspectos: a definição de diretrizes de gestão, a especificação de requisitos funcionais e os metadados para os sistemas informatizados.

Assim, aborda-se a seguir as características e definições dos metadados no domínio arquivístico e, posteriormente, os tipos metadados arquivísticos identificados na norma ISO 23081-1:2006 e no Modelo de Requisitos e-ARQ Brasil, foco da análise deste trabalho.

3 METADADOS ARQUIVÍSTICOS

Um ponto de vista importante considerado neste trabalho é de que os metadados são produzidos e padronizados há muito tempo em áreas específicas, porém sem ser utilizada esta denominação. Seu estabelecimento em domínios específicos antecede o surgimento dos padrões de metadados, ocorrendo por meio de uma abordagem manual de construção de representações. Portanto, os metadados não são necessariamente digitais ou elementos exclusivos do ambiente digital e *Web*, embora sejam elementos inerentes aos sistemas de informação (MILSTEAD; FELDMAN, 1999; MÉNDEZ RODRÍGUEZ, 2002; GILLILAND, 2008; ALVES, 2010).

Os metadados significam literalmente “dados sobre dados”¹⁰⁶. De modo geral, podemos entender que os metadados são atributos que caracterizam um recurso informacional (ALVES, 2005).

Também existem diferentes tipologias que os caracterizam e essa variedade depende do domínio de aplicação, das necessidades dos usuários desse domínio e dos tipos de recursos informacionais (entidades a serem representadas). De modo geral, os metadados podem ser: intrínsecos ou extrínsecos às entidades descritas (MÉNDEZ RODRÍGUEZ, 2002); administrativos, descritivos, de preservação, técnicos e de uso (GILLILAND, 2008); podem ser criados de modo associado internamente ao recurso ou externamente; podem ser criados automaticamente ou de forma manual; podem ser metadados estáticos ou dinâmicos, de curta duração ou longa duração; podem ser metadados estruturados, semiestruturados e livres; com valores controlados ou não por padrões de conteúdo, que são preenchidos por especialistas do domínio ou não especialistas; podem ser metadados de coleções de recursos ou de um item específico, entre outras características (GILLILAND, 2008).

Quanto às funções que desempenham identificam-se duas perspectivas: uma tradicional e uma emergente. A perspectiva tradicional compreende a realização de funções

¹⁰⁶ Miller (1996); Souza, Catarino e Santos (1997); Milstead e Feldman (1999); Gilliland-Swetland (1999); Souza, Vendrusculo e Melo (2000); Takahashi (2000); Senso e Rosa Piñero (2002); Méndez Rodríguez (2002); Zeng e Qin (2008) entre outros.

como identificação, descrição, busca, recuperação e localização dos recursos; já a perspectiva emergente, comprehende a realização de funções como autoria, propriedade intelectual, formas de acesso, atualização da informação, preservação e conservação, restrição de uso, valoração do conteúdo, visibilidade da informação e acessibilidade dos conteúdos (MÉNDEZ RODRÍGUEZ, 2002).

É importante observar que os metadados de funções tradicionais e emergentes podem ser utilizados conjuntamente. A primeira função dos metadados seria garantir o gerenciamento da informação, por meio da identificação e da descrição para o acesso, busca, localização e recuperação da informação. Em relação aos sistemas de informação, os metadados tem a função de facilitar o gerenciamento e o acesso às informações, o intercâmbio de dados e a interoperabilidade entre sistemas. Já em relação aos usuários dos sistemas, os metadados identificam a informação acessível, sua localização e as condições para se chegar até ela (MÉNDEZ RODRÍGUEZ, 2002).

No contexto do domínio arquivístico, os metadados cumprem papéis tradicionais e emergentes, dentre as funções que desempenham destaca-se: identificação, descrição, busca, recuperação, localização e preservação do recurso; contextualizam os recursos nos sistemas de informação, em especial, nos processos de gestão arquivística de documentos em sistemas informatizados; gerenciam os recursos e outros metadados nos sistemas; otimizam a recuperação; promovem a interoperabilidade entre sistemas, dentre outras funções (GILLILAND, 2008).

Há uma variedade de definições para o termo metadados, por isso, é necessário contextualizá-lo de acordo com o domínio onde está sendo utilizado.

Em sistemas de banco de dados os metadados são considerados atributos¹⁰⁷, características que identificam uma entidade (objeto do mundo real) que pode ser, por exemplo, um recurso informacional ou um documento arquivístico¹⁰⁸ (CHEN, 1990; COAD, YOURDON, 1991).

De acordo com os estudos de Alves (2010, p. 47), os metadados para o domínio bibliográfico ou o domínio de bibliotecas são

¹⁰⁷ “Um atributo é um dado (informação de estado) para qual cada Objeto em uma Classe tem seu próprio valor” (COAD, YOURDON, 1991, p. 117). Essa definição provém da área de Análise Orientada a Objeto e se constitui com conceitos adotados em sistemas de informações.

¹⁰⁸ Um documento arquivístico pode ser tradicional, eletrônico ou digital e é definido como “[...] um documento produzido e/ou recebido e mantido por pessoa física ou jurídica, no decorrer das suas atividades, qualquer que seja o suporte, e dotado de organicidade.” (CONSELHO NACIONAL DE ARQUIVOS, 2011, p. 09).

[...] atributos que representam uma entidade (objeto do mundo real) em um sistema de informação. Em outras palavras, são elementos descritivos ou atributos referenciais codificados que representam características próprias ou atribuídas às entidades; são ainda dados que descrevem outros dados em um sistema de informação, com o intuito de identificar de forma única uma entidade (recurso informacional) para posterior recuperação.

Embora essa definição tenha sido estabelecida para o domínio bibliográfico, é possível utilizá-la para compreender os metadados do Universo Bibliográfico como um todo e, portanto, entendê-los do ponto de vista arquivístico.

No domínio arquivístico o termo metadados vem sendo definido pela CTDE – Câmara Técnica de Documentos Eletrônicos - como “dados estruturados que descrevem e permitem encontrar, gerenciar, compreender e/ou preservar documentos arquivísticos ao longo do tempo” (CONSELHO NACIONAL DE ARQUIVOS, 2011, p. 90). Percebe-se que esta definição explica o significado dos metadados vinculando-os às funções que desempenham no domínio arquivístico.

Com base nos estudos sobre metadados anteriormente desenvolvidos (ALVES, 2010) e na definição estabelecida pela Câmara Técnica de Documentos Eletrônicos, considera-se para neste trabalho que **metadados arquivísticos** são: atributos que representam características próprias ou atribuídas a uma entidade que podem estar em meio digital ou não, com o objetivo de encontrar, gerenciar, compreender e/ou preservar dados e documentos arquivísticos ao longo do tempo em um sistema informatizado de gestão arquivística de documentos ou outro ambiente arquivístico (CONSELHO NACIONAL DE ARQUIVOS, 2011; ALVES, 2010).

Neste sentido, constatou-se que os metadados arquivísticos estão presentes em todo o ciclo de vida dos documentos nos arquivos, mais especificamente, nos processos de gestão arquivística de documentos refletidos nos sistemas informatizados.

3.1 Metadados Da Norma ISO 23081-1:2006

A norma ISO 23081 representa um marco no domínio arquivístico, pois estabelece a criação, gerenciamento e uso dos metadados para a gestão arquivística de documentos. A norma é composta por três partes:

- a) *ISO 23081-1:2006 – Information and documentation – Records management processes – Metadata for records – Part 1: Principles*: aborda os princípios que sustentam e regem a determinação dos metadados para os processos de gestão arquivística de documentos (*records management*) (INTERNATIONAL ORGANIZATION OF STANDARDIZATION, 2013a).

- b) *ISO 23081-2:2009 – Information and documentation – Managing metadata for records -- Part 2: Conceptual and implementation issues:* estabelece um framework que: define os elementos de metadados; permite a descrição padronizada das entidades; fornece uma compreensão comum para a interoperabilidade entre os sistemas e permite a reutilização dos metadados (INTERNATIONAL ORGANIZATION OF STANDARDIZATION, 2013b).
- c) *ISO 23081-3:2011 – Information and documentation – Managing metadata for records -- Part 3: Self-assessment method:* fornece orientações para a realização de uma autoavaliação sobre os metadados da gestão arquivística de documentos (INTERNATIONAL ORGANIZATION OF STANDARDIZATION, 2013c).

A norma ISO 23081-1, é abordada neste trabalho, pois se constitui como um guia para entender, implantar e utilizar metadados próprios da gestão arquivística de documentos e dos processos de negócios (qualquer outro tipo de atividade: comercial, administrativa, atividades de organizações sem fins lucrativos etc.) (INTERNATIONAL ORGANIZATION OF STANDARDIZATION, 2006, 2008).

Os tipos e categorias de metadados da norma ISO 23081-1 foram determinados de acordo com as principais entidades e relações identificadas na análise do domínio arquivístico, em especial, com a análise dos processos de gestão arquivística de documentos e processos de negócios. Assim, a norma está baseada nas seguintes entidades do domínio arquivístico:

- **Entidade documento:** refere-se não só do documento, mas também de seus agrupamentos;
- **Entidade agente:** os responsáveis pelos documentos, atividades e processos;
- **Entidade processos de negócios:** trata-se das funções, atividades, transações, segurança, acesso dos documentos;
- **Entidade processos de gestão:** refere-se aos processos de gestão de documentos (INTERNATIONAL ORGANIZATION OF STANDARDIZATION, 2006, 2008).

Com base nessas entidades a ISO 23081-1 estabelece cinco categorias de metadados, que se subdividem em metadados criados no momento da incorporação do documento no sistema e metadados criados e incorporados posteriormente à inclusão do documento no sistema (INTERNATIONAL ORGANIZATION OF STANDARDIZATION, 2006, 2008). As categorias de metadados contempladas na norma são apresentadas a seguir:

- 1) **Metadados sobre o próprio documento:** a) metadados sobre o documento no momento da captura, b) metadados sobre o documento posterior a captura, c) metadados para o acesso aos documentos, d) metadados para segurança dos documentos;
- 2) **Metadados sobre regras de negócios, política e regulamentações:** a) metadados sobre regras de negócios, políticas e regulamentações no momento da captura, b) metadados sobre regras de negócios, políticas e regulamentações posterior a captura;
- 3) **Metadados sobre agentes:** a) metadados sobre agentes no momento da captura, b) metadados sobre agentes posterior a captura;
- 4) **Metadados sobre processos de negócios:** a) metadados do processo de negócios no momento da captura, b) metadados do processo de negócios posterior a captura;
- 5) **Metadados sobre processos de gestão de documentos:** a) metadados sobre processos de gestão de documentos no momento da captura, b) metadados sobre processos de gestão de documentos posterior a captura.

É importante ressaltar que a norma estabelece ainda tipos mais específicos de metadados, entretanto, optou-se por não apresentá-los devido a sua extensão. Além disso, a norma não define um conjunto obrigatório (esquema de metadados), pois essa escolha deve ser feita com base nos requisitos de cada organização, de suas atividades e dos processos de gestão que realizam. Também existem diferentes níveis de aplicação dos metadados, de acordo com cada organização, que devem ser escolhidos entre: os metadados aplicados aos documentos individuais; os metadados aplicados ao conjunto ou agrupamentos dos documentos; ou os metadados aplicados aos sistemas completos de gestão arquivística de documentos. A decisão sobre os níveis de aplicação dos metadados deverão ser especificadas nas políticas e regras elaboradas pelas instituições para a gestão de metadados nos sistemas informatizados (INTERNATIONAL ORGANIZATION OF STANDARDIZATION, 2006, 2008). A ISO 23081-1 determina que um sistema informatizado deve ser projetado para permitir a incorporação de metadados em qualquer um desses níveis, além disso, é preciso que os metadados sejam atualizados quando necessário para evitar perda de contexto durante sua gestão como, por exemplo, no momento de transferência de uma organização para outra.

Além dos metadados próprios da gestão arquivística de documentos, é possível que outros tipos sejam incorporados ao sistema, vindo de diferentes fontes como, por exemplo: metadados herdados de sistemas de workflow; metadados de correio eletrônico; metadados do comércio eletrônico (*e-business*, *e-commerce* e *e-government*); metadados de preservação;

metadados para descrição do recurso (normas de descrição arquivística); metadados para descoberta de recursos (incluindo metadados de acessibilidade); metadados para o gerenciamento de direitos entre outros (INTERNATIONAL ORGANIZATION OF STANDARDIZATION, 2006, 2008). Por esta razão, é importante considerar um contexto mais geral de criação e incorporação dos metadados na gestão arquivística de documentos, assegurando o estabelecimento de suas relações sem a duplicação de metadados.

A norma ISO 23081-1 não só aborda os princípios para a determinação dos metadados dos processos de gestão arquivística como também determina um modelo conceitual para o domínio arquivístico. Portanto, constitui-se como um instrumento relevante para a determinação padronizada de esquemas de metadados no domínio arquivístico, além de servir de base para avaliação dos esquemas de metadados já existentes.

3.2 Metadados do modelo de requisitos e-arq brasil

O Modelo de Requisitos para os Sistemas Informatizados de Gestão Arquivística de Documentos e-ARQ Brasil constitui-se como uma especificação de requisitos a serem cumpridos pela organização produtora e recebedora de documentos nos sistemas informatizados de gestão arquivística para garantir a confiabilidade, autenticidade e acessibilidade dos documentos arquivísticos. O modelo estabelece requisitos mínimos para os Sistemas Informatizados de Gestão Arquivística de Documentos - SIGAD¹⁰⁹ independente da plataforma tecnológica e seu objetivo é orientar a implantação da gestão arquivística de documentos digitais e não digitais em sistemas informatizados, por meio de especificações técnicas, funcionais e por meio do estabelecimento de um esquema de metadados conceitual. O desenvolvimento do modelo teve início em 2004 e em 2011 foi lançada a versão que contempla o esquema de metadados conceitual para os sistemas informatizados (CONSELHO NACIONAL DE ARQUIVOS, 2011).

A metodologia para o estabelecimento do esquema foi baseada em quatro etapas: a primeira etapa identificou os metadados referidos no e-ARQ Brasil, em especial, na descrição das etapas de gestão arquivística (Parte I do modelo) e nos requisitos apresentados na seção aspectos de funcionalidade do SIGAD (Parte II). A segunda etapa complementou os metadados identificados, a partir de normas e referências bibliográficas das áreas de

¹⁰⁹ Um SIGAD pode ser definido como [...] um conjunto de procedimentos e operações técnicas que visam o controle do ciclo de vida dos documentos, desde a produção até a destinação final, seguindo os princípios da gestão arquivística de documentos e apoiado em um sistema informatizado (CONSELHO NACIONAL DE ARQUIVOS, 2011, p. 11).

arquivologia e diplomática. A terceira etapa confrontou o levantamento inicial com esquemas, normas e padrões de metadados semelhantes, nacionais e internacionais. A quarta etapa analisou, definiu e aprovou o esquema (CONSELHO NACIONAL DE ARQUIVOS, 2011).

O esquema de metadados do Modelo de Requisitos e-ARQ Brasil apresenta uma estrutura de entidades e relacionamentos diferente do modelo conceitual da norma ISO 23081-1. Além disso, os atributos conceituais são definidos de modo mais específico e não apenas categorias e subcategorias de metadados. O esquema de metadados do modelo foi baseado em seis entidades do domínio arquivístico: documento; evento de gestão; classe; agente; componente digital e evento de preservação. Apresenta-se a seguir cada uma das entidades, uma breve explicação e os tipos de metadados contemplados em cada uma delas:

- 1) **ENTIDADE DOCUMENTO:** são os documentos arquivísticos gerenciados pelo SIGAD, esses documentos estabelecem relações entre si, formando agregações denominadas processos ou dossiês; estabelece relação com a classe no momento de sua captura no SIGAD e com uma série de eventos de gestão arquivística ao longo de seu ciclo de vida. Os documentos arquivísticos podem ser classificados e gerenciados de duas formas: individualmente ou agregados em processos ou dossiês. Se o documento arquivístico for digital ele será composto por um ou mais componentes digitais. De modo geral, os metadados desta entidade tem como característica: descrever, identificar, proporcionar o acesso, garantir a integridade e a preservação dos documentos etc. Além disso, podem ser aplicados apenas em alguns níveis de agregação (processo/dossiê, volume e documento), conforme as especificações do esquema. Os metadados da entidade documento especificados no esquema conceitual do modelo de requisitos são: 1 Identificador do documento; 2 Número do documento; 3 Número do protocolo; 4 Identificador do processo/dossiê; 5 Número do processo/dossiê; 6 Identificador do volume; 7 Número do volume; 8 Tipo de meio; 9 Status; 10 Identificador de versão; 11 Título; 12 Descrição; 13 Assunto; 14 Autor; 15 Destinatário; 16 Originador; 17 Redator; 18 Interessado; 19 Procedência; 20 Identificador do componente digital; 21 Gênero; 22 Espécie; 23 Tipo; 24 Idioma; 25 Quantidade de folhas/página; 26 Numeração sequencial dos documentos; 27 Indicação de anexos; 28 Relação com outros documentos; 29 Níveis de acesso; 30 Data de produção; 31 Classe; 32 Destinação prevista; 33 Prazo de guarda; 34) Localização (CONSELHO NACIONAL DE ARQUIVOS, 2011).
- 2) **ENTIDADE EVENTO DE GESTÃO:** são as ações de gestão arquivística que ocorrem com os documentos ao longo de seu ciclo de vida (captura, tramitação,

abertura e encerramento de processo/dossiê, classificação, desclassificação, eliminação, transferência, recolhimento, entre outros). O evento de gestão estabelece relações com o documento e com o agente responsável pela ação. De modo geral, os metadados desta entidade referem-se aos eventos de captura, movimentação e controle do ciclo de vida dos documentos arquivísticos. Os metadados da entidade evento de gestão especificados no esquema conceitual do modelo de requisitos são os seguintes:

1 Captura; 2 Tramitação; 3 Transferência; 4 Recolhimento; 5 Eliminação; 6 Abertura_processo/dossiê; 7 Encerramento_processo/dossiê; 8 Reabertura_processo/dossiê; 9 Abertura_volume; 10 Encerramento_volume; 11 Juntada_anexação; 12 Juntada_apensação; 13 Desapensação; 14 Desentranhamento; 15 Desmembramento; 16 Classificação_sigilo; 17 Desclassificação_sigilo; 18 Reclassificação_sigilo (CONSELHO NACIONAL DE ARQUIVOS, 2011).

- 3) **ENTIDADE CLASSE:** refere-se aos níveis de agregação do plano de classificação organizados de forma hierárquica: classe, subclasse, grupos e subgrupos. Em cada classe estão associadas informações a respeito da temporalidade e destinação dos documentos, sendo que todas as alterações devem ser registradas em metadados específicos da classe. As classes estão relacionadas aos documentos arquivísticos nelas classificados. De modo geral, os metadados desta entidade referem-se à configuração e administração do plano de classificação. Os metadados da entidade classe especificados no esquema conceitual do modelo de requisitos são agrupados em duas categorias, conforme apresentados a seguir: 1 Descrição da Classe (1.1 Classe_nome; 1.2 Classe_codigo; 1.3 Classe_subordinação; 1.4 Registro de abertura; 1.5 Registro de desativação; 1.6 Reativação da classe; 1.7 Registro de mudança de nome de classe; 1.8 Registro de deslocamento de classe; 1.9 Registro de extinção; 1.10 Indicador de classe ativa/inativa); 2 Temporalidade associada à classe (2.1 Classe_codigo; 2.2 Prazo de guarda na fase corrente; 2.3 Evento que determina a contagem do prazo de guarda na fase corrente; 2.4 Prazo de guarda na fase intermediária; 2.5 Evento que determina a contagem do prazo de guarda na fase intermediária; 2.6 Destinação final; 2.7 Registro de alteração; 2.8 Observações) (CONSELHO NACIONAL DE ARQUIVOS, 2011).
- 4) **ENTIDADE AGENTE:** são os usuários que acessam o SIGAD, que podem ser usuários que desempenham funções específicas (ou cargos com responsabilidade e autoridades bem definidas) e como grupo (conjunto de usuários reunidos para a realização de uma atividade em comum, por um determinado tempo). Os agentes se relacionam: entre si (usuários podem ser agregados em papéis e grupos e um usuário

pode estar associado com um ou mais papéis) e, com o evento de gestão e evento de preservação pelo qual é responsável. Os metadados da entidade agente especificados no esquema conceitual do modelo de requisitos são: 1 Nome; 2 Identificador; 3 Autorização de acesso; 4 Credenciais de autenticação; 5 Relação; 6 Status do agente (CONSELHO NACIONAL DE ARQUIVOS, 2011).

- 5) **ENTIDADE COMPONENTE DIGITAL:** são os objetos digitais que compõem o documento arquivístico digital como, por exemplo: arquivos de computadores com informações de conteúdo, forma e composição necessárias à apresentação do documento arquivístico. As ações de preservação são realizadas nos componentes digitais, sendo que cada documento pode estar relacionado com um ou mais componentes digitais que, por sua vez, pode conter informações de um ou mais documentos. Os metadados da entidade componente digital especificados no esquema conceitual do modelo de requisitos são: 1 Identificador do componente digital; 2 Nome original; 3 Características técnicas; 4 Formato de arquivo; 5 Armazenamento; 6 Ambiente de software; 7 Ambiente de hardware; 8 Dependências; 9 Relação com outros componentes digitais; 10 Fixidade (CONSELHO NACIONAL DE ARQUIVOS, 2011).
- 6) **ENTIDADE EVENTO DE PRESERVAÇÃO:** são as ações de preservação realizadas nos documentos arquivísticos digitais, como por exemplo: migração, compressão, validação, decifração, entre outras. O evento de preservação relaciona-se com o componente digital e com o agente responsável pela ação de preservação. Ressalta-se aqui os eventos mais importantes a serem registrados, podendo ser incluídos outros de acordo com a necessidade. Os metadados da entidade evento de preservação especificados no esquema conceitual do modelo de requisitos são: 1 Compressão; 2 Decifração; 3 Validação de assinatura digital; 4 Verificação de fixidade; 5 Cálculo hash; 6 Migração; 7 Replicação; 8 Verificação de vírus; 9 Validação (CONSELHO NACIONAL DE ARQUIVOS, 2011).

O esquema de metadados do Modelo de Requisitos e-ARQ Brasil constitui-se como um esquema conceitual no qual apresenta: os metadados, suas definições, seus refinamentos (atributos que representam informações mais específicas), as especificações que os normalizam, os requisitos que atendem no SIGAD, notas de aplicação e o nível de obrigatoriedade dos atributos. O esquema caracteriza-se por ser flexível, pois cada órgão ou

instituição deve eleger os metadados que mais se adéquam às suas necessidades, de acordo com os níveis de obrigatoriedade.

Os esquemas de metadados apresentados: esquema de metadados da norma ISO 23081-1:2006 e o esquema de metadados do Modelo de Requisitos e-ARQ Brasil, apresentam semelhanças e diferenças entre si. Primeiramente pode-se destacar que os dois esquemas de metadados constituem-se como esquemas conceituais que podem ser implementados em qualquer sistema informatizado de gestão arquivística de documentos – SIGAD. Entretanto, para isso necessitam de uma estrutura automatizada, de uma sintaxe de codificação e persistência (armazenamento) dos atributos e seus valores. Além disso, devem seguir um modelo conceitual para estruturas as entidades e relações no banco de dados e ao mesmo tempo fazer existir a estrutura lógica de descrição das entidades.

Deste modo, além de escolher qual esquema de metadados seguir, qual conjunto de metadados e qual nível será utilizado em um órgão ou instituição específica, é preciso adotar outros instrumentos que auxiliarão no estabelecimento e padronização dos metadados e seus valores. Neste sentido, apresenta-se a seguir os instrumentos existentes no domínio arquivístico que auxiliarão no estabelecimento padronizado dos metadados e seus valores:

- a) **INSTRUMENTOS PRINCIPAIS DA GESTÃO ARQUIVÍSTICA DE DOCUMENTOS:** esses instrumentos determinam alguns metadados e padronizam seus valores nos esquemas de metadados ou padrões de metadados já existentes. Os instrumentos principais da gestão arquivística são: Plano de classificação, baseado nas funções e atividades do órgão ou entidade; tabela de temporalidade e destinação; Manual de gestão arquivística de documentos; Esquema de classificação referente à segurança e ao acesso aos documentos.
- b) **INSTRUMENTOS ADICIONAIS DA GESTÃO ARQUIVÍSTICA DE DOCUMENTOS:** esses instrumentos padronizam os valores dos metadados. Os instrumentos adicionais são: Glossário; Vocabulário controlado e Tesauro.
- c) **INSTRUMENTOS DE APOIO DA GESTÃO ARQUIVÍSTICA DE DOCUMENTOS:** esses instrumentos fornecem informações sobre o gerenciamento dos metadados e seus valores no sistema, informações sobre o contexto da gestão arquivística, informações sobre usuários e permissões de acesso. Os instrumentos de apoio da gestão arquivística são: Relatório de análise do contexto jurídico-administrativo do órgão ou entidade; Relatório dos riscos que envolvem as atividades desenvolvidas pelo órgão ou entidade; Plano de contingência e plano de prevenção contra desastres; Estrutura organizacional e delegação de competências do órgão ou

entidade; Registro dos funcionários e das permissões de acesso aos sistemas do órgão ou entidade.

- d) **MODELOS CONCEITUAIS:** os modelos conceituais são responsáveis por definir as entidades, os atributos e os relacionamentos do domínio arquivístico, além de estruturar os metadados e seus relacionamentos nos bancos de dados dos sistemas informatizados. Os modelos conceituais do domínio arquivístico são: MoReq – *Modular Requirements for Records Systems* (Modelo de Requisitos para a Gestão de Registros Eletrônicos); e-ARQ Brasil - Modelo de Requisitos para Sistemas Informatizados de Gestão Arquivística de Documentos; *Design criteria standard for electronic records management software applications*: DOD 5015.2-STD, 2002; *Requirements for electronic records management systems: Functional requirements, United Kingdom*, 2002.
- e) **ESQUEMAS DE CODIFICAÇÃO:** dividido em esquemas de codificação de conteúdo e esquemas de codificação de valores que, por sua vez, são subdivididos em tipos mais específicos. **Esquemas de codificação de conteúdo**, composto por *padrões que determinam a estrutura de dados* (determinam uma estrutura codificada para os metadados, ex.: um esquema de metadados estruturado na linguagem XML apresenta um metadado (atributo codificado) e o espaço do valor, que em alguns casos são padronizados de acordo com esquemas externos); e *padrões de intercâmbio e interoperabilidade de dados* (como, por exemplo, o protocolo OAI-PMH - *Open Archives Initiative Protocol for Metadata Harvesting*). **Esquemas de codificação de valores**, composto por *padrões de conteúdo de dados* (são padrões ou códigos elaborados para a representação arquivística, como por exemplo: ISAD (G) - Norma geral internacional de descrição arquivística; ISAAR (CPF) - Norma internacional de registro de autoridade arquivística para entidades coletivas, pessoas e famílias; ISDF - Norma internacional para descrição de funções; ISDIAH - Norma internacional para descrição de instituições com acervo arquivístico; NOBRADE - Norma brasileira de descrição arquivística); e *padrões de valores de dados* (padronizam os valores no espaço do valor e os instrumentos são: Tesauros, vocabulário controlado, listas de autoridade, esquemas de classificação, tabelas de temporalidade e destinação, Plano de classificação, Esquema de classificação referente à segurança e ao acesso aos documentos, entre outros).
- f) **PADRÕES DE METADADOS UTILIZADOS NO DOMÍNIO ARQUIVÍSTICO:** esses padrões de metadados são ou podem ser utilizados no domínio arquivístico,

entretanto, é importante destacar que alguns deles pertencem a outros domínios como, por exemplo, o padrão MARC 21 do domínio bibliográfico e o padrão Dublin Core do domínio Web. Os padrões de metadados do domínio arquivístico são: MARC 21 e desdobramentos, em especial MARC 21 HOLDINGS e MARC 21 AUTORIDADES (incluindo a versão em MARC XML); EAD - Encoded Archival Description; DC - Dublin Core; e-Government Metadata Standard – e-GMS, United Kingdom, v. 3.0, 2004; Metainformação para Interoperabilidade de Portugal – MIP, Lisboa, 2006; MoReq 2 – Model requirements for the management of electronic records update and extension, 2007; Padrão de Metadados do Governo Eletrônico – e-PMG, Brasil (minuta); PREMIS Data Dictionary for Preservation Metadata – version 2.

- g) ESQUEMAS DE METADADOS CONCEITUAIS UTILIZADOS NO DOMÍNIO ARQUIVÍSTICO:** esses esquemas constituem-se como esquemas conceituais, ou seja, não apresentam estrutura codificada e implementável para os metadados, entretanto, servem de base para avaliação dos requisitos e metadados arquivísticos de outros padrões e esquemas de metadados existentes no domínio. Os esquemas de metadados conceituais são: Esquema do Modelo de Requisitos e-ARQ Brasil; ISO23081-1 *Information and documentation – Records management processes – Metadata for records – Part 1: Principles*, 2006.

Esses instrumentos são responsáveis não só pelo estabelecimento dos metadados, mas também da padronização de seus valores e estruturação do banco de dados do sistema informatizado. A adoção desses instrumentos, segundo os requisitos e necessidades da instituição arquivística, garante uma representação estruturada e padronizada, oferecendo maior qualidade ao sistema informatizado.

4 CONCLUSÕES

Pode-se dizer que os metadados são atributos que representam características próprias ou atribuídas a uma entidade arquivística ao longo dos processos de gestão arquivística de documento em sistemas informatizados. Durante seu ciclo de vida em sistemas informatizados, os documentos arquivísticos vão adquirindo novos contextos de uso que necessitam ser registrados. Assim, novas camadas de metadados são adicionadas aos documentos conforme passam pelos processos de gestão arquivística.

Existem diferentes tipos e níveis de metadados distribuídos entre os processos de gestão arquivística. Os metadados aqui apresentados foram estabelecidos com base em

modelos conceituais do domínio arquivístico que identificam as entidades, seus atributos (características essenciais que devem ser representadas) e os relacionamentos que realizam no contexto dos processos de gestão arquivística.

Os esquemas de metadados apresentados constituem-se como esquemas conceituais e flexíveis, que podem ser implementados nos sistemas informatizados de qualquer instituição arquivística. Nesse sentido, ressalta-se nesse estudo a importância da gestão de metadados na gestão arquivística de documentos e a utilização de outros instrumentos arquivísticos (instrumentos da gestão arquivística; instrumentos adicionais e de apoio à gestão arquivística; modelos conceituais; esquemas de codificação; padrões de metadados e esquemas conceituais) que padronizam o estabelecimento de metadados arquivísticos e seus valores.

Com este estudo foi possível verificar que os metadados arquivísticos são estabelecidos com base em entidades do domínio arquivístico, modelos conceituais e processos de gestão arquivística. Foi possível verificar também que as teorias sobre os metadados estão em consonância com o domínio arquivístico, pois os metadados arquivísticos são estabelecidos de acordo com os aspectos tecnológicos (modelos conceituais, requisitos do sistema e esquemas) necessários para sua consolidação nos sistemas informatizados de gestão arquivística, juntamente com os aspectos representacionais (esquemas externos de codificação de dados e valores dos metadados arquivísticos), que garantem a padronização dos metadados e seus valores. Essas características são uma tendência na comunidade de metadados e ao mesmo tempo uma exigência do domínio arquivístico, por se tratar de um domínio específico.

REFERÊNCIAS

- ALVES, R. C. V. **Metadados como elementos do processo de catalogação.** 2010. 132f. Tese (Doutorado em Ciência da Informação)-Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília, 2010.
- ALVES, R. C. V. **Web Semântica:** uma análise focada no uso de metadados. 2005. 180f. Dissertação (Mestrado em Ciência da Informação)-Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília, 2005.
- ARQUIVO NACIONAL (BRASIL). **Dicionário brasileiro de terminología arquivística.** Rio de Janeiro: Arquivo Nacional, 2005.
- CHEN, P. **Modelagem de dados:** a abordagem Entidade-Relacionamento para projeto lógico. São Paulo: McGraw-Hill, MAKRON Books, 1990.
- COAD, P. YOURDON, E. **Análise baseada em objetos.** 2. ed. Rio de Janeiro: Campus, 1991.
- CONSELHO NACIONAL DE ARQUIVOS (BRASIL). CÂMARA TÉCNICA DE DOCUMENTOS ELETRÔNICOS. **e-ARQ Brasil:** modelo de requisitos para sistemas

informatizados de gestão arquivística de documentos - Versão 1.1. Rio de Janeiro: Arquivo Nacional, 2011. Disponível em:
<http://www.conarq.arquivonacional.gov.br/cgi/cgilua.exe/sys/start.htm>. Acesso em: 23 jun. 2011.

FONSECA, M. O. **Arquivologia e ciência da informação**. Rio de Janeiro: Editora FGV, 2005.

GILLILAND, A. J. Setting the stage. In: BACA, M. (Ed.). **Introduction to metadata**. 2. ed. Los Angeles: Getty Research Institute, 2008. p. 1-19.

GILLILAND-SWETLAND, A. J. La definición de los metadatos. In: **INTRODUCCIÓN a los metadatos: vías a la información digital**. [S. l.]: GETTY, 1999. p. 1-9.

IFLA. **Declaração dos Princípios Internacionais de Catalogação**. 2009. Disponível em: http://archive.ifla.org/VII/s13/icp/ICP-2009_pt.pdf. Acesso em: 27 jan. 2010.

INTERNATIONAL ORGANIZATION OF STANDARDIZATION. **Información y documentación** - Procesos de gestión de documentos - Metadatos para la gestión de documentos. Parte 1: Principios. ISO 23081-1:2006. Revista Española de Documentación Científica, Madrid, v. 31, n. 2, p. 27-301. 2008. Disponível em:
<http://redc.revistas.csic.es/index.php/redc/article/view/429/441>. Acesso em: 10 dez. 2013.

INTERNATIONAL ORGANIZATION OF STANDARDIZATION. **ISO 23081-1: Information and documentation – Records management process – Metadata records – Part 1: Principles**, 2006.

INTERNATIONAL ORGANIZATION OF STANDARDIZATION. **ISO 23081-1:2006 – Information and documentation - Records management processes - Metadata for records - Part 1: Principles**, 2013a.

INTERNATIONAL ORGANIZATION OF STANDARDIZATION. **ISO 23081-2:2009 - Information and documentation - Managing metadata for records - Part 2: Conceptual and implementation issues**, 2013b.

INTERNATIONAL ORGANIZATION OF STANDARDIZATION. **ISO/TR 23081-3:2011 Information and documentation - Managing metadata for records - Part 3: Self-assessment method**, 2013c.

MÉNDEZ RODRÍGUEZ, E. **Metadatos y recuperación de información**: estándares, problemas y aplicabilidad en bibliotecas digitales. Gijón: Trea, 2002. 429 p.

MILLER, P. **Metadata for the masses**. [S. l.: S. n.], 1996. Disponível em:
<http://www.ariadne.ac.uk/issue5/metadata-masses/>. Acesso em: 16 abril 2004.

MILSTEAD, J.; FELDMAN, S. **Metadata**: cataloging by any other name. *Online*, [S. l.], january 1999. Disponível em: <<http://www.online.com/online/ol1999/milstead1.html>>. Acesso em: 22 jun. 2004.

RODRIGUES, A. C. **Tipologia documental como parâmetro para a gestão de documentos de arquivos**: um manual para o Município de Campo Belo, MG. 2003. Dissertação (Mestrado em História Social)-Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo, São Paulo, SP. Disponível em: <<http://www.teses.usp.br/teses/disponiveis/8/8138/tde-25042003-181526/publico/tesaAnaCelia.pdf>>. Acesso em: 26 jun 2007.

RONDINELLI, R. C. **Gerenciamento arquivístico de documentos eletrônicos:** uma abordagem teórica da diplomática contemporânea. 4. ed. Rio de Janeiro: FGV, 2005.

SENSO, J. A.; ROSA PIÑERO, A. de la. El concepto de metadato. Algo más que descripción de recursos electrónicos. **Ciência da Informação**, Brasília, v. 32, n. 2, p. 95-106, maio/ago. 2003.

SOUZA, M. I. F.; VENDRUSCULO, L. G.; MELO, G. C. Metadados para a descrição de recursos de informação eletrônica: utilização do padrão Dublin Core. **Ciência da Informação**, Brasília, v. 29, n. 1, p. 93-102, jan./abr. 2000. Disponível em: <<http://www.ibict.br/cionline/290100/29010010.pdf>>. Acesso em: 6 jul. 2000.

SOUZA, T. B. de; CATARINO, M. E.; SANTOS, P. C. dos. Metadados: catalogando dados na Internet. **Transinformação**, Campinas, v. 9, n. 2, maio/ago. 1997. Disponível em: <<http://www.puccamp.br/~biblio/tbsouza92.html>>. Acesso em: 12 jul. 2000.

TAKAHASHI, T. (Org.). **Sociedade da Informação no Brasil:** o livro verde. Brasília: Ministério da Ciência e Tecnologia, 2000.

ZENG, M. L.; QIN, J. **Metadata.** New York: Neal-Schuman Publishers, 2008.

UM ROTEIRO PARA MODELAGEM CONCEITUAL DE SISTEMAS DE INFORMAÇÃO BASEADA EM PRINCÍPIOS ONTOLÓGICOS

A ROADMAP FOR CONCEPTUAL MODELING OF INFORMATION SYSTEMS BASED ON ONTOLOGICAL PRINCIPLES

Stefane de Melo Silva

Fernanda Farinelli

Mauricio Barcellos Almeida

Resumo: Sistemas de Informação devem ser baseados em representações consistentes da realidade. Modelos construídos de maneira *ad-hoc*, sem diretrizes que orientem as ações de modelagem, podem resultar em representações inadequadas, gerar dificuldade de comunicação e impossibilidade de integração e interoperabilidade entre sistemas. As ontologias são instrumentos fundamentais para a melhoria da modelagem conceitual, uma vez que são capazes de fornecer semânticas bem definidas. Este artigo apresenta um passo a passo para a modelagem conceitual, em forma de um roteiro, com base nos princípios ontológicos extraídos da *Unified Foundational Ontology - UFO*. Para isto, contextualiza-se a atividade de modelagem conceitual e os princípios ontológicos utilizados, através das categorias básicas de entidades e relações definidas pela ontologia de fundamentação. Em seguida, são descritas as atividades do roteiro proposto, bem como seus insumos e produtos gerados. Ao final, de forma a testar a viabilidade do roteiro proposto, descreve-se um estudo de caso de um modelo real do domínio da administração pública de Minas Gerais.

Palavras-chave: Modelagem conceitual. Ontologia. *Unified Foundational Ontology - UFO*. Roteiro.

Abstract: Information Systems should be based on consistent representations of reality. Conceptual models made by ad-hoc way, without directives that guide the actions of modeling can result in poor representations, generating communication difficulties and impossibilities of integration and interoperability between systems. Ontologies are fundamental tools for the improvement of conceptual modeling, since they are able to provide well-defined semantics. This paper presents a walkthrough to support conceptual modeling in the form of a roadmap, based on ontological principles extracted from the Unified Foundational Ontology. For this, we contextualizes the activity of conceptual modeling and ontological principles used by the basic categories of entities and relationships defined by foundational ontology specification. Then, the activities of the proposed roadmap are described, as well as its inputs and outputs. Finally, in order to test the viability of the roadmap, we describe a case study of a real domain model of government of Minas Gerais.

Keywords: Conceptual modeling. Ontology. Unified Foundational Ontology - UFO. Roadmap.

1 INTRODUÇÃO

Empresas que desenvolvem sistemas de informação estão inseridas em um mercado altamente competitivo e dinâmico. É perceptível, então, a necessidade crescente de investimentos em suas atividades de engenharia de software, na tentativa de melhorar os seus processos e produzir produtos e serviços de melhor qualidade. Para isto, existe uma grande variedade de padrões disponíveis para utilização por empresas de tecnologia da informação

(TI), que variam de processos tradicionais como o Rational Unified Process - RUP (KRUCHTEN, 2001) a metodologias ágeis, centradas em dados ou em arquitetura, etc. A atividade de modelagem conceitual está presente na maioria dessas metodologias de desenvolvimento e é de fundamental importância, pois uma modelagem bem definida é a base para um bom sistema de informação (SÁNCHEZ, CAVERO e MARCOS, 2005).

Modelos são estruturas de representação criadas a partir de abstrações da realidade. Podem variar de estruturas conceituais, como modelos mentais, a representações gráficas, como os modelos comumente elaborados em projetos de software utilizando a Unified Modeling Language – UML (OMG, 2012). Construir modelos consiste em descrever a realidade, reproduzindo conceitos e seus relacionamentos relevantes para os usuários, a partir de uma construção mental especificada por uma linguagem (SÁNCHEZ, CAVERO e MARCOS, 2005). Os modelos são aplicados na Ciência da Informação no âmbito da organização e representação da informação como forma de representação de um domínio. Na Ciência da Computação, os modelos são utilizados como mecanismos de captura e comunicação de informações pertinentes à construção de um sistema de informação, passando de modelos conceituais a modelos de dados implementados em um banco de dados.

Segundo Guarino (1998), uma ontologia é um artefato de engenharia de software constituído por um vocabulário específico, que descreve certa realidade através de uma hierarquia de conceitos conectados por relações de subordinação. Dentre os tipos de ontologias existentes na literatura cabe destacar as ontologias de fundamentação, que são baseadas em estudos, teorias e ferramentas e buscam identificar as entidades básicas do mundo, sendo capazes de se adequar a vários domínios (OLIVEIRA, 2009).

Para fins de modelagem computacional, modelos e ontologias são conceitos similares, pois ambos têm a finalidade de representar um domínio de conhecimento a partir de uma conceitualização da realidade (MENDONÇA e ALMEIDA, 2012). Porém, a aplicação de ontologias adiciona regras e restrições para possibilitar a padronização e a melhoria semântica da modelagem.

A escolha por representar determinadas características em um modelo, geralmente, é feita de maneira ad-hoc, ou seja, parte da conceitualização particular de cada modelador, sem um embasamento consistente ou algum padrão que o oriente. A aplicação de ontologias à modelagem pode proporcionar mais subsídios para o modelador, fazendo com que a sua tomada de decisão passe a ser baseada em um padrão e deixe de ser ad-hoc. Outro benefício dessa aplicação é facilitar o entendimento sobre “o que é o sistema”, na medida em que a semântica é trabalhada na definição do domínio de negócio. Assim, acredita-se que a

aplicação de princípios ontológicos para definição de conceitos e relações em um modelo conceitual pode proporcionar uma aproximação entre a realidade modelada e o modelo produzido.

Este artigo apresenta um passo a passo para a modelagem conceitual, em forma de um roteiro, baseado em princípios ontológicos extraídos da Unified Foundational Ontology - UFO como proposta para a melhoria semântica de modelos conceituais de sistemas de informação e também para a melhoria dos processos comunicacionais e melhor entendimento de um domínio do conhecimento, conforme prerrogativas da Ciência da Informação. Para a proposta do roteiro, este trabalho contextualiza a atividade de modelagem conceitual e detalha os princípios ontológicos utilizados através das categorias básicas de entidades e relações definidas pela UFO. Além disso, de forma a validar a viabilidade do roteiro proposto, descreve-se um estudo de caso com um modelo real do domínio da administração pública de Minas Gerais.

O restante deste trabalho está organizado como se segue. A seção dois introduz a modelagem conceitual, contextualizando-a para a Ciência da Informação e para a Ciência da Computação. A seção três apresenta as ontologias e detalha os principais conceitos relevantes para o roteiro proposto. A seção quatro propõe o roteiro e apresenta o estudo de caso. A seção cinco apresenta as conclusões e perspectivas futuras.

2 Modelagem conceitual

A atividade de criação de modelos, ou modelagem, figura entre diversas áreas da ciência, caracterizando-se multidisciplinar. Para este trabalho, que transita entre a Ciência da Computação e a Ciência da Informação, entende-se importante a apresentação de definições sobre modelagem e também as diversas abordagens de sua aplicação encontradas na literatura, pertinentes a essas duas áreas, principalmente. Também são apresentadas aqui, e relacionadas aos conceitos de modelagem, as definições e abordagens sobre ontologias revisadas, que completam parte da base da fundamentação teórica desta pesquisa.

Modelos são estruturas que possibilitam a representação de informações a partir de abstrações do mundo real. Podem variar de modelos abstratos, como representações mentais de forma descritiva, como os modelos matemáticos e diagramas gráficos, a modelos concretos, que representam todas as características e os requisitos necessários a uma ideia ou conceito (SÁNCHEZ, CAVERO E MARCOS, 2005). Exemplos de modelos concretos pode ser uma maquete de uma construção ou uma miniatura de um objeto.

A construção dessas abstrações da realidade que tomam forma a partir dos modelos, cada pessoa toma como base a sua conceitualização particular e o seu conhecimento prévio sobre o mundo. Um modelo construído significa, então, abstrações da realidade representadas a partir da perspectiva de uma pessoa ou grupo de pessoas.

Para Frigg (2006), existem três tipos de questões relativas a modelos: questões de natureza ontológica, de natureza epistemológica e de natureza semântica. As questões de natureza ontológica dizem respeito a definir o que são modelos dentre, por exemplo: objetos físicos, estruturas da teoria dos conjuntos, descrições, equações, ou combinações dos anteriores. Nas questões de natureza epistemológica, os modelos são considerados meios de aprendizado e, a partir deles, é possível descobrir características do sistema representado. As questões de natureza semântica dizem respeito à função representacional dos modelos. Nesse contexto, cabe distinguir modelos de fenômenos, teorias e modelos de dados. Um modelo de fenômeno lida com problemas de representação científica. Esse modelo é uma estrutura que torna verdadeiras as sentenças de uma teoria. Por sua vez, um modelo de dados é uma visão idealizada de dados obtidos por observação direta. Para Evermann (2005), os modelos possuem finalidade de meio de comunicação: é necessário entender a realidade, o que pode ser feito com a ajuda da modelagem.

Para a Ciência da Informação, a modelagem está relacionada à representação do conhecimento, onde os mecanismos para tal representação possibilitam a elaboração de linguagens documentárias verbais e notacionais com o objetivo de organização e de recuperação (CAMPOS, 2001). Segundo Oliveira (2009), a representação constitui um processo de conhecimento, o qual consiste na assimilação das coisas, sendo os seres humanos os conucedores e observadores da realidade externa. Como a Ciência da Informação necessita de estudos sobre representação, codificação e uso racional da informação (CAPURRO, 1991), os modelos e a atividade de modelar se apresentam como uma alternativa.

2.1 Modelagem de sistemas de informação

Na Ciência da Computação, os modelos possuem objetivos distintos, de representar e de auxiliar na implementação das estruturas computáveis e estão relacionados principalmente a modelos de dados, estruturas muito utilizadas em empresas de TI para representar a implementação de um sistema de informação através de estruturas de banco de dados (OLIVEIRA, 2009). Tais modelos são construídos de maneira que o armazenamento e recuperação ocorram de maneira eficiente e, em geral, são representados através da linguagem

entidade relacionamento (ER), conforme exemplo da Figura 1. Cougo (1997) afirma que os modelos de dados são compostos por níveis de abstração distintos, denominados modelo físico, modelo lógico e modelo conceitual. Cada um desses níveis possui diferentes aspectos a representar. A área da Ciência da Computação que se preocupa com o desenvolvimento e manutenção de sistemas, a Engenharia de Software, define disciplinas específicas para abordagem da modelagem ao longo de um processo de software, sendo chamadas de análise e desenho de software (PÁDUA, 2001). Interessados em maiores detalhes sobre estas disciplinas podem consultar Pressman (2011), Sommerville (2003).

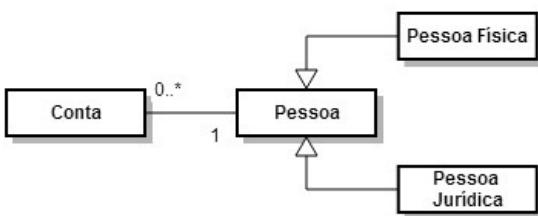
FIGURA 4 - Exemplo de um modelo de dados ER



Fonte: Elaborado pelos autores.

Na atividade de modelagem de sistemas, representam-se os principais processos de um contexto específico objetivando a construção do sistema. A etapa em que são criados modelos sob a perspectiva do entendimento humano é conhecida como modelagem conceitual. Os modelos conceituais são obtidos a partir de abstrações de aspectos da realidade, isentos de decisões ou estruturas que possibilitam a implementação de um sistema e são, geralmente, representados através da UML. Estas últimas são aplicadas em níveis posteriores de modelagem de SI. As abstrações são formas de especificar as entidades e as relações entre as entidades, em um domínio do conhecimento de interesse para o sistema em desenvolvimento.

FIGURA 5 - Exemplo de modelo UML



Fonte: Elaborado pelos autores.

A modelagem conceitual de SI, como conhecida hoje, é resultado de pesquisa produzida nos últimos 50 anos. Os primeiros modelos de dados (YOUNG; KENT, 1958; BOSAK ET AL., 1962) deram lugar aos modelos semânticos (ABRIAL, 1974; JARDINE, 1976; CHEN, 1976), aos quais se seguiram os modelos orientadas a objeto (BOOCH, 1993; RUMBAUGH ET AL., 1991; JACOBSON ET AL., 1992).

Ao longo dos anos, a pesquisa sobre modelos conceituais tem sido motivada pela busca de formas cada vez mais apuradas de representar a realidade em SI. Entretanto, os modelos semânticos dispõem de um conjunto limitado de constructos para a tarefa de modelagem conceitual. O modelo ER, por exemplo, pressupõe que a realidade pode ser articulada por dois tipos de constructos apenas, as entidades e os relacionamentos. Como alternativa Guarino (1998) destaca a necessidade de uma conceitualização comum entre sistemas de informação. A interoperabilidade entre sistemas é possível apenas se as linguagens subjacentes aos modelos possuem conceitualizações que se sobrepõem em algum nível. Em última instância, apenas uma conceitualização compartilhada, produto de maior esforço na comunicação humana, pode proporcionar melhorias na modelagem. Ontologias são alternativas para especificar tal conceitualização.

3 ONTOLOGIAS

O significado do termo ontologia é apresentado de diferentes formas em áreas diversas e tem se caracterizado pela coexistência de abordagens interdisciplinares, que abrangem Ciência da Computação, Filosofia, Ciência da Informação, Linguística, dentre outros campos.

Na filosofia, onde se encontram as suas raízes, ontologia é um sistema particular de categorias para uma determinada visão do mundo. Nas áreas de computação e sistemas de informação, refere-se ao conjunto de formalismos que representam conceitos e suas relações, fundamentados na semântica de um domínio de conhecimento (OLIVEIRA, 2009).

Para Guarino (1998), a ontologia é um artefato de engenharia de software, constituído por um vocabulário específico que descreve certa realidade, mais um conjunto de prerrogativas sobre o significado pretendido para os termos do vocabulário. Já Gruber (1993) define ontologias como um esquema conceitual para sistemas de banco de dados, o qual fornece a descrição lógica dos dados.

Para Borst (1997), uma ontologia é uma especificação formal e explícita de uma conceitualização. Nesse contexto, conceitualização refere-se a um modelo abstrato de algum fenômeno do mundo; formal significa que pode ser lida por máquinas e explícita diz respeito a restrições explicitamente definidas. Segundo Sowa (2000), uma ontologia é um catálogo de tipos de coisas que se assume existir em um domínio, na perspectiva de uma pessoa que usa uma linguagem com a finalidade de representação.

Em Benevides e Guizzardi (2009), ontologia é definida como um sistema formal e filosoficamente bem fundamentado de categorias que pode ser usado para articular conceitualizações e modelos em domínios específicos do conhecimento, para áreas como

Modelagem Conceitual e Modelagem Organizacional. Para a Inteligência Artificial, a Engenharia de Software e a Web Semântica, uma ontologia é um artefato concreto de engenharia, projetado com um propósito específico e sem prestar nenhuma ou quase nenhuma atenção a aspectos teóricos de fundamentação. Significa um modelo de um domínio específico do conhecimento, por exemplo, um modelo da área de finanças, logística, doenças infecciosas, etc., e expresso em uma linguagem de representação específica, como UML, ER, SDM, entre outras.

Fonseca (2007) afirma que é importante distinguir as conotações existentes sobre o termo ontologia, dada a sua utilização em diferentes campos de pesquisa. Uma distinção apresentada pelo autor classifica ontologias de sistemas de informação e ontologias para sistemas de informação. No primeiro caso, as ontologias de SI são usadas para criar modelos conceituais. A segunda distinção se concentra no uso de ontologias como parte do sistema de informação e descrevem o vocabulário de um domínio. Através de ontologias de SI, é possível avaliar a capacidade de representação de linguagens de modelagem. Comparam-se os constructos da linguagem aos constructos ontológicos de uma ontologia utilizada como referência, também denominada ontologia de fundamentação, nomenclatura utilizada por esta pesquisa.

3.1 Ontologias de fundamentação para modelagem conceitual

Baseando-se em princípios filosóficos consolidados, é possível identificar e reparar modelagens deficientes, conduzidas de forma ad-hoc e sem correspondência com o mundo real através das ontologias de fundamentação.

De acordo com Guizzardi et. al. (2008), ontologias, no sentido filosófico, têm sido desenvolvidas em filosofia desde Aristóteles com sua teoria de Substância e Acidentes e, mais recentemente, várias dessas teorias têm sido propostas sob o nome de ontologias de fundamentação (Foundational Ontologies). Desde o fim da década de oitenta, observa-se um crescente interesse no uso dessas ontologias de fundamentação no processo de avaliação e (re)engenharia de linguagens de modelagem conceitual. A hipótese inicial, e que foi posteriormente confirmada por várias evidências empíricas, pode ser explicada através da seguinte argumentação: (i) modelos conceituais são artefatos produzidos com o objetivo de representar uma determinada porção da realidade segundo uma determinada conceitualização; (ii) ontologias de fundamentação descrevem as categorias que são usadas para a construção dessas conceitualizações. Pode-se, portanto, concluir que uma linguagem adequada de modelagem conceitual deve possuir primitivas de modelagem que reflitam as categorias

conceituais definidas em uma ontologia de fundamentação (BENEVIDES E GUIZZARDI, 2009).

Exemplos de ontologias de fundamentação são: a Basic Formal Ontology - BFO (GRENON; SMITH, 2004), a Descriptive Ontology for Linguistics and Cognitive Engineering - DOLCE (MASOLO et al., 2003), a General Formal Ontology - GFO (HERRE et al., 2006) e a Unified Foundational Ontology - UFO (GUIZZARDI; WAGNER, 2004). Em sistemas de informação, cabe ainda citar a pesquisa pioneira sobre ontologias de SI conhecida como Bunge-Wand-Weber Ontology - BWW (WAND; WEBER, 1990) (WAND; STOREY; WEBER, 1999).

3.1.1 Ontologia de fundamentação unificada

A Ontologia de Fundamentação Unificada (Unified Foundational Ontology - UFO) (GUIZZARDI e WAGNER, 2004) é resultado da combinação da GFO (HERRE et. al. 2006) e da DOLCE (MASOLO et. al., 2003), representando assim, uma síntese de ontologias de fundamentação. Foi criada na tentativa de tentar suprir as limitações de captura de conceitos básicos de linguagens de modelagem conceitual de outras ontologias de fundamentação, aproveitando as características positivas e sanando as limitações detectadas (GUIZZARDI; WAGNER, 2004).

A UFO foi proposta inicialmente em por Guizzardi (2005) e é dividida em três camadas: UFO-A, que define o núcleo ontológico de indivíduos duradouros (endurants); UFO-B, que define um incremento da UFO-A incluindo os termos relacionados a processos (perdurants); UFO-C, que define termos relacionados às esferas sociais incluindo aspectos linguísticos, construída sobre as partes A e B. O escopo deste trabalho se limita, neste momento, ao estudo da UFO-A, a partir da qual foi proposta a linguagem de modelagem OntoUML. Esta linguagem integra a UFO-A e o Meta-Object Facility (MOF) da UML 2.0 (OMG, 2012), que é reparado para garantir um isomorfismo em seu mapeamento para a estrutura definida pela ontologia. Este trabalho limita-se também ao mapeamento das entidades e relações definidas pela UFO consideradas como mais importantes para a modelagem conceitual. O mapeamento completo da UFO está em Guizzardi (2005).

O

QUADRO 2 apresenta o mapeamento das principais categorias de entidades determinadas pela UFO.

QUADRO 2- Categorias de entidades da UFO

Categoria	Descrição
Quantity	Refere-se à quantidade de matéria. Ex.: água, açúcar, vinho, etc. A subQuantidade de matéria sempre se refere ao mesmo tipo de matéria. Ex.: Qualquer subQuantidade de água também é água.
Collective	Podem ter partes que não são do mesmo tipo, como árvore e floresta. Neste caso, um conjunto que não é infinitamente divisível, ou seja, uma floresta não pode ser composta apenas de uma árvore, mas um conjunto de árvores.
Kind	Representa uma substância cujas instâncias são complexos funcionais. Os exemplos incluem as instâncias de coisas naturais (pessoa, cão, árvore) e de artefatos (carro, televisão).
Subkind	Representa um subconjunto do tipo Kind, ou seja, uma especialidade de Kind. Ex.: Homem é um tipo específico de Pessoa.
FunctionalComplex	Os complexos são constituídos por peças que desempenham papéis múltiplos no contexto de um conjunto. As partes de um complexo têm em comum o fato de ter uma ligação funcional com o complexo todo.
Role	Representa um papel desempenhado por uma entidade do tipo Kind. Ex.: Estudante é um papel de uma Pessoa.

Fonte: Guizzardi (2005).

O QUADRO 3 apresenta o mapeamento das principais categorias de relações determinadas pela UFO.

QUADRO 3- Classificação de relações da UFO.

Categoria	Descrição
Generalization (is-a)	Relação de generalização/especialização entre Kind e Role; entre Roles; entre Subkind e Role; entre Kind e Subkind
Formal Relation	Associação formal entre entidades do tipo Kind
ComponentOf	Relação meronímica entre FunctionComplex
MemberOf	Relação meronímica entre Collectives ou entre FunctionComplex e Collectives
SubCollectionOf	Relação meronímica entre Collectives
SubQuantityOf	Relação meronímica entre Quantities

Fonte: Guizzardi (2005).

4 ROTEIRO PROPOSTO: ELABORAÇÃO DO MODELO CONCEITUAL

A presente seção descreve o roteiro para a modelagem conceitual, contemplando as suas etapas, insumos e produtos (seção 4.1). Além disso, apresenta-se um estudo de caso para avaliar a viabilidade da proposta por meio de uma necessidade de negócio real (Seção 4.2).

4.1 Descrição das etapas

O roteiro proposto para elaboração do modelo conceitual de SI é baseado em princípios ontológicos e organiza os procedimentos em quatro atividades principais. Para a realização de cada atividade é necessário um ou mais artefatos que vão subsidiar a sua execução, estes artefatos são chamados de entrada. Por sua vez, cada atividade produz um resultado, que é chamado de saída. O QUADRO 4 apresenta a relação das entradas e saídas de cada atividade proposta. Em seguida, cada subseção descreve uma atividade específica.

QUADRO 4 – Relação de entradas e saídas das atividades propostas

Atividade	Entrada	Saída
Identificação de termos e fatos	Especificação de requisitos, especificação de caso de uso	Lista de termos e fatos
Elaborar glossário do domínio	Lista de termos e fatos	Glossário
Classificação dos termos e fatos	Lista de tipos de entidades e relações, Glossário, Lista de termos e fatos	Lista de termos e fatos classificados, Modelo semi-formal
Construção do diagrama conceitual	Modelo semi-formal	Diagrama Conceitual

Fonte: Elaborado pelos autores.

4.1.1 Identificação de termos e fatos do domínio

A atividade de identificação de termos e fatos do domínio consiste na transposição dos conceitos e características observados no ambiente de domínio que se pretende modelar. Entende-se como termos os conceitos que representam entes físicos ou entes que se pode descrever por características. Os fatos se referem às ações que ocorrem no ambiente, que denotam comportamento dos entes ou interligação entre entes. A seguir, sugerem-se algumas técnicas para a seleção de termos e fatos candidatos. Para exemplificar considere a sentença a seguir: O cliente folheia o catálogo e seleciona itens para adicionar na cesta de compras.

Seleção de termos candidatos: Seleção de palavras que tem função substantiva na sentença (substantivos, verbos substantivados, expressões, etc). Normalmente, estas palavras são determinadas por artigo, numeral ou pronome possessivo. Os atores do caso de uso também são termos candidatos. Na sentença exemplo, correspondem as palavras destacadas em negrito e itálico. Os termos geralmente são elementos que representam coisas tangíveis, funções exercidas por pessoas ou coisas, eventos, ou interações.

Seleção de fatos candidatos: Seleção de verbos que designam ações que afetam os termos, ou verbos que não designam ações, mas que servem para ligar o sujeito a seu

predicado quando estes dois equivalem a termos. Os casos de uso também são fatos candidatos. Na sentença exemplo, correspondem as palavras sublinhadas.

4.1.2 Elaborar glossário do domínio

Um glossário refere-se a uma organização de todas as palavras (termos e fatos) com sua respectiva definição, importantes para o entendimento do negócio que está sendo modelado. Esta atividade consiste em definir cada um dos termos e fatos identificados na atividade anterior e se necessário de novos termos e fatos que sejam importantes para a compreensão do ambiente que está sendo observado e modelado.

A obtenção dos conceitos ou definições dos termos e fatos do glossário pode ser pela análise de documentação pertinente ao ambiente em questão, por meio de entrevistas ou seções de brainstorm com especialistas no assunto. O importante a ressaltar é que a definição dada a um termo ou fato deve ser consensual entre os participantes do projeto e clientes. Para orientar a elaboração das definições, sugere-se o entendimento da Teoria do Conceito de Dahlberg (1978) e da Teoria da Terminologia Dahlberg (1983).

4.1.3 Classificação dos termos e fatos quanto às categorias de entidades e relacionamento

Esta atividade versa a classificar os termos e fatos segundo as categorias de entidades e relacionamento que são propostas pela UFO (seção 3.1.1). A seguir são apresentadas categorias propostas para este roteiro.

Classificação para termos: Determinar para cada termo identificado na primeira atividade (seção 4.1.1) qual das categorias do QUADRO 5 o termo se refere.

QUADRO 5- Proposta de categorias de termos para modelagem conceitual

Categoría	Descrição
Quantidade	Quantitativo de matéria. Algo que possa ser dimensionado.
Coletivo	Algo que faz parte de um conjunto de elementos, não necessariamente do mesmo tipo.
Coisa	Algo que fornece um princípio de identidade para suas instâncias. Representa uma espécie na sua forma mais primitiva.
Complexo Funcional	Todo constituído de partes que desempenham múltiplos papéis.
Papel	Função específica desempenhada por uma entidade do tipo Coisa.

Fonte: Elaborado pelos autores.

Classificação para termos: Determinar para cada fato identificado na primeira atividade (seção 4.1.1) qual das categorias do

QUADRO 6 o fato se refere.

QUADRO 6- Proposta de categorias de fatos para modelagem conceitual

Categoría	Descrição
É um/ É uma	É uma relação de generalização que remete a uma hierarquia entre termos onde os termos específicos possuem as mesmas características do termo mais genérico e outras características específicas dele.
É Sub quantidade de	É uma relação meronímica entre termos da categoria quantidade.
É Sub coleção de	É uma relação meronímica entre termos que são coletivos (plurais).
É Membro de	É uma relação meronímica entre termos complexos funcionais, singulares ou coletivos, como parte de termos que são coletivos.
É Componente de	É uma relação meronímica entre termos que são complexos funcionais.

Fonte: Elaborado pelos autores.

Documentação semi-formal de termos e fatos: Consiste em escrever sentenças em linguagem natural utilizando os termos e fatos categorizados a partir da sua definição semiformal com base ontológica. O QUADRO 7 apresenta a coluna da esquerda corresponde à definição semiformal existente entre uma categoria de termo e uma categoria de fato. A coluna da direita refere-se a um exemplo de escrita semiformal.

QUADRO 7 – Exemplo de documentação semi-formal

Definição semiformal	Exemplo
Papel é uma Coisa	Cliente é uma Pessoa
Quantidade é sub quantidade de Quantidade	200 ml de água é sub quantidade de 1 litro de água
Coletivo é sub coleção de Coletivo	Um conjunto de árvores é sub coleção de Uma floresta
Papel é membro de Coleção	Um jogador é membro de Um Time
Coisa é componente de Coisa	Coração é componente de Pessoa

Fonte: Elaborado pelos autores.

4.1.4 Construção do diagrama conceitual

O diagrama conceitual é uma representação visual estrutura por uma abordagem de modelagem conceitual. Conforme discutido na seção 2.1, duas abordagens são comunmente usadas na engenharia de software, a abordagem Entidade-Relacionamento que origina os diagramas E-R e a abordagem orientada a objeto que utiliza a linguagem UML para elaborar os diagramas de classes. Na seção 3.1.1, é introduzido a abordagens ontológica e a modelagem OntoUML como uma extensão da UML que tem suas classes e relacionamentos baseados na ontologia UFO-A que também produz um diagrama de classe.

Para a construção do diagrama conceitual, optou-se pela abordagem UML, utilizando como representação visual o diagrama de classes, porém sem determinar estereótipos de

classes e relacionamentos. Em linhas gerais, as entidades ou classes equivalem aos termos categorizados conforme QUADRO 5 e os relacionamentos são os fatos categorizados conforme

QUADRO 6. A seguir são descritos os passos para a elaboração do diagrama conceitual.

Passo 1: Definição de Classes ou Entidades.

As entidades são os elementos ou termos que persistem ao tempo e espaço. Selecionar os termos identificados e categorizados e criar uma classes no diagrama para cada um dos termos. As classes são representadas no diagrama como retângulos.

Observe que alguns termos que foram identificados na atividade da seção 4.1.1 podem não ter sido categorizados, porém, estes termos podem representar características ou propriedades de termos categorizados, também conhecidos como os atributos das entidades.

Passo 2: Definição de relacionamentos associativos.

Os relacionamentos associativos equivalem aos fatos que denotam as seguintes categorias: É Sub quantidade de, É Sub coleção de, É Membro de, É Componente de. Estes relacionamentos são representados no diagrama de classe como associações, agregações ou composições.

Passo 3: Definição de relacionamentos hierárquicos.

Os relacionamentos hierárquicos equivalem aos fatos que denotam a categorias “É um/é uma”, e são representados no diagrama de classe como herança.

Passo 4: Definir a multiplicidade entre associações.

A multiplicidade de uma associação refere-se ao número de instâncias da classe de uma das extremidades da associação em relação a classe da outra extremidade. Pode-se indicar o número mínimo e máximo das instâncias.

4.2 Estudo de caso: gestão da estrutura administrativa estadual

O objetivo do estudo de caso foi avaliar a viabilidade do roteiro proposto, aplicando-o a um requisito real de SI. No estudo de caso foi utilizado um cenário de negócio real da administração pública do Estado de MG. A seção 4.2.1 apresenta uma parte da especificação de requisitos do cenário “Gerir estrutura administrativa estadual”. Cada uma das seções seguintes (4.2.2 à 4.2.5) equivale ao resultado da realização de uma das atividades propostas no roteiro da seção 4.1.

4.2.1 Especificação de requisitos

A FIGURA 6 equivale à parte da especificação de requisitos do cenário de negócio da gestão da estrutura organizacional da Administração Pública.

FIGURA 6 - Especificação de requisitos do estudo de caso

A Administração Pública é um conjunto de estruturas estatais indispensáveis ao atendimento de necessidades da coletividade. Fazem parte da Administração Pública a Administração Pública Direta e a Administração Pública Indireta.

- Administração Pública Direta faz parte da Administração Pública e é formada por órgãos. Os órgãos são estruturas estatais que não possuem personalidade jurídica própria e autonomia administrativa. Seu patrimônio e despesas são realizados diretamente pelo orçamento estadual.
- Administração Pública Indireta faz parte da Administração Pública e é formada por entidades que possuem personalidade jurídica própria, patrimônio e autonomia administrativa e cujas despesas são realizadas através de orçamento próprio.
- Os Órgãos da Administração Pública Direta são: **Governandoria, Vice-governandoria, Secretarias de Estado e Órgãos Autônomos**.
- As Entidades da Administração Pública Indireta são: **Autarquias, Fundações Públicas, Empresas Públicas e Sociedade de Economia Mista**.

Fonte: Elaborado pelos autores.

4.2.2 Identificação de termos e fatos do domínio

No texto de especificação de requisitos (FIGURA 6) os termos identificados correspondem aos substantivos e verbos substantivados e foram marcados como negrito e itálico, os fatos correspondem aos verbos sublinhados.

4.2.3 Elaborar glossário do domínio

Dentre as experiências obtidas ao longo da pesquisa, cabe destacar a discussão com a equipe acerca das definições dadas a cada termo ou fato identificado. A FIGURA 7 demonstra um extrato do glossário criado nesta atividade, ressalta-se que não foi apresentado o artefato como um todo em função da limitação de espaço para o artigo.

FIGURA 7 - Extrato do glossário do estudo de caso

Termo ou Fato	Definição
Administração Pública	Conjunto de estruturas estatais indispensáveis ao atendimento de necessidades da coletividade.
Administração Pública Direta	Estruturas da Administração Pública que não possuem personalidade jurídica própria, patrimônio e autonomia administrativa e cujas despesas são realizadas diretamente através do orçamento estadual.
Administração Pública Indireta	Estruturas da Administração Pública, vinculadas à Administração Pública Direta, com personalidade jurídica própria, patrimônio e autonomia administrativa e cujas despesas são realizadas através de orçamento próprio.
Órgão	Estrutura da Administração Pública Direta, ligados diretamente ao poder estadual, que reúne atribuições exercidas pelos agentes públicos que o integram com o objetivo de expressar a vontade do Estado.
Entidade	Estrutura estatal da Administração Pública Indireta que executa as Atividades Administrativas de maneira descentralizada.

Fonte: Elaborado pelos autores.

4.2.4 Classificação dos termos e fatos quanto às categorias de entidades e relacionamento

Classificação de termos: Os termos foram classificados de acordo com a categoria e organizados no QUADRO 8.

QUADRO 8- Categorização dos termos do estudo de caso

Categoria de Termo	Termos do Estudo de Caso
Quantidade	Nenhum termo identificado
Coletivo	Administração Pública, Administração Pública Direta, Administração Pública Indireta
Coisa	Órgãos, Entidades, Governadoria, Vice-governadoria, Secretarias de Estado, Órgãos Autônomos, Autarquia, Fundação Pública, Empresa Pública, Sociedade de Economia Mista
Complexo Funcional	Nenhum termo identificado
Papel	Unidade Administrativa

Fonte: Elaborado pelos autores.

Classificação de fatos: Os fatos categorizados foram organizados no QUADRO 9.

QUADRO 9- Categorização dos fatos do estudo de caso

Categoria de Fato	Fatos do Estudo de Caso
É um/ É uma	É, são
É Sub quantidade de	
É Sub coleção de	Faz parte
É Membro de	É formada
É Componente de	

Fonte: Elaborado pelos autores.

Documentação semi-formal de termos e fatos: Um extrato da documentação semiformal é apresentado no

QUADRO 10.

QUADRO 10 – Exemplo de documentação semi-formal

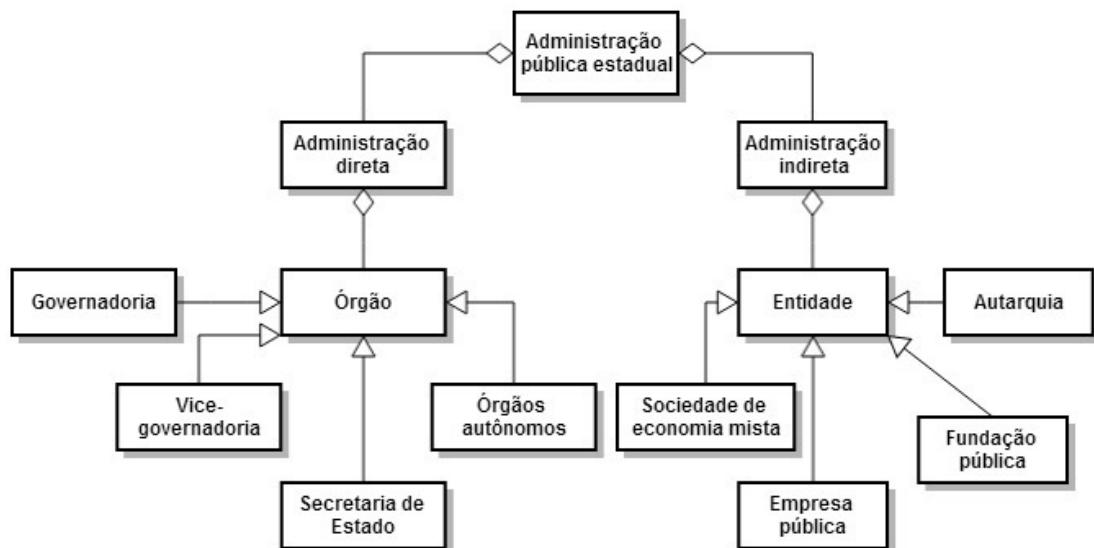
Definição semiformal	Exemplo
Coisa é uma Coisa	Governadoria é um órgão. Autarquia é uma Entidade.
Coletivo é sub coleção de Coletivo	Administração Pública Direta é sub coleção de Administração Pública Administração Pública Indireta é sub coleção de Administração Pública
Coisa é membro de Coleção	Órgão é membro de Administração Pública Direta Entidade é membro de Administração Pública Indireta

Fonte: Elaborado pelos autores.

4.2.5 construção do diagrama conceitual

A FIGURA 8 refere-se ao diagrama conceitual baseado no extrato da documentação semiformal apresentado.

FIGURA 8 - Diagrama conceitual do estudo de caso



Fonte: Elaborado pelos autores.

5 CONSIDERAÇÕES FINAIS

Este artigo discutiu a aplicação de princípios ontológicos à modelagem conceitual de sistemas de informação e teve como objetivo apresentar a proposta de um roteiro que possa orientar a atividade de modelagem com base na ontologia, possibilitando a geração de modelos semanticamente bem definidos. O roteiro foi elaborado com base nos princípios ontológicos extraídos da UFO e a partir das categorias de entidades e relações definidas em sua especificação. As premissas sobre a atividade de modelagem conceitual encontradas na

literatura também orientaram a proposta. Ao final, foi realizado um estudo de caso sobre o domínio da Administração Pública Estadual de Minas Gerais para a verificação de viabilidade do roteiro proposto.

Através do estudo de caso realizado pode-se concluir que o roteiro proposto é viável de ser aplicado à modelagem conceitual para diversos domínios de informação. Conclui-se também que a proposta é caracterizada como uma forma “leve” de uso de ontologias para pessoas que não estão ambientadas com os seus conceitos e ainda assim possibilita diversos benefícios para o desenvolvimento de um sistema de informação, tais como melhoria da interoperabilidade entre sistemas, a diminuição de ambiguidades em modelos e a melhoria dos processos comunicacionais.

A proposta aqui apresentada pode ser vista como uma oportunidade para promover a aplicação prática de ontologias no âmbito da Ciência da Computação, a qual se percebe ainda pouco explorada e também como oportunidade de esclarecimento da abordagem de modelagem conceitual de sistemas de informação para os profissionais da Ciência da Informação. Buscou-se contribuir para a interdisciplinaridade de saberes e aproximação das áreas de Ciência da Informação e Ciência da Computação, a partir do estudo de um problema real de interesse acadêmico e gerencial na tecnologia da informação.

Por fim, podem ser pontuadas algumas oportunidades de trabalhos futuros, como: incremento de outras categorias de entidades e relações especificadas pela UFO para um roteiro mais completo; elaboração de novas versões do roteiro para outras ontologias de fundamentação aplicáveis à modelagem conceitual; aplicação do roteiro proposto para a modelagem de outros domínios de informação; validação do roteiro com profissionais que atuam na prática de modelagem conceitual.

REFERÊNCIAS

ABRIAL, J. R. Data semantics. In: J.W. Klimbie & K.L. Koffeman (Eds.). **Proceedings of the IFIP Working Conference Data Base Management** (p. 1–60). Amsterdam: North-Holland, 1974.

BENEVIDES, A. B.; GUIZZARDI, G. A model-based tool for conceptual modeling and domain ontology engineering in OntoUML. In FILIPE, J.; CORDEIRO, J. (Ed.). ICEIS. Heidelberg: Springer, 2009. **Lecture Notes in Business Information Processing**. vol. 24, p. 528–538. ISBN 978-3-642-01346-1. Disponível em:
http://www.inf.ufes.br/~gguizzardi/ICEIS_2009.pdf Acesso em: 10 fev. 2013.

BOOCH, G. **Object-oriented analysis and design with applications**, 2 nd ed. Redwood: Benjamin Cummings, 1993.

BORST, W. N. **Construction of engineering ontologies**; PhD Thesis (1997). Disponível em: <<http://www.ub.utwente.nl/webdocs/inf/1/t0000004.pdf>>. Acesso em: 26 Jan 2013.

BOSAK, R., et al. **An information algebra: phase 1 report - language structure group of the CODASYL development committee**. Communications of the ACM, New York, v. 5, n. 4, p. 190-204, Apr. 1962.

CAMPOS, M. L. A. **A organização de unidades do conhecimento em hiperdocumentos**: o modelo conceitual como um espaço comunicacional para a realização da auditoria. 2001. 198 p. Tese (Doutorado em Ciência da Informação) – Departamento de Ensino e Pesquisa do IBICT – Instituto Brasileiro de Informação em Ciência e Tecnologia, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2001.

CAPURRO, R. What is information science for? A philosophical reflection. In: Vakkari, Perti, Cronin, Blaise. **Conceptions of library and information science**. Tempere, Taylor Graham, p. 82-93, 1991.

CHEN, P. **The entity-relationship model**: towards a unified view of data. ACM Transactions on Database Systems, 1(1), p. 9–36, 1976.

COUGO, P. S. **Modelagem Conceitual e Projeto de Bancos de Dados**. Campus, 1997.

DAHLBERG, Ingetraut. Teoria do conceito. **Ciência da informação**, v. 7, n. 2, 1978.

DAHLBERG, Ingetraut. Terminological definitions: characteristics and demands. In: **Problèmes de la définition et de la synonymie en terminologie**. Québec, GIRSTERM, 1983. p. 13-51.

EVERMANN, J. Thinking Ontologically: Conceptual vs. Design Models in UML. Business Systems Analysis with Ontologies. Idea Group, p. 82-104, 2005.

FARRADANE, J. E. L. Relational Indexing: part I. **Journal of information science**, n.1, p. 267-276, 1980a.

FONSECA, F. The double role of ontologies in information science research. **Journal of the American Society for Information Science and Technology**, v. 58, n. 6, p. 786-793, 2007.

FRIGG, R. **Models in Science**. Disponível em: <<http://plato.stanford.edu/entries/models-science/>>. Acesso em: 26 Jan 2012.

GRENON P., SMITH B., GOLDBERG L. Biodynamic ontology: Applying BFO in the Biomedical Domain, in Pisanelli DM (ed). **Ontologies in Medicine. Proceedings of the Workshop on Medical Ontologies**, Rome. October 2003. IOS Press, Studies in Health Technology and Informatics, vol 102, p. 20-38, 2004.

GRUBER, T. **What is an ontology?** (1993). Disponível em: <<http://wwwksl.stanford.edu/kst/what-is-an-ontology.html>>. Acesso em: 26 Jan 2012.

GUARINO, N. **Formal ontology and information systems**. (1998). Disponível em: <<http://citeseer.ist.psu.edu/guarino98formal.html>>. Acesso em: 15 Dez 2011.

GUIZZARDI, G. **Ontological Foundations for Structural Conceptual Models**. Universal Press, The Netherlands, 2005.

GUIZZARDI, G., FALBO, R.A., GUIZZARDI, R.S.S. A importância de Ontologias de Fundamentação para a Engenharia de Ontologias de Domínio: o caso do domínio de Processos de Software. **Revista IEEE América Latina**, v.6, p. 244-251, 2008.

GUIZZARDI, G; WAGNER, G. A Unified Foundational Ontology and Some Applications of it in Business Modeling. In: **Proceedings of the Enterprise Modelling and Ontologies for Interoperability workshop**, 2004.

HERRE, H.; HELLER, B.; BUREK, P.; HOEHNDORF, R.; LOEBE, F.; MICHALEK, H.. **General Formal Ontology (GFO)**: A Foundational Ontology Integrating Objects and Processes. Part I: Basic Principles. Research Group Ontologies in Medicine (Onto-Med), University of Leipzig, 2006.

JACOBSON, I., CHRISTERSON, M., JONSSON, P., & OVERGAARD, G. **Object-oriented software engineering: A use case driven approach**. Boston: Addison-Wesley, 1992.

JARDINE, D. A. **The ANSI/SPARC DBMS model**. Proceedings of the second SHAREWorking Conference on Database Management Systems. Amsterdam: North Holland, 1976.

KRUCHTEN, P. **The Rational Unified Process**: An Introduction, Second Edition. Addison-Wesley, Boston, 2001.

MASOLO, C.; BORGO, S.; GANGEMI, A.; GUARINO, N.; OLTRAMARI, A.; SCHNEIDER, L. (2003) **WonderWeb Library of Foundational Ontologies – Preliminary Report**. Disponível em:
[<http://wonderweb.semanticweb.org/deliverables/documents/D17.pdf>](http://wonderweb.semanticweb.org/deliverables/documents/D17.pdf). Acesso em: 26 jan 2012.

MENDONÇA, F.M.; ALMEIDA, M.B. **Modelos e teorias para a representação**: uma teoria ontológica sobre o sangue humano. Anais do XIII Enancib. Rio de Janeiro, 2012.

OLIVEIRA, V.N.P. **Uma investigação sobre a avaliação de modelagem conceitual baseada em ontologias**: estudo de caso de modelos para sistemas de informação desenvolvidos na Universidade Federal de Minas Gerais. 2009. 177 p. Dissertação (Mestrado em Ciência da Informação) – Escola de Ciência da Informação, Universidade Federal de Minas Gerais, Belo Horizonte, 2009.

OMG - Object Management Group. **Unified Modeling Language (UML)**, Infrastructure, V2.1.2. Disponível em: [<http://www.omg.org/docs/formal/07-11-04.pdf>](http://www.omg.org/docs/formal/07-11-04.pdf). Acesso em: 26 jan 2012.

PADUA, W. D. **Engenharia de software**. Rio de Janeiro: LTC, 2001.

PRESSMAN, R. S. **Engenharia de software**. McGraw Hill Brasil, 2011.

RUMBAUGH, J., BLAHA, M., PREMERLANI, W., EDDY, F., & LORENSEN, W. **Object-oriented modeling and design**. New York: Prentice Hall, 1991.

SÁNCHEZ, D. M.; CAVERO, J. M., MARCOS, E. **On models and ontologies**. (2005). Disponível em:
[<http://kybele.escet.urjc.es/PHISE05/papers/sesionIV/SanchezCaveroMarcos.pdf>](http://kybele.escet.urjc.es/PHISE05/papers/sesionIV/SanchezCaveroMarcos.pdf). Acesso em: 10 Dez 2011.

SMITH, B., WELTY C. A. Ontology: Towards a new synthesis. *Formal Ontology in Information Systems*, **ACM Press**, 2001.

SOMMERVILLE, I. **Engenharia de Software**. 6a. edição, Addison-Wesley/Pearson, 2003.

SOWA, J. F. **Ontology**. (2000). Disponível em: <<http://www.jfsowa.com/ontology/>>. Acesso em: 26 jan 2012.

Wand, Y.; Weber, R. Mario Bunge's ontology as a formal foundation for information systems concepts. In: **Studies on Mario Bunge's Treatise**. p. 123-150. Amsterdam: Radopi, 1990.

WAND, Y; STOREY, V. C.; WEBER, R. An ontological analysis of the relationship construct in conceptual modeling. **ACM Transactions on Database Systems (TODS)**, v. 24, n. 4, p. 494-528, 1999.

YOUNG JR, John W.; KENT, Henry K. **An abstract formulation of data processing problems**. In: Preprints of papers presented at the 13th national meeting of the Association for Computing Machinery. p. 1-4. ACM, 1958.

SISTEMA PARA GERENCIAMENTO E RECUPERAÇÃO DE INFORMAÇÃO

MANAGEMENT SYSTEM AND INFORMATION RETRIEVAL

Renata Maria Abrantes Baracho
 Cátia Rodrigues Barbosa
 Christiano Benvindo dos Santos

Resumo: Atualmente, em diferentes áreas tanto o gerenciamento quanto a recuperação da informação constituem um desafio. São comuns problemas de recuperação da informação quando temos pessoas que dela necessitam em locais distantes de onde a mesma é mantida. Essa pesquisa desenvolve um modelo e um protótipo de organização e recuperação de informação referente a Objetos Arquitetônicos Edificados (OAE). Os sistemas de ambientes informacionais turísticos, bibliotecários, arquivísticos, museológicos, patrimoniais, hoteleiros, governamentais necessitam cada vez mais de sistema de informação eficiente na busca, disseminação e recuperação da informação sobre objeto arquitetônico. No modelo proposto, recursos tecnológicos foram utilizados para facilitar o acesso à informação, garantir sua integridade e permitir que a mesma esteja disponível sempre que possível. Buscou-se facilitar a gerência do produto permitindo que usuários sem grandes conhecimentos de informática pudessem manter o conteúdo atualizado de acordo com as necessidades. O processo de indexação é automático e a estrutura utilizada para seu armazenamento é a de índice invertido em que para cada termo indexado existe uma lista de documentos onde o mesmo ocorre. O modelo de recuperação é o vetorial sendo a ponderação dos termos realizada utilizando-se o método tf-idf e os documentos recuperados são ordenados mediante cálculo de similaridade. O modelo prevê a correção dos termos digitados pelo usuário no momento da consulta buscando aumentar a eficácia da recuperação. Os resultados obtidos são refinados mediante a utilização de filtros os quais ajudam o usuário a encontrar o que deseja de forma mais rápida e direta. Outro ponto importante é o uso das coordenadas dos objetos arquitetônicos edificados (OAE) de forma a localizá-los em um mapa, bem como permitir encontrar outros objetos próximos a ele fazendo inferências a determinadas informações. Tem-se como resultado um protótipo para recuperação e disponibilização de informações referentes aos OAE de forma simples e direta.

Palavras-chave: Recuperação da informação. Arquitetura da informação. Gestão da informação. Sistemas de informação. Organização da informação.

Abstract: In different areas knowledge the recovery and management of information is a challenge. Problems are common with regard to recovery of information when we have people who are in need in locations distant from where the same is maintained. In the proposed model the technology was used to facilitate the access to information, to ensure their integrity and allow it to be available whenever possible. We also sought to facilitate the management of the product allowing users without extensive knowledge of information technology could, without too much effort, keep the content updated in accordance with the needs of the present or planned. The systems of informational environments tourist, librarians, archival, museological, property, hotels, government increasingly need a system of efficient information search and dissemination and retrieval of information about architectural object. The indexing process adopted is automatic and the structure used for storage is the index reversed where for each term indexed there is a list of documents where the same occurs. The recovery model is the vector with the weighting of the terms using the method tf-idf and the recovered documents are sorted by calculation of similarity. The model also provides for the correction of the terms entered by the user at the time of consultation looking to increase the

effectiveness of recovery. The results obtained can still be refined through the use of filters which help the user to find what you want more quickly and directly. Another important point of the research is the use of the co-ordinates of architectural objects built (EOA) in order to locate them on a map, as well as find other objects close to it. In accordance with the above placements we have as a result a prototype for recovery and availability of EAO so simple and direct.

Keywords: Information retrieval. Information architecture. Information management. Information systems. Organization of information.

1 INTRODUÇÃO

Em todas as áreas o gerenciamento e recuperação da informação são um desafio. Um problema recorrente é como encontrar a informação mais adequada à necessidade do usuário. Tendo isso em mente, buscou-se propor uma solução no âmbito da gestão dos Objetos Arquitetônicos Edificados (OAE), termo utilizado na pesquisa para definir casas, igrejas, palácios, palacetes, construções militares, mausoléus, edifícios dentre outras construções, sendo selecionados aqueles que possuam alguma característica histórica ou artística que os destaque.

OAE tombados possuem as características históricas e artísticas supracitadas. “Os tombamentos são realizados com o intuito de preservar bens de valor histórico, cultural, arquitetônico, ambiental e também de valor afetivo para a população, impedindo a destruição e/ou descaracterização de tais bens.” (IPHAN). Suas informações despertam interesse nas comunidades locais e, por vezes, até mesmo internacionais quando tratamos daqueles de grande repercussão, o que justifica o corpus da pesquisa. A proposta é um modelo simples para gerenciar e disseminar a informação acerca dos mesmos.

No Brasil, a gerência pelos bens tombados em âmbito nacional pertence ao Instituto do Patrimônio Histórico e Artístico Nacional (IPHAN) e em âmbito estadual ao Instituto Estadual do Patrimônio Histórico e Artístico (IEPHA). Internamente às organizações supracitadas existem os arquivos referentes ao tombamento de cada bem os quais são ricos em informação de qualidade. No caso do IEPHA, vemos a preocupação referente a essa documentação no art. 11 da portaria nº 29, de 03 de Julho de 2012

O dossiê técnico, mencionado no inciso II do art. 10, deverá ser elaborado por equipe técnica interdisciplinar especialmente formada em função da categoria do bem e que deverá atuar em todas as etapas de pesquisa e redação dos textos de acordo com a metodologia adotada pelo IEPHA/MG.

No entanto nem sempre é simples ter acesso a tal informação que em alguns casos deve ser solicitada e, somente após um certo período, pode ser acessada uma vez que o pedido é enviado a outra cidade, é copiado e enviado de volta. Em outros casos, a consulta é permitida apenas presencialmente, o que dificulta o acesso de uma parcela pública.

Existe informação acerca dos OAE na internet em maior ou menor grau de acordo com a sua visibilidade. As iniciativas para sua divulgação costumam ser em nível local, onde cada região busca valorizar as suas obras para aumentar a visibilidade. Nem todos os OAE possuem informações disponibilizadas via web fazendo com que seja necessário ir fisicamente ao local a procura das mesmas. Considera-se a potencialidade dos recursos computacionais para por meio deles ampliar a disseminação da informação referentes a objetos arquitetônicos. Faz-se um paralelo com o conceito de expor objetos de forma virtual vencendo limitações de tempo e espaço. Essa situação pode ser exemplificada com o conceito de museu uma vez que o público interessado deve ir até eles para conhecer suas peças, no entanto a distância é por vezes impeditiva. De acordo com isso “A implantação do museu virtual pode disponibilizar as principais edificações e objetos em âmbito mundial além de reconstituir os espaços e os objetos para estudos avançados com uso de tecnologias” (BARACHO; BARBOSA, 2011, p. 201).

A tecnologia vem para auxiliar nos processos de forma a permitir que os mesmos tenham um maior alcance.

Não se pode compreender a atualidade, sem confrontar no dia a dia com a importância crescente dos recursos informacionais nas suas várias dimensões. As formas atuais de se produzir, transferir e usar informações, associadas aos padrões tecnológicos cada vez mais sofisticados, traduzem e provocam alterações significativas nos conceitos e práticas das organizações, método de trabalho, parâmetro de disseminação e difusão (CASTRIOTA, 2007, p. 202)

Os dados da pesquisa apontam que diversos sites, que divulgam o patrimônio arquitetônico permitem a busca a partir da navegação entre os itens, entretanto torna-se necessária uma estratégia de busca que vise facilitar o processo permitindo que pessoas, mesmo sem treinamento, consigam encontrar a informação que desejam. Procurando novas soluções para a recuperação da informação dos OAE de forma mais eficiente, tem-se como base conceitos de informação e serviços de informação. Para Tarapanoff (2006, p. 23), “O Objeto da área é a informação e o trabalho com a informação, transformando-a em produtos e serviços de utilidade (com valor agregado) para o cliente usuário”. Borges (2007), em seu artigo, deu as seguintes ideias sobre o que poderia compor um serviço de informação: Processo de auxílio ao usuário na busca de informação ou na satisfação de suas necessidades informacionais, interface direta entre a informação e o usuário e como atividade destinada à identificação, aquisição, processamento e transmissão de informação e ao seu fornecimento em um serviço ou produto de informação. No mesmo artigo, a autora realiza a revisão de alguns trabalhos citando o que seria, de acordo com os mesmos, essencial ao gerenciamento

de um produto de informação: levantar as necessidades de informação dos usuários, estabelecer fontes de informação relevantes, disseminar e divulgar os produtos de informação, definir as tecnologias de informação apropriadas para os produtos e ter uma equipe de profissionais de informação que tenham as competências necessárias para o gerenciamento dos produtos de informação.

Levando em conta esses conceitos, a pesquisa tem o intuito de servir como modelo para criação de um produto de informação, não somente para OAE, mas que também possa ser utilizado em outros domínios.

No modelo proposto a tecnologia é utilizada para facilitar o acesso à informação, garantir sua integridade e permitir que a mesma esteja sempre disponível. Também buscou-se facilitar a gerência do produto permitindo que usuários sem grandes conhecimentos de informática pudessem, sem muito esforço, manter o conteúdo atualizado de acordo com as necessidades.

No intuito de facilitar a busca da informação para atender às necessidades do usuário, o protótipo desenvolvido utilizou-se do modelo de busca vetorial e do cálculo de similaridade. As respostas são ordenadas de acordo com a similaridade entre os documentos e a consulta. Para utilização do modelo, a base deverá ser indexada seguindo o processo descrito na pesquisa.

Outras técnicas utilizadas para facilitar o acesso à informação incluem um corretor ortográfico na entrada da consulta do usuário, exibição de objetos fisicamente próximos ao objeto procurado e formas para refinar a pesquisa através de filtros aplicados sobre os documentos recuperados.

Para facilitar o gerenciamento do produto, foi proposto que o responsável pelo conteúdo não tenha que criar as páginas web diretamente. Ao responsável é computada a responsabilidade de alimentar e manter o sistema com informações pertinentes e atualizadas sem que sejam necessários conhecimentos em linguagens para criação de websites. Desta forma a única preocupação é com a gestão do conteúdo e não do produto em si.

A proposta é a criação de um modelo de gestão e disseminação de informação que se utilize de tecnologias apropriadas para facilitar a gerência e a recuperação da informação sobre Objetos Arquitetônicos Edificados e que possa ser aplicada também em outros domínios com poucas alterações.

2 FUNDAMENTAÇÃO TEÓRICA CONCEITUAL

De acordo com a conceituação apresentada os OAE se aproximam do conceito de monumento, sendo:

A constituição federal brasileira de 1988, no seu artigo 216, dispõe sobre o Patrimônio Cultural nos seguintes termos: constituem patrimônio cultural brasileiro os bens de natureza material e imaterial, tomados individualmente ou em conjunto, portadores de referência à identidade, à ação, à memória dos diferentes grupos formadores da sociedade brasileira, nos quais se incluem: as formas de expressão; os modos de criar, fazer e viver; as criações científicas, artísticas e tecnológicas; as obras, objetos, documentos, edificações e demais espaços destinados às manifestações artístico-culturais; os conjuntos urbanos e sítios de valor histórico, paisagístico, artístico, arqueológico, ecológico e científico.

Torna-se premente um estudo sobre a gestão da informação e em que medida a arquitetura da informação contribui para uma eficaz disseminação e recuperação da informação sobre os objetos arquitetônicos.

Quanto à gestão da informação Davenport (1998, p. 170) diz tratar-se de um conjunto estruturado de atividades que incluem o modo como as empresas obtêm, distribuem e usam a informação e o conhecimento.

Os sistemas de recuperação de informação (SRI) apresentam-se aptos a cumprir o papel de disseminação dos objetos. Nesse contexto, os SRI são a interface entre uma coleção de recursos de informação, em meio impresso ou não, e uma população de usuários. (LANCASTER;WARNER, 1993, p.4-5).

Souza (2006, p. 163) afirma que os SRI trabalham com a organização e viabilização do acesso aos itens de informação desempenhando as atividades de representação e recuperação das informações e documentos de forma a satisfazer a necessidade dos usuários. Para Hiemstra(2000, p. 2) os SRI devem possuir três características: a capacidade de receber termos em linguagem natural sem a necessidade de operadores, ordenação dos resultados de acordo com a relevância e permitir o refinamento dos resultados de forma a auxiliar o usuário.

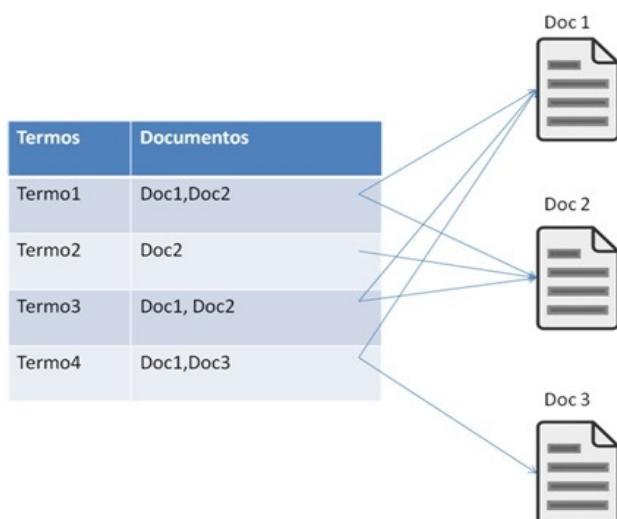
A representação é parte importante nos sistemas de informação e reflete diretamente na recuperação da informação. De acordo com Alvarenga (2003), podemos considerar como representação primária aquela feita por autores a partir da observação dos fenômenos naturais e sociais e representação secundária, que geralmente é utilizada nos sistemas de informação, como pontos de acesso fundamentais na representação dos documentos para futura recuperação.

Como pontos de acesso aos documentos, os metadados são fundamentais para a organização do sistema de informação. O termo surgiu em Dublin durante um simpósio, o qual deu origem à iniciativa Dublin Core. Segundo Alvarenga (2001), pode ser conceituado como “dado sobre dado”; dado que descreve, a essência, atributos e contexto de emergência de um recurso (documento, fonte, etc.) e caracteriza suas relações, visando o acesso e o uso potencial. Ainda segundo Alvarenga (2001), no novo contexto de produção, organização e recuperação de objetos digitais, as metas de trabalho não se restringem à criação de representações simbólicas dos objetos físicos constantes de um acervo, compreendem estabelecimento dos denominados metadados(...) constituindo-se esses em chaves de acesso a serviço dos usuários da internet.

Os metadados constituem a base da indexação e recuperação da informação. Nessa pesquisa a indexação dos termos é realizada de forma automática por meio da extração destes e a estrutura utilizada para armazenamento é a do índice invertido. “Um índice invertido é um mecanismo orientado para a indexação de uma coleção de texto a fim de acelerar a tarefa de busca.” (NAVARRO, 2013, p. 339). "Cada documento no sistema possui um identificador único e a forma de encontrar uma palavra é por via dicionário, o qual é uma lista de palavras que não se repetem e um apontador para a lista dos documentos onde os termos ocorrem." (KOWALSKY; MAYBURY, 1945 p. 82). A FIGURA 1 apresenta essa representação.

Para realizar a indexação várias etapas devem ser seguidas para tratar o texto e dele extrair os termos que irão compor o índice. A FIGURA 2 apresenta as etapas seguindo um fluxo para terminar com a indexação do documento.

FIGURA 1 - Índice invertido



Fonte: Elaborado pelo autor

FIGURA 2 - Processo de indexação



Fonte: Adaptado pelo autor

No pré-processamento, o texto é normalizado. No caso das páginas da internet, por exemplo, as tags html são retiradas, a codificação do arquivo é verificada e normalizada para permitir o correto tratamento de palavras acentuadas por exemplo.

A filtragem e quebra dos termos considera pontuação, dígitos, hifens e diferenciação entre maiúsculas e minúsculas. Para cada um desses itens tem-se um problema peculiar. Fox(1992) demonstra diversos problemas com os quais nos deparamos durante esta tarefa como, por exemplo, quando separar palavras com hifens. Podemos ter o hífen utilizado para unir palavras separadas em final de linha. Os dígitos podem fazer ou não diferença durante a recuperação, não podemos, por exemplo, excluir o numeral 6 do termo B6 (vitamina). Dessa forma vemos que a cada um dos aspectos apresentados cabe uma tratativa que geralmente difere em relação ao objeto tratado pela base estudada. Após a filtragem para a indexação a quebra do texto é realizada passando-se à próxima etapa.

As stopwords podem ser consideradas como palavras com frequência alta entre os documentos. Lancaster (1993, p.232) cita como exemplo os artigos, preposições, conjunções e assemelhados que estão presentes em grande parte dos documentos de qualquer base. E apesar de sua retirada diminuir bastante o tamanho do índice facilitando sua manutenção diversas máquinas de busca não o fazem. O Google é um exemplo, sabemos disso pois há diferença em pesquisar por “os outros” ou apenas “outros”, sendo que no primeiro caso os resultados mais pertinentes são relativos ao filme que leva esse nome.

Por fim é realizada a indexação dos termos que permanecerão após aplicados os filtros do processo. Será necessário armazenar o número de vezes que o termo ocorre na base como um todo e em cada documento da mesma. Caso se trabalhe com um índice completo também será necessário armazenar a posição do termo no documento de forma a permitir a busca levando em consideração a proximidade entre os termos.

Para permitir a recuperação e ordenação dos resultados, optou-se por utilizar o modelo vetorial utilizando o tf-idf (Term Frequency-Inverse Document Frequency) como forma de ponderar os termos.

De acordo com Souza (2006) “nesse modelo, os documentos são tratados como “sacos de palavras” (bags of words) e são representados como vetores no espaço n-dimensional, no qual n é o total de termos índices”. Esse espaço, segundo Salton, Wong e Yang (1975, p.613)

consiste de documentos, que serão identificados por um ou mais termos de indexação sendo que os termos podem ser ponderados de acordo com sua importância.

No modelo vetorial, a pesquisa é feita a partir de texto livre na interface de busca, sem a utilização de conectores como na busca booleana (Manning, Raghavan, Schutze 2009, p.107). Esse estilo, segundo os autores, é muito comum em sistemas web devido em parte a facilidade no uso, bastando apenas que os usuários entrem com uma lista de palavras.

Quanto a ponderação dos termos, podemos dizer que existem diversos métodos que podem ser utilizados, sendo um dos mais utilizados o tf-idf (Term Frequency * Inverse Document Frequency).

A forma mais simples de ponderação é a tf que considera a freqüência do termo no documento, ou seja, quanto maior o número de vezes que determinado termo aparecer em um documento, maior será a relevância do documento para um usuário que pesquisar por aquele termo. O problema de se utilizar apenas essa medida conforme descrito por Lancaster (1993, p.232 – 233) está no fato de que “algumas palavras que ocorrem freqüentemente num documento podem não ser bons discriminantes – que sirvam para diferenciar este documento de outros da base de dados – porque também ocorrem com freqüência na base de dados como um todo”.

Sendo assim, para complementar foi necessário encontrar um fator que contrabalançasse o tf. O idf cumpre esse papel uma vez que varia inversamente ao número de documento que contém determinado termo. Um fator idf típico pode ser calculado como $\log(N/n)$, onde N é o número de documentos na coleção e n a freqüência do termo nesta mesma coleção. (SALTON,BUCKLEY 1988, p.516). A freqüência com que uma palavra ocorre na base de dados como um todo é ainda mais importante que a freqüência da palavra nos documentos. (LANCASTER, 1993, p.233). Ainda de acordo com o autor as palavras que aparecem pouco em uma coleção são boas discriminantes para um documento ou no máximo um pequeno grupo, com isso é possível afirmar que podemos aumentar a precisão da recuperação ao utilizar esse fator.

Os pesos tf e idf apresentam comportamento de lei de potência que se equilibram. Valores altos de tf tendem a estar associados a baixos valores idf e vice-versa. Como consequência os maiores valores de tf-idf são alcançados por valores de idf intermediários. Assim tanto termos que aparecem com freqüência muito alta nos documentos, como as stopwords, ou termos raros como palavras estrangeiras ou erros de digitação não ficarão no topo com o maior valor uma vez que não serão de grande valia para o ranqueamento, o que comprova a eficácia do método.

Com o índice pronto o próximo passo é referente a busca onde o usuário poderá realizar sua consulta utilizando linguagem natural e sem necessidade de conectores entre os termos. Os que forem utilizados na pesquisa irão passar por um filtro como aqueles dos documentos indexados. Os que forem utilizados receberão um peso e comporão o vetor da consulta. Esse vetor será comparado com os vetores dos documentos presentes na base e o documento que apresentar a menor distância entre seu vetor e o da consulta será o primeiro a ser devolvido e o mais distante o último. Essa distância é medida pelo valor obtido a partir do cálculo da similaridade entre os vetores, sendo realizado a partir da seguinte equação.

$$\text{similarity}(Q, D) = \frac{\sum_{i=1}^t (w_{iq} * w_{ij})}{\sqrt{\sum_{i=1}^t (w_{iq})^2 * \sum_{i=1}^t (w_{ij})^2}}$$

onde

$$w_{iq} = \left(0.5 + \frac{0.5 freq_{iq}}{maxfreq_q} \right) * IDF_i$$

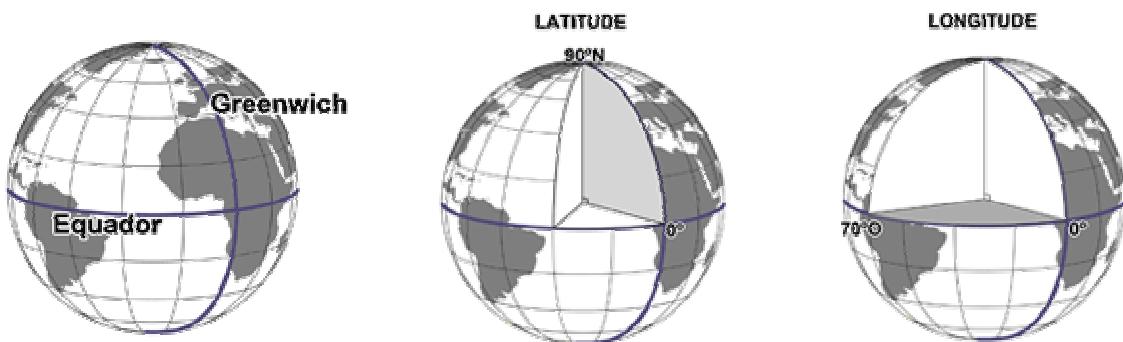
e

$$w_{ij} = \frac{freq_{ij} * IDF}{\sqrt{\sum_{vector} (freq_{ij} * IDF)^2}}$$

Outro ponto importante da pesquisa é a geolocalização dos OAE bem como o cálculo da distância entre eles. Para localizar um ponto na terra é necessário sabermos sua latitude e longitude. Latitude é a distância ao Equador medida ao longo do meridiano de Greenwich. Esta distância mede-se em graus, podendo variar entre 0° e 90° para Norte(N) ou para Sul(S). Longitude é a distância ao meridiano de Greenwich medida ao longo do Equador. Esta distância mede-se em graus, podendo variar entre 0° e 180° para Leste(E) ou para Oeste(W). A FIGURA 3 ilustra o que foi explicado.

Ou seja, dada uma latitude e longitude é possível identificar qualquer ponto no planeta.

FIGURA 3 - Latitude e Longitude



Fonte: Laboratório de Cartografia da UFSM

Uma das formas existentes para o cálculo da distância entre dois pontos é a partir da fórmula de Haversine. Apesar desta fórmula considerar a terra como uma esfera perfeita a aproximação do resultado é muito bom, sofrendo pequena degradação apenas em distâncias muito grandes.

Abaixo vemos a fórmula de Harversine

$$d = 2r * \arcsin \left(\sqrt{\sin^2\left(\frac{\theta_2 - \theta_1}{2}\right) + \cos(\theta_1) \cos(\theta_2) \sin^2\left(\frac{\theta_1 - \theta_2}{2}\right)} \right)$$

d é a distância entre dois pontos com latitude e longitude (θ , φ) e r é o raio estimado da terra tendo sido considerado como 6371 km.

3 DESENVOLVIMENTO

A pesquisa foi desenvolvida em etapas entre seleção de fontes, aquisição, organização de informação e desenvolvimento do protótipo.

Segundo Sapiro (1993), uma escolha acertada das fontes de dados é crítica para a coleta e a classificação das informações. A primeira fase da pesquisa foi o levantamento das informações acerca dos OAE selecionados. Foi realizada nos institutos responsáveis pelas informações e em publicações oficiais. A documentação que se encontrava em meio digital foi copiada, e a que estava em papel foi fotografada para posterior pesquisa. Como os OAE escolhidos estão em Belo Horizonte, mesma cidade em que se encontram os escritórios dos institutos responsáveis pelas informações foi possível o acesso direto às mesmas. Somente para alguns OAE como aqueles que fazem parte do complexo da Pampulha foi necessário solicitar a documentação de outro escritório que não o local.

Além da documentação recolhida nos institutos foram tiradas fotografias de cada um dos OAE estudados para fim de documentação de seu estado atual bem como para ilustração

dos mesmos no protótipo desenvolvido. Para capturar as imagens foi utilizada uma câmera Nikon D3100 na sua maior resolução e com o método automático ligado de forma a ter sempre uma imagem de qualidade e com certo padrão entre si. Não foi utilizado nenhum tipo de filtro ou efeito uma vez que apesar da imagem ser utilizada de forma a valorizar o objeto frente aos usuários ela também será utilizada como documento arquitetônico. Vemos um exemplo na FIGURA 4, a qual faz parte do acervo criado para a pesquisa.

FIGURA 4 - Instituto de Educação



Fonte: Foto tirada pelo próprio autor

Para as fotos adotou-se a licença da Creative Commons do tipo Atribuição-NãoComercial-CompartilhaIgual CC BY-NC-SA. Segundo o próprio site essa licença permite que outros remixem, adaptem e criem a partir do seu trabalho para fins não comerciais, desde que atribuam ao autor o devido crédito e que licenciem as novas criações sob termos idênticos.

De posse da documentação coletada foram selecionadas as informações com o intuito de demonstrar a funcionalidade do modelo e protótipo. As informações coletadas não foram disponibilizadas na íntegra. “A variedade de informação deve ser administrada de modo que as informações coletadas reflitam a complexidade do ambiente sem sobrecarregar os usuários com excesso de informação” (CHOO 2006, p.419).

Os metadados principais utilizados foram baseados nas fichas do SICG M301 e M302. SICG - Sistema Integrado de Conhecimento e Gestão, essa é uma iniciativa do IPHAN para gerir informação acerca do patrimônio material cujo processo vem desde 2006. Até o momento, o sistema era composto apenas de fichas em Word e Excel, contudo no inicio de 2014 foi marcado o primeiro treinamento do sistema sendo que o mesmo pode ser visto, sem os dados oficiais inseridos, em: <http://portal.iphan.gov.br/sicg/pesquisarBem>.

Com base nas fichas foram elencados os seguintes metadados para fazerem parte do cadastro como informações principais a serem extraídas da documentação coletada.

Título: Nome oficial do Objeto Arquitetônico Edificado.

Título Alternativo: Outros nomes pelos quais o OAE seja reconhecido.

Descrição: Descrição histórica e ou arquitetônica acerca do objeto. Foi retirada dos documentos coletados nos institutos e do livro de Bens tombados do IEPHA-MG.

Localização: UF / Município / Logradouro / Número / Complemento / Código Postal: Os dados referentes ao estado e município foram cadastrados no sistema para que o usuário realize apenas a escolha em uma lista. Os dados foram extraídos do site do IBGE para poder compor a listagem no protótipo. As demais informações deverão ser informadas pelo usuário.

Coordenadas Geográficas: As coordenadas são consideradas no padrão decimal, utilizando ponto (.). Exemplo: -19.865960, -43.962208, como as duas coordenadas estão negativas significa que o local indicado fica ao sul do equador e a oeste de Greenwich.

Data construção: Data em que o objeto arquitetônico foi construído.

Uso Original: Qual foi a primeira utilização do bem. Como exemplo podemos citar o Cassino da Pampulha

Uso Atual: Qual a utilização atual. Seguindo o exemplo anterior o Cassino hoje é utilizado como Museu de Arte.

Tipologia: (Religiosa, civil, oficial, militar, industrial, outra)

Estilo Arquitetônico: O estilo de cada obra foi retirado da documentação, mas antes da inserção cada estilo foi verificado em um thesauro de arquitetura e livro específico acerca de estilos arquitetônicos. Os estilos são inseridos sob demanda, ou seja, um estilo só é inserido quando necessário. O thesauro utilizado foi o do getty o qual trata justamente de arquitetura.

Pessoas: Indicar as pessoas identificadas que trabalharam na construção do OEA, tais como engenheiros, arquitetos, paisagistas e quaisquer outros que tiveram relação com a obra. Algumas informações foram pré-cadastradas no sistema para facilitar o cadastro e evitar duplicidades e erros no momento do cadastro. Dessa forma busca-se padronizar a informação evitando o uso de nomes diferentes para o mesmo autor. Além de provocar dúvidas ao usuário iria atrapalhar ao realizar um agrupamento por pessoas. Teríamos um agrupamento com "Aleijadinho", por exemplo, e outro com "Antônio Francisco da Costa Lisboa" sendo que deveria ser apenas um. Para esta pesquisa foi utilizado a própria documentação referente aos tombamentos. Em alguns casos foi utilizado o serviço "Union List of Artists Names" o qual pode ser acessado em <http://www.getty.edu/research/tools/vocabularies/ulan/>. O serviço permite a pesquisa pelo nome e retorna o nome preferencial e outras variações do mesmo. È

um thesaurus on-line muito útil para pesquisa, no mesmo site é possível encontrar ainda informações referentes a material utilizado em construções bem como outras informações referentes à arquitetura. O mesmo está em inglês e existe uma versão em espanhol, até o momento parece que nenhuma parceria foi feita para realizar a tradução do mesmo para o português.

Os dados necessários para o cadastro das imagens são:

Título: Título da obra, necessários para casos como o de complexos arquitetônicos. Como exemplo, podemos citar o complexo da Pampulha o qual terá fotos de diversos objetos tais como a Igreja de São Francisco de Assis e do Museu de Artes, nesse caso é necessário identificar qual imagem pertence a qual objeto.

Autor Imagem: Quando possível identificar o autor de determinada imagem

Tipo da licença: Informar se a imagem pode ou não ser reproduzida e sobre quais termos, todas as fotos que foram tiradas para a pesquisa poderão ser utilizadas por terceiros desde que a fonte seja identificada.

Origem: Informar se a foto foi tirada para compor a documentação ou sua fonte.

Por referência a indexação. Esse passo será feito conforme descrito no referencial teórico.

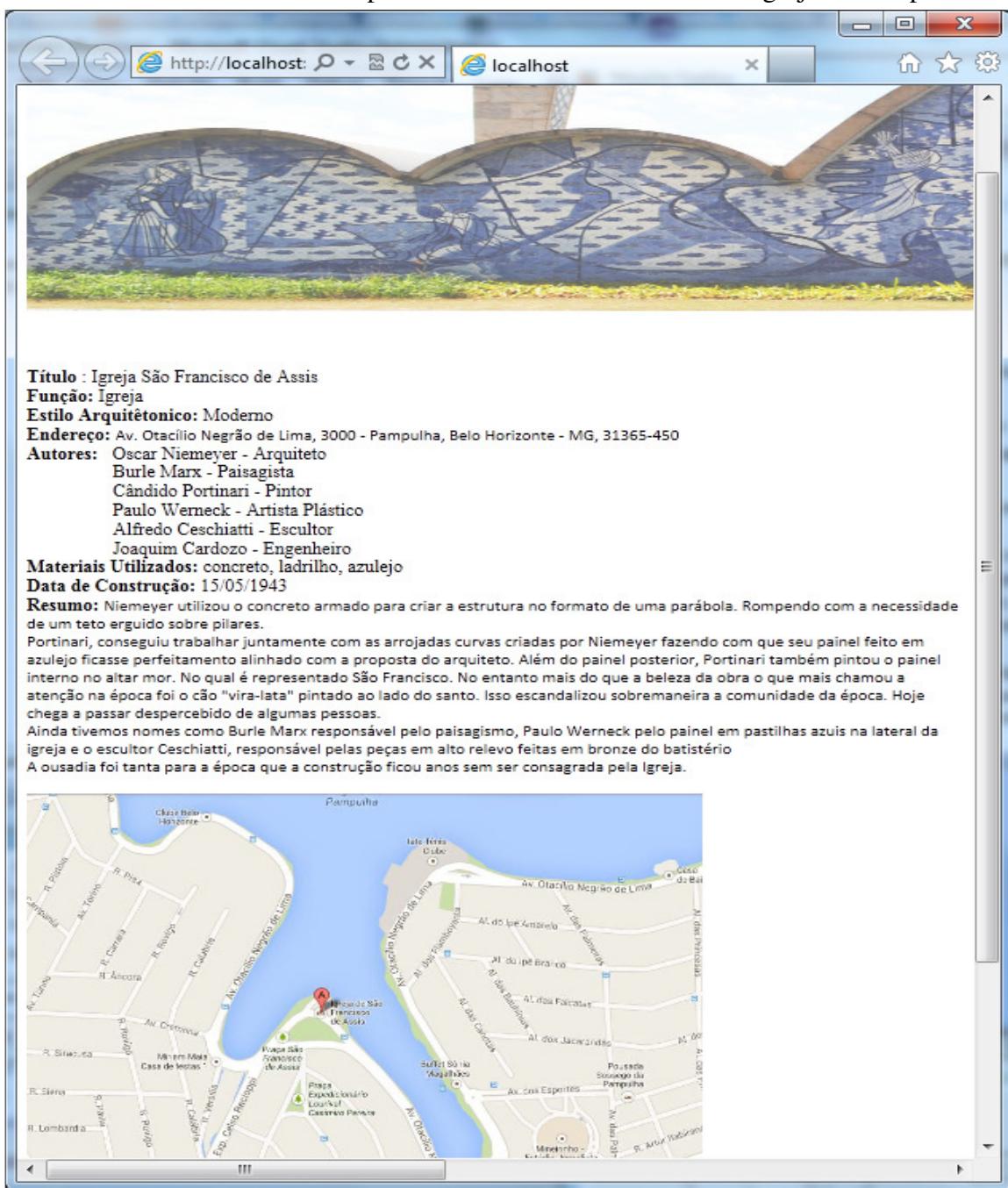
A FIGURA 5 apresenta uma tela do sistema com os metadados propostos de um dos objetos arquitetônicos cadastrados.

Após a definição dos metadados e a indexação, a base está preparada para a recuperação da informação. Para recuperar a informação desejada o usuário poderá entrar com um texto em linguagem natural sem necessidade da utilização de termos específicos ou prévios conhecimentos em lógica de conjuntos como acontece na busca booleana.

A consulta passa pelos mesmos processos dos documentos antes da indexação. A diferença é que ao final destes os termos são verificados frente a um dicionário da língua portuguesa o qual é incrementado com palavras presentes na base e que por ventura ainda não fazem parte do mesmo. Isso pode ocorrer com nomes próprios por exemplo.

Caso o termo não possua uma ocorrência no dicionário, sofrerá mutações, procurando encontrar outros para os quais sejam necessárias até duas alterações. Se com até duas alterações, seja de inclusão, alteração ou exclusão de letras se consiga chegar a um exemplar válido, este será considerado como sugestão. Ao final é retornado uma lista de palavras que atendem ao requisito sendo que a escolhida é aquela com maior número de aparições na base de dados. Ao retornar o resultado será informado ao usuário quais termos foram utilizados para evitar quaisquer enganos.

FIGURA 5 – Tela do Protótipo com os metadados referentes à Igreja da Pampulha



Fonte: Foto tirada pelo próprio autor

Com o intuito de aproximar o protótipo ao que foi descrito nas palavras de Hiemstra durante o referencial, foi criada uma forma para o refinamento dos resultados por parte do usuário. Eles podem ser refinados a partir de filtros pré definidos, que serão exibidos ao usuário e o ajudarão na tarefa de restringir os resultados facilitando assim no processo da recuperação da informação de forma mais precisa e menos trabalhosa.

A estrutura dos documentos inseridos na base é normalizada para todos os OAE, facilitando assim a identificação de características comuns entre os mesmos. Consideremos

uma coleção composta por diversos documentos que compartilham a mesma estrutura. Essa estrutura além de título e descrição contem algumas características específicas, características que são comuns a todos os documentos da base. Em alguns documentos o conteúdo dessas características são semelhantes e em outros não, fazendo com que um filtro aplicado ao resultado ajude o usuário a encontrar de forma mais rápida o que deseja.

Por exemplo, se o usuário digitar igreja, ele poderá, hipoteticamente, receber dois resultados. Um referente à igreja de São Francisco de Assis e outra referente à igreja da Boa Viagem. Os filtros então serão carregadas com informações alusivas a essas duas igrejas. Sendo:

Cidade: Belo Horizonte (2)

Pessoa: Oscar Niemeyer (1), João Morandi (1)

Estilo: Moderno (1), Eclético (1)

No filtro Pessoa, teríamos mais resultados no caso real, aqui foi reduzido apenas como efeito didático. Os números logo após o termo indicam em quantos documentos os termos estão presentes. Seguindo o exemplo, os dois resultados pertencem a “Belo Horizonte” mas apenas um tem o estilo Eclético e apenas em um consta que João Morandi trabalhou na obra. Se o usuário selecionar qualquer um desses termos os resultados serão filtrados passando a exibir apenas aquele OAE referente à Igreja da Boa Viagem.

Essa funcionalidade utiliza o conceito de interseção entre os conjuntos, sem que o usuário precise ter domínio do mesmo. Dessa forma sempre que um usuário clicar em um termo ele estará restringindo o resultado da consulta textual que ele efetuou. Ao digitar o termo “igreja” e depois dos resultados visualizados selecionar o termo “João Morandi” no filtro Pessoa é compatível a dizer que o mesmo quer ter acesso aos OAE referentes à igrejas onde João Morandi tenha trabalhado.

O protótipo se baseia nas linguagens asp.net MVC e C# com banco de dados MySQL e servidor IIS7. O asp.net é o padrão da Microsoft para o desenvolvimento de páginas web enquanto que o c# é utilizado para o desenvolvimento de rotinas e regras de negócio. O banco de dados MySQL foi escolhido por ser estável, prova são as diversas aplicações que utilizam-no na web, e gratuito. IIS7 (Internet Information Services) é um servidor de web que pode hospedar páginas, streaming, aplicativos web. Para que um site, por exemplo, funcione é necessário que o mesmo esteja hospedado em um servidor web, seja ele apache, IIS ou outro qualquer.

Um dos pontos colocados como objetivo para o desenvolvimento do protótipo foi o fato de que não deveria ser necessário conhecimento em desenvolvimento web por parte do

usuário. O que gerou um problema, devido à forma como os sites são desenvolvidos. Um website comum é composto por páginas as quais estão presentes no servidor web. O usuário faz uma requisição ao servidor e esse por sua vez devolve uma página ou algum outro recurso como resposta.

Aqui faz-se necessário entender o que é Uniform Resource Locator (url) ou como também é chamado “endereço da página”, a url é o caminho que indica onde está o recurso que o usuário deseja. Como exemplo de url considera-se: <HTTP://exemplo.com/software/windows.html>. Se quebrarmos o nosso exemplo em partes teremos:

HTTP – Protocolo de transferência de hipertexto

<Exemplo.com> – endereço do nosso site ou domínio. Podemos ter o site exemplo com a extensão .com, .com.br, org, .edu. Mas não faz parte dessa pesquisa o aprofundamento nesse ponto.

Software – iremos considerar nessa pesquisa que software é uma pasta em nosso servidor e que essa pasta está na raiz de nosso site exemplo. Dentro dela podemos ter diversas páginas relacionadas com software.

<Windows.html> – essa seria a página com informações referentes a Windows. Dentro da pasta software poderíamos ter diversas outras páginas que resultariam em endereços como <HTTP://exemplo.com/software/linux.html> por exemplo.

Isso demonstra é que a estrutura dos sites se parecem com a estrutura de arquivos que temos em nossas máquinas. Onde temos um diretório raiz com diversos outros diretórios ou pastas dentro desse e em cada um diversos arquivos.

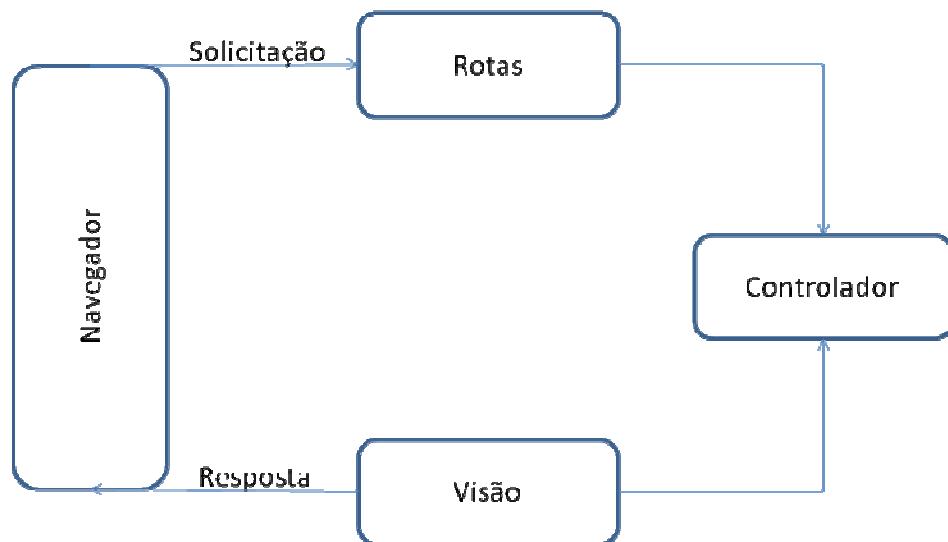
Para que o usuário não tenha que criar páginas e nem inserir arquivos diretamente no servidor foi preciso encontrar uma forma para que o endereço não aponte diretamente para um arquivo. Parte da solução foi dada com a utilização de rotas e páginas dinâmicas. Rota nesse contexto é uma forma que nos possibilita mapear um endereço (uma url) para um controlador/ação e qualquer parâmetro adicional bem como construir essa url.

No protótipo ao invés de um endereço mapear um arquivo ou diretório mapeia uma ação e seus parâmetros. Considerando uma parte do exemplo anterior temos <http://exemplo.com> (domínio) / software (controle / ação) o termo software deixa de representar um diretório e passa a ser um controlador que pode receber como parâmetro uma string por exemplo, no exemplo esse parâmetro é Windows (agora sem o .html) pois não representa um arquivo. A url ficou então <http://exemplo.com/software/windows>. Ao receber o parâmetro "Windows" o controlador irá solicitar as informações referentes a esse software e receberá,

caso o mesmo exista na base de dados, as informações necessárias para preencher a página que o usuário visualizará.

A FIGURA 5 mostra de forma simplificada o processo explicado. É feita uma solicitação, é verificado a partir das definições de rotas qual o controlador/ação será chamado. Esse controlador sabe quais as operações serão executadas e que informações devem ser passadas para a camada de visão de forma a enviar uma resposta ao usuário.

FIGURA 6 - Modelo utilizado no protótipo (resumido)



Fonte: elaborado pelo autor

A visão retornada ao usuário é um modelo que será preenchido com as informações do software que ele procura, ou com as informações do OAE que ele deseja visualizar.

As rotas são tratadas com cuidado, da mesma forma que ao mudar uma pasta de lugar, ao alterar uma rota endereços antigos que o usuário considerava válidos deixarão de funcionar tornando os usuários insatisfeitos.

Considera-se, para simplificar que o endereço do site seja <http://pesquisa.com.br>, o endereço não é válido e não pretende-se adquirir esse domínio. Foi escolhido de forma aleatória visando apenas facilitar o entendimento. Para esse domínio cria-se as rotas, conforme TABELA 1.

TABELA 1 - rotas do site

	URL	Descrição
1	http://pesquisa.com.br/	Página inicial do protótipo com a tela de busca
2	<a href="http://pesquisa.com.br/oaе/<param>">http://pesquisa.com.br/oaе/<param>	“param” será o título do objeto arquitetônico
3	<a href="http://pesquisa.com.br/cad/<param>">http://pesquisa.com.br/cad/<param>	“param” será o cadastro a ser tratado

Fonte: Elaborado pelo autor

Um exemplo utilizando a rota 2 seria <http://pesquisa.com.br/oaе/palacetedantas> assim que o servidor web receber essa requisição ela será enviada para um controlador que entendera “palacetedantas” como um parâmetro e realizará as tratativas necessárias para devolver as informações necessárias para que um modelo seja preenchido e devolvido ao usuário. E é ai que está a beleza, não existe amarração com arquivos e sim processos.

O padrão de arquitetura utilizado no sistema foi o MVC (Model-View-Controller). Padrão consolidado e aceito no desenvolvimento de sistemas. Baseando o aplicativo nessa arquitetura temos três componentes principais: modelo, visão e controle. Sendo que cada um é responsável por determinadas tarefas sem a necessidade de conhecer como as demais são executadas nos outros componentes. Apesar de parecer não é o mesmo paradigma do desenvolvimento em camadas.

Os três componentes do MVC da seguinte forma:

Modelo: As classes que representam a regra de negócios e o acesso aos dados da base de dados. Muitas vezes encapsula os dados que estão armazenados na base bem como a lógica para manipulá-los.

Visão: É responsável por gerar a interface que será apresentada ao usuário. Isso é realizado recebendo um objeto de modelo a partir do controle renderizando isso em html, ou outro tipo de conteúdo, para ser apresentado ao usuário.

Controlador: Essa classe é responsável pelo relacionamento entre a visão e o modelo, bem como responder às entradas do usuário. Além disso, a decisão quanto a rotas e chamadas a visões para atender a demanda do usuário é realizada neste componente.

A idéia é facilitar a manutenção uma vez que cada um dos componentes é responsável por uma tarefa. É mais fácil dividir essas tarefas de forma a serem tratadas de forma separada por pessoas ou equipes diferentes dentro do projeto.

4 CONCLUSÃO

Com o desenvolvimento da pesquisa procurou-se atender aos objetivos de permitir o acesso da informação ao maior número de usuários independentemente da distância do OAE, e isso foi realizado mediante a utilização de plataforma web.

Outro ponto importante é a tentativa de desvincular a necessidade de um profissional de tecnologia para a inserção de conteúdo na base. Dessa forma o usuário tem maior independência para realizar essa tarefa. As informações entram no sistema mediante cadastro simples e as páginas que são disponibilizadas ao usuários externos tem um suporte visual mais elaborado sendo preenchidas a partir de um modelo com as informações disponibilizadas no cadastro.

Buscou-se facilitar a recuperação da informação a partir de técnicas como correção ortográfica dos termos da consulta baseando-se em dicionário da língua portuguesa e termos presentes na base, refinamento da pesquisa a partir de filtros, ordenação dos resultados baseando-se no cálculo de similaridades.

De acordo com as colocações acima temos como resultado um protótipo para recuperação e disponibilização de objetos arquitetônicos edificados. Desenvolvido com técnicas de representação e indexação para recuperação. Apresenta recursos de filtros e facilidade de uso para melhorar o processo de mediação entre o usuário e o sistema.

Esse fatores são essenciais para ampla divulgação da informação, possibilitando de amplo acesso e insumos para gestão da informação.

REFERÊNCIAS

ALVARENGA, Lidia. A Teoria do Conceito Revisitada em Conexão com Ontologias e Metadados no Contexto das Bibliotecas Tradicionais e Digitais. **DataGramZero - Revista de Ciência da Informação**, v.2, n.6, dez/01. Disponível em: <http://www.dgz.org.br/dez01/Art_05.htm>. Acessado em 27/08/2013.

ALVARENGA, Lídia. Representação do conhecimento na perspectiva da ciência da informação em tempo e espaço digitais. **Encontros Bibli: Revista Eletrônica de Biblioteconomia e Ciência da Informação**, Florianópolis, n. 15, 1º sem. 2003.

BARACHO, Renata Abrantes; BARBOSA, Cátia Rodrigues. O objeto museal em diferentes contextos e mídias. **Em Questão**, Porto Alegre, v. 17, n. 2, p. 195-208, jul./dez. 2011.

BORGES, Mônica Erichsen Nassif. O essencial para a gestão de serviços e produtos de informação. **Revista Digital de Biblioteconomia e Ciência da Informação**, Campinas, v.5, n. 1, p. 115-128, jul/dez. 2007

BRASIL. Constituição (1988). **Constituição da República Federativa do Brasil**. Brasília, DF, Senado, 1998.

CASTRIOTA, L. B. ; ANGELO, C. V. S. . Tecnologia digital e acessibilidade: a Rede Latino-americana de Acervos de Arquitetura e Urbanismo (RELARQ). In: XI Congreso de Sociedad Iberoamericana de Gráfica Digital (SIGraDi), 2007, Mexico. Anais do XI Congreso de Sociedad Iberoamericana de Gráfica Digital (SIGraDi). Mexico : SIGRADI / Universidad La Salle, 2007. p. 200-204. acessado em 05/2013 em: site

CHOO, Chun Wei. **A organização do conhecimento: como as organizações usam a informação para criar significado, construir conhecimento e tomar decisões.** tradução Eliana Rocha. 2. ed - São Paulo: Editora Senac São Paulo, 2006.

DAVENPORT, Thomas H. **Ecologia da informação: por que só a tecnologia não basta para o sucesso na era da informação.** Laurence Prusak; tradução Bernadette Siqueira Abrão. — São Paulo : Futura, 1998.

FOX, Christopher. Lexical analysis and stoplists. In **Information Retrieval: Data Structures & Algorithms.** edited by William B. Frakes and Ricardo Baeza-Yates. Prentice Hall; 1992.

Grafisch Centrum Twente, 2000. Originalmente apresentada como tese de doutorado, Centre for Telematics and Information Technology, 2000.

HIEMSTRA, Djoerd. **Using Language Models for Information Retrieval.** Enschede:

KOWALSKI, Gerald J. MAYBURY, Mark T. **Information Storage and Retrieval Systems.** Springer 2000.

LANCASTER, F. Wilfred. WARNER, Amy J. **Information Retrieval Today.** Information Resources Press. United States of America. 1993.

MANING, Christopher D., RAGHAVAN, Prabakar; SHUTZE, Hinrich. **Introduction to Information Retrieval.** Cambridge University Press 2009.

NAVARRO, Gonzalo. Indexação e Busca. In: BAEZA-YATES, Ricardo; RIBEIRO-NETO, Berthier. Recuperação de Informação: **Conceitos e tecnologia das máquinas de busca.** 2 ed. Porto Alegre : Bookman, 2013.

SALTON, G. WONG, A. YANG,C.S. A vector space for automatic indexing. **Communications of the ACM**, v. 18, n. 11, november 1975.

SALTON, Gerald. BUCKLEY, Christopher. Term-weighting approaches in automatic text retrieval. **Information Processing & Management**, v. 24, n. 5, 1988 Printed in Great Britain.

SOUZA, Renato Rocha. Sistemas de recuperação de informações e mecanismos de busca na Web: panorama atual e tendências. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 11, n. 2, p. 161-173, maio/ago. 2006.

TARAPANOFF, Kira (Org.). **Inteligência, Informação e Conhecimento.** Brasília: IBICT, UNESCO, 2006.

BIG DATA: UMA INVESTIGAÇÃO COM USO DE DADOS ABERTOS SOBRE ACIDENTES DE TRABALHO

BIG DATA: AN INVESTIGATION WITH OPEN DATA ON ACCIDENTS AT WORK

Cláudio José Silva Ribeiro

Resumo: O tema Big Data tem despertado interesse nos profissionais que trabalham com a Gestão da Informação, pois trata de insumo essencial no processo de criação do conhecimento. Este relato apresenta o tema e explora os fundamentos que auxiliam no entendimento da abordagem de Big Data. Discute a explosão informacional e a avalanche de dados, chegando aos elementos que compõem o tema. Aborda os 4 Vs do Big Data e as fases de *Discovery*, *Data Preparation*, *Model Planning* e *Analytics*. O processo de pesquisa se baseia em um estudo exploratório desenvolvido por meio de pesquisa bibliográfica com análise documental. O marco empírico para análise dos elementos do domínio foi baseado no site de Dados Abertos da Dataprev, bem como com outras informações disponíveis em redes sociais e blogs especializados. Apresenta os resultados preliminares sobre estudos nas fases iniciais de projetos de Big Data, de forma a viabilizar alternativas para representação e disseminação dos grandes volumes de dados presentes na Internet. Ao final, apresenta alguns aspectos ligados ao perfil do profissional que está participando destes projetos.

Palavras-chave: Big Data. Gestão da Informação. Análise de Dados. Dados Abertos. Profissional de Informação

Abstract: The theme of Big Data has been rising interest in professionals working with information management as it's an important factor for knowledge creation process. This article presents this theme and explores the fundamentals which help in understanding Big Data approach. Also discusses the information explosion and data deluge, reaching elements that make up the theme. Moreover, it covers the 4 Vs of Big Data and the phases of Discovery, Data Preparation, Model Planning and Analytics. The research process uses a exploratory study with literature search and documental review. The empirical milestone for analysis of elements of the domain was based on the Open Data Dataprev website as well as other information available on social networks and specialized blogs. The paper presents preliminary results about studies in early stages of Big Data projects in order to make possible alternatives for the representation and distribution of huge volumes of data present on the Internet. At it's end, this paper points out some aspects related to the professional who takes part in these projects.

Keywords: Big Data. Information Management. Data Analysis. Open Data. Information Professional.

1 INTRODUÇÃO

O uso de dados e informação sempre foi objeto de estudos para a área de Ciência da Informação. Desde a década de 1940, quando, principalmente em decorrência da 2^a. Guerra Mundial, novas pesquisas que foram impulsionadas pela indústria bélica geraram o “Caos Documental” ou ainda a “Explosão Informacional” (RIBEIRO, 2008), os profissionais envolvidos com pesquisas na área de CI sempre estiveram presentes em trabalhos ligados à temática (SARACEVIC, 1996a).

Aliado a isso, percebe-se um outro deslocamento nos estudos sobre Gestão de Dados e Informação. O uso cada vez maior de dispositivos móveis tem contribuído para incrementar a “avalanche de dados”, uma vez que celulares e tablets que atuam como geradores e consumidores de dados e informação estão presentes no nosso cotidiano. Esse movimento tem sido conhecido como Big Data e vem despertando o interesse por todas as pessoas que tem algum envolvimento com atividades para Gestão da Informação¹¹⁰.

Outro tema que tem motivado debates e investigações é a possibilidade de fazer análise deste vasto volume de dados. Um exemplo recente sobre o uso dessa abordagem vem do esporte. Em reportagem veiculada por meios de comunicação sobre a seleção da Alemanha, vencedora da Copa do Mundo no Brasil, houve registro sobre a utilização de solução que permitisse a análise de dados sobre jogos, jogadores, táticas de jogo e a possibilidade de avaliar comentários especializados. Ou seja, o uso de dados e informação em quantidade como forma de melhorar o desempenho do time alemão (SAP, 2014).

Pode-se afirmar que não está se propondo algo totalmente novo com o tema Big Data, pois o uso de informação para a obtenção de resultados não é uma coisa nova. As abordagens exploradas nos últimos anos pela Ciência da Informação vêm contemplando diferentes visões em Sistemas de Apoio à Decisão (EIS), Armazéns de Dados (Data Warehouses e Data Marts), Desempenho dos Negócios (Business Intelligence), soluções para Mineração de Dados (Data Mining), além de informação para planejamento estratégico, gestão de recursos informacionais e ativos de informação na Web. Fazendo uma análise preliminar destes movimentos, é lícito supor que o profissional de informação deve refletir um pouco sobre como poderá se envolver nas discussões sobre o tema Big Data (RIBEIRO, 2014).

Dentro deste contexto, este relato apresenta alguns resultados preliminares sobre o envolvimento de profissionais da informação nas fases iniciais de projetos de Big Data, de forma a viabilizar alternativas para representação e disseminação dos grandes volumes de dados presentes na Internet.

2 A AVALANCHE DE DADOS COMO CATALISADOR DE BIG DATA

O tratamento e uso da informação pela sociedade têm se modificado nas últimas décadas como consequência do surgimento de novos modelos sociais, econômicos ou tecnológicos. A crescente utilização de meios de comunicação com alto grau de mobilidade e o uso cada vez maior da Internet, definem outros espaços e demarcam novas fronteiras para a

¹¹⁰ O relato aqui apresentado é parte do projeto de pesquisa do autor e está apoiada no Referencial Teórico desenvolvido em (RIBEIRO, 2014)

sociedade contemporânea (RIBEIRO, 2008). Ficou evidenciado o estado de insatisfação dos usuários nas pesquisas desenvolvidas por Wurman (2005) - onde reafirmou-se um sentimento de “Ansiedade da Informação” – e por Gopinath e Das (1997) – onde estabeleceu-se uma percepção sobre a “Explosão Informacional”.

Heath e Bizer (2011) reforçam essa ideia quando observam que na atualidade estamos cercados por uma grande quantidade de dados e informação. São registros sobre o cotidiano – desempenho da educação, produção de bens e serviços, investimentos e impostos governamentais, estatísticas sobre a economia e dados sobre o consumo - que nos auxiliam a tomar decisões e gerar conhecimento.

Mas afinal, o que está impulsionando esta avalanche?

Alvin Tofler, em seu livro publicado em 1984 sobre as mudanças do comportamento da sociedade pós-moderna, cunhou o termo prosumer¹¹¹ (TOFLER, 1984), para definir o novo perfil de interação da sociedade de consumo. É possível afirmar que com o avanço do uso de dispositivos móveis, esta forma de interação começou a se concretizar. O uso de sensores industriais e biomédicos, fotos, vídeos, e-mails, redes sociais, além do comércio eletrônico, interações via call centers, dados públicos imagens médicas e outros dados científicos, câmeras para monitoramento, medidores inteligentes, GPS, aplicativos para troca de mensagens, aplicações que nos ajudam a pegar táxis, outras que nos ajudam na locomoção urbana evitando engarrafamentos, ou ainda no monitoramento de ônibus e até de aviões, são exemplos claros dos efeitos desta avalanche (RIBEIRO, 2014).

Complementarmente, é possível perceber também uma mudança no funcionamento das aplicações de comércio eletrônico. A ampliação do uso de sistemas de recomendação¹¹² na Web, permite que sejam indicados dezenas de opções de compras aos clientes usuários destes serviços. Aliado a isto, tem-se que a previsão da expansão das fontes de dados é de aproximadamente 50 vezes nos próximos 10 anos. Segundo previsões apresentadas pela empresa EMC, instituição especializada em armazenamento de dados, o crescimento de dados

¹¹¹ Neologismo criado por Alvin Tofler que representa a união dos conceitos de produtor e consumidor

¹¹² Em um sistema de recomendação, parte-se de perfis de usuários específicos, que podem ser agrupados e relacionados a outros perfis que, quando incrementados com seus respectivos históricos de compras e com os dados originados pelas redes sociais, possibilitam a descoberta de produtos a serem ofertados (FLORISSI, 2014).

e informações digitais no mercado brasileiro crescerá de 212 Exabytes¹¹³ em 2014, alcançando a marca de 1.6 Zettabytes (1.600 Exabytes) em 2020 (EMC, 2014).

Fruto deste cenário, rico em volume e variedade de fontes, tem surgido uma disciplina que, apesar de não ser apenas um tema essencialmente tecnológico, vem sendo impulsionado pelos projetos de tecnologia: a vertente de Big Data (RIBEIRO, 2014).

3 REFERENCIAL TEÓRICO

Fox e Hendler (2011) também anteciparam que estamos vivendo com uma nova abordagem chamada de “Big Data”. Esta abordagem está surgindo em decorrência da geração, e, consequentemente, da necessidade da coleta de grande volume de dados com formatos variados. Ademais, estes dados ainda precisam ser geridos e, neste sentido, Fox e Hendler continuam e observam que a gestão destes recursos possibilitará a resolução de problemas que nem sabíamos que existiam. No entanto, vale ressaltar que não podemos prescindir de ferramentas, pois a capacidade do ser humano de analisar dados e informações com múltiplas facetas é limitada. Logo, são necessários alguns instrumentos que nos auxiliem a executar estas tarefas.

3.1 A noção de big data

A necessidade de vencer o desafio, reunindo e analisando fontes de diversas naturezas, deu origem a pesquisas que nos levaram ao tema “Big Data”. Estas pesquisas foram desenhadas a partir de três aspectos iniciais (DAVENPORT, 2014):

A múltipla natureza dos dados – aspecto relacionado com as diferentes fontes disponíveis.

O uso de processamento em nuvem – aspecto relacionado ao uso ilimitado de recursos computacionais e com processamento em larga escala, com a possibilidade de redução de custos (economia de escala – é o aspecto econômico-financeiro).

Uso de tecnologias específicas, tais como processamento de rotinas em paralelo e ferramentas para otimização como *Hadoop*¹¹⁴ e *MapReduce*¹¹⁵, *HDFS*¹¹⁶, além de abordagens de *MachineLearning*¹¹⁷ e *Analytics*¹¹⁸.

¹¹³ Em valores aproximados a unidade Exabyte é equivalente a cerca de 1.000 Petabytes, ou ainda, a 1.000.000 de Terabytes.

¹¹⁴ *Hadoop* é tecnologia *open source* desenvolvida pela Google e Yahoo para processar muitos dados em servidores, usando a noção de processamento em paralelo e uso de *clusters* (conjuntos) de computadores no processamento. Pode ser chamado de Apache Hadoop e também foi desenvolvido e customizado por outros fabricantes (EMC, Intel, Microsoft, dentre outros).

A abordagem de Big Data está apoiada em quatro outros fatores de sustentação, conhecidos como os 4 Vs do Big Data: Volume, Variedade, Velocidade e Veracidade (DUMBILL, 2012). A seguir será apresentado um breve esclarecimento do papel de cada um desses componentes:

O primeiro V é de Volume e está ligado ao grande quantitativo de dados e informações que nos cercam no cotidiano. Já o segundo V está ligado à variedade destes recursos.

Devido a forte relação entre Volume e Variedade, estes fatores serão comentados em conjunto. A multiplicidade de dispositivos e a capacidade destes dispositivos interagirem em rede está promovendo a verdadeira inundação de dados. Cada um de nós carrega junto de si um celular, que agindo como um sensor, pode enviar informação de localização das pessoas e permitir a realização de negócios direcionados¹¹⁹. Ao levarmos em consideração que o mundo tem cerca de 7 bilhões de habitantes (WIKIPEDIA, 2014) e que aproximadamente 6 bilhões possuem celulares (ONUBR, 2014), pensem no volume e na variedade de dados que pode ser gerado, captado, processado, (re)utilizado e entregue.

As cidades estão repletas de câmeras de monitoramento nos prédios, lojas, ruas e avenidas. Qualquer cidadão pode gravar e postar um vídeo em mídias sociais ou no Youtube. Estima-se que a quantidade de vídeos produzidos diariamente ultrapassa a produção dos primeiros 50 anos de televisão (DAVENPORT, 2014).

Saindo do cotidiano e observando o ambiente de ciência e tecnologia, temos muitos outros exemplos. Os projetos de pesquisa de perfuração de petróleo em águas profundas, incluindo o pré-sal (CIARINI, 2013; SANTOS, 2014), além de projetos de pesquisa em astronomia, estão impulsionando o uso da abordagem de Big Data (PORTO, 2013).

¹¹⁵ *MapReduce* é o framework arquitetural que deu origem à tecnologia de *Hadoop*. Usa a estratégia de dividir para conquistar, ou seja, distribui e aloca um problema muito grande em *clusters* de armazenamento, usando registros serializáveis do tipo <chave, valor>.

¹¹⁶ *HDFS* é a sigla de *Hadoop File System*. É uma estrutura de armazenamento de arquivos que utiliza blocos de 128 Mbytes, que são muito menores do que os blocos de particionamento tradicionais, utilizados em dispositivos de armazenamento.

¹¹⁷ *Machine Learning* trata o uso de algoritmos que identificam o melhor modelo para ser aplicado ao conjunto de dados.

¹¹⁸ *Analytics* é a essência de Big Data. Trata a análise dos dados e será apresentada mais à frente neste relato.

¹¹⁹ Um exemplo para a prática de negócios direcionados é o e-*e-couponing*. Esta prática possibilita o envio de cupons de desconto em tempo real para os usuários, quando os mesmos estão nas proximidades das lojas, utilizando-se as coordenadas GPS dos celulares dos usuários (FLORISSI, 2014).

Adiciona-se a esse cenário, uma vasta coleção de outras fontes e formas para geração de unidades documentárias. O crescimento do uso de documentos digitais e páginas Web nas organizações, recursos estes estruturados por meio de ferramentas para Gestão de Conteúdo (RIBEIRO, 2012), bem como o desenvolvimento de propostas de uso da Web of Data e Linked Data (RIBEIRO, ALMEIDA, 2011; RIBEIRO, VIEIRA, 2014) também têm contribuído para um aumento em Volume e Variedade de dados e informação.

Voltando aos 4 Vs do Big Data, chega-se agora ao terceiro V, de Velocidade. A melhoria dos canais de transmissão, com redes em fibra ótica e emissores de sinais de alta capacidade, o uso de satélites, o uso de outras bandas para a telefonia celular, as comunicações em tempo real para controle de processos na internet, os workflows científicos com processamento paralelo e cluster de processamentos vem possibilitando atingir uma maior velocidade para troca de dados e informação (MATTOSO, 2013). Ademais, é possível afirmar que a velocidade continuará crescendo, pois o desenvolvimento da tecnologia de processadores, dos canais e do hardware para armazenamento (discos rígidos e memória rápida – *flash memory*), duplica o seu poder a cada período de 2 anos (FLORISSI, 2012).

O quarto V é de Veracidade. A qualidade dos dados e informação é característica essencial para que os usuários interessados (executivos, gestores públicos e a sociedade em geral) usem e (re)usem os dados de maneira apropriada e real, gerando informações críveis para eles mesmos.

Para concluir a noção de Big Data ainda vale explorar um componente que faz parte do terceiro aspecto relacionado anteriormente por Davenport. A discussão sobre o trabalho de análise dos dados, entendidos pela noção de Big Data *Analytics*.

3.2 Big data analytics

O objetivo da tarefa de Analytics é executar a análise preditiva dos dados por meio da execução de mining (minerações)¹²⁰. Segundo os autores Oliveira (2013) e Tavares (2014), inicialmente, serão tratados os dados com o uso de técnicas estatísticas, para separação e reunião de conjuntos (denominado de fase de discovery).

Adicionalmente, para executar a tarefa também pode-se fazer uso de técnicas para categorização, limpeza e transformação dos dados, utilizando, inclusive, a visão da proveniência (fontes de origem) dos dados para auxiliar no processo de categorização. Ao final desta fase é possível chegar à definição e preparação de modelos (fase de *data*

¹²⁰ A noção de *mining* de dados passa pela extração e análise de grandes volumes de informação em busca de padrões e comportamentos.

preparation e model planning) que serão úteis na construção do grande conjunto de dados, chamado de lago de dados (*data lake*).

A carga de dados (denominada fase de *ingest*) ocorrerá em seguida e será realizada para povoar o lago de dados. No lago estarão reunidos todos os dados que serão alvo de análise. Por fim, os resultados que serão obtidos a partir do tratamento e análise do conteúdo do lago serão apresentados com uso de ferramentas de visualização e deverão estar associados ao contexto de negócios (OLIVEIRA, 2013; TAVARES, 2014).

A análise de dados que atendem aos requisitos descritos anteriormente (lembrem-se dos 4 Vs), precisará ser desenvolvida segundo uma nova arquitetura de análise, onde dados serão obtidos de múltiplas fontes e em tecnologias diversas. O ponto central desta análise está ligado à capacidade de correlacionar dados, pois, como já observado, o ser humano possui limitações para fazer análises associadas a múltiplas dimensões. Em essência, quando temos uma pequena quantidade de dados (*little data*) não temos muita dificuldade de correlacioná-los, pois existem poucas inter-relações. Mas, com uma grande quantidade (*big data*), temos muitos dados sendo gerados em paralelo, logo, surge a dificuldade para correlacioná-los (SEYMOUR, 2014).

Para Sathi (2013), o trabalho com os processos de negócio de uma organização começa a prescindir de um profissional que saiba executar tarefas ligadas à fase de *Analytics*, de forma que seja possível desenvolver novos produtos e serviços para os clientes e/ou usuários.

3.3 O cientista de dados e a fase de discovery

O surgimento desta nova arquitetura de análise, impulsionou o processo para formação de um perfil profissional que passou a ser denominado de Cientista de Dados (*Data Scientist*). A característica principal deste profissional é ter a capacidade de aplicar ferramentas analíticas e algoritmos para gerar previsões sobre produtos e serviços (DAVENPORT; PATIL, 2012). Oliveira (2013) complementa e detalha que este perfil deve ter forte conhecimento em disciplinas como a matemática e a estatística, com treinamento avançado em estratégias para tratamento de grandes conjuntos de dados, fazendo uso de modelos matemáticos, formulação de hipóteses e técnicas de regressão. Já Brietman (2014) observa que o Cientista de Dados deve ter capacidade de levantar requisitos dos usuários, buscando não apenas nas necessidades destes usuários, mas também nos outros envolvidos no ambiente sob análise (clientes, parceiros de negócio, informações de mercado, feeds de notícias, redes sociais, blogs, dentre outros).

Para Oliveira (2013), o cientista de dados deve ser um técnico cético, curioso, criativo, comunicativo e deve saber trabalhar em colaboração. Ademais, o cientista de dados deve sempre (re)avaliar questões durante as primeiras fases do desenvolvimento do trabalho.

Oliveira continua e apresenta questões que podem auxiliar na revisão das fases de *Discovery*, *Data preparation* e *Model Planning*. No entanto, daremos ênfase na fase de *Discovery*, por ser o recorte adotado para esta pesquisa.

As principais questões apresentadas para a fase de *Discovery* são: eu posso o conhecimento suficiente do ambiente de dados e informação? Eu tenho informação suficiente para esboçar um plano analítico e compartilhar com meus pares? Eu consigo desenvolver trabalhos para organização para tipos de problemas? Categorizações e classificações de dados? Projeto de conjuntos (*clusters*) de dados? Eu consigo esboçar e realizar entrevistas para conhecer o contexto e domínio que será trabalhado? Eu posso identificar as diferentes fontes de dados?

Em suma, os projetos de Big Data são desenvolvidos com os objetivos de criar novos produtos, compreender necessidades dos clientes e seus comportamentos, bem como perceber novos mercados. Para isto, é necessário desenvolver teorias para tratar com clientes e usuários, construindo hipóteses e identificando dados e informações relevantes. Este processo deve ser repetido e refinado, de acordo com os experimentos realizados e as respostas obtidas (MARCHAND; PEPPARD, 2013).

4 METODOLOGIA

O recorte deste estudo contempla o uso de recursos ligados a organização do conhecimento para auxiliar no desenvolvimento de projetos de Big Data, conforme requisitos apontados (ver 3.1 e 3.2).

Trata-se de um estudo exploratório e tendo em vista o recorte adotado, buscou-se identificar na primeira fase da pesquisa os conjuntos de dados e de informação que possuem correlação com a temática de Acidentes de Trabalho. Esta primeira fase foi desenvolvida sobre o conteúdo do portal de Dados Abertos da Dataprev¹²¹.

Com o intuito de complementar os conjuntos levantados, na segunda fase serão reunidos outros sítios específicos na internet e rede sociais que debatem temas correlatos com o assunto.

¹²¹ Disponível em <http://dadosabertos.dataprev.gov.br>

As questões sobre a relevância da informação tratadas por Saracevic (1970; 1996b) são utilizadas para auxiliar na definição dos conjuntos de interesse. Os itens relevantes, em geral, responderam a questões que podem ser de interesse para o cliente e/ou usuário. Estas questões devem ser desenvolvidas com o apoio de critérios, sintetizados a seguir (RIBEIRO, 2012):

- As informações ou os conjuntos de dados selecionados ocasionam sentimentos de excitação e satisfação aos clientes e/ou usuários;
- Especificidade das pesquisas realizadas pelos clientes e/ou usuários estão contempladas nos conjuntos selecionados;
- As respostas esperadas cobrem conceitos que estão presentes nas suas pesquisas.

Com uma primeira versão de escopo definido, passou-se a observar as questões apresentadas (ver 3.3) com o intuito de avaliar os conjuntos e correlações identificadas.

5 O EXPERIMENTO SOBRE OS DADOS DE ACIDENTES DE TRABALHO

A Previdência Social é detentora de muitos dados e informações da sociedade brasileira. Com o esforço para publicação de dados em formato aberto para poderem ser reutilizados pela sociedade (RIBEIRO, ALMEIDA, 2011), surgiram muitas possibilidades de reuso, tanto em projetos de pesquisa específicos quanto em trabalhos acadêmicos (FERREIRA; SANTANA; VIDOTTI, 2012; RODRIGUES, 2012; CAMPOS; CAMPOS; CARVALHO; LIMA, 2012; ROCHA; CHAVES, 2013; GERMANO, 2013).

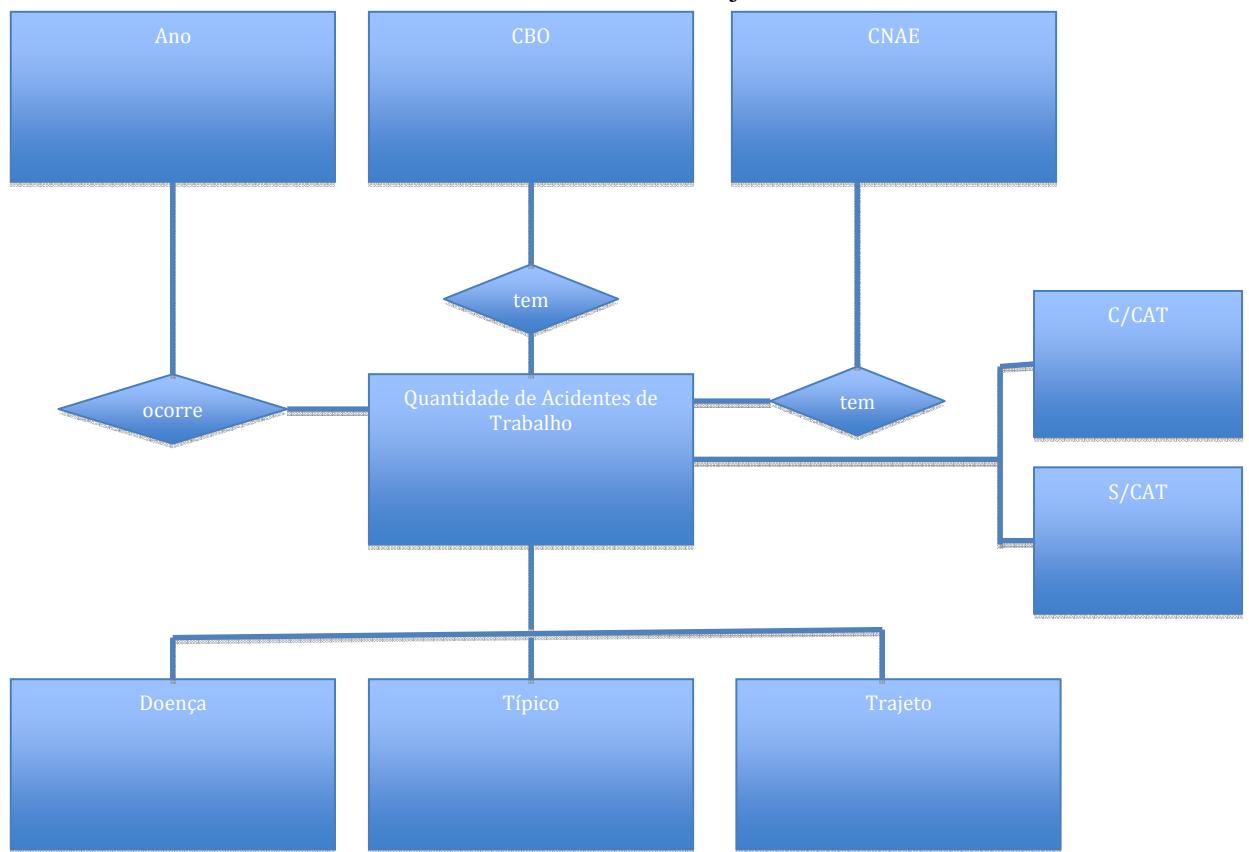
Esta investigação está sendo desenvolvida no âmbito de um projeto de pesquisa em curso na Unirio e está apoiada em trabalhos anteriores desenvolvidos pelo autor. O objetivo deste estudo é investigar um conjunto de atividades que viabilizem a participação do profissional da informação na fase de *Discovery* de projetos de Big Data. Assim, a delimitação do campo empírico se deu a partir da questão: é possível analisar os dados abertos sobre Acidentes de Trabalho disponíveis no portal da Previdência e correlacioná-los com comentários de blogs específicos e/ou redes sociais, para gerar análises que possibilitem a construção de políticas de prevenção?

Para Oliveira (2013) as duas primeiras etapas para executar o trabalho de *Discovery* e entendimento são: a identificação das informações que serão necessárias para possibilitar interpretações sobre os resultados, correlações, implicações e causalidade; e determinar o(s) tipo(s) de organização para estas informações (classificações e clusterizações).

Na tarefa de entendimento a elaboração de modelos de dados é útil, pois permite avaliar, inclusive, a possibilidade de realizar integrações de dados e associações. Sukumar e Ferrel (2013) convalidam esta estratégia quando afirmam que conhecer estrutura dos dados e seus relacionamentos, podem facilitar o trabalho de análise de conjuntos de dados.

Assim, a investigação partiu da identificação dos dados e informações relevantes no Portal de Dados abertos da Dataprev e, neste sentido, foram identificados dois conjuntos de dados que poderiam ser úteis para responder a questão de partida: Acidentes de Trabalho por CBO¹²² e Acidentes de Trabalho por CNAE 2.0¹²³. A representação a seguir (Figura 1) foi gerada a partir de uma extensão do modelo obtido em (RIBEIRO, ALMEIDA, 2011).

FIGURA 1: Modelo de Dados dos conjuntos escolhidos



Fonte: o autor

Os componentes desse modelo (entidades, relacionamentos e atributos) serão utilizados na avaliação do conteúdo dos comentários, pois permitem alguma representação semântica dos dados e informação contemplados pela modelagem.

¹²² Classificação Brasileira de Ocupações

¹²³ Classificação Nacional de Atividade Econômica

Cabe registrar que o modelo de dados permite representação semântica na medida em que este modelo apresenta elementos (entidades e atributos) e relações em um determinado domínio (CAMPOS, 2001). Campos, Campos, Carvalho e Lima (2012) reafirmam a estratégia de modelagem conceitual, pois permite uma boa forma de efetuar a representação do conhecimento em um domínio.

Vale ressaltar que conforme apresentado (ver 4) a complementação dos dados e demais correlações será desenvolvida na próxima etapa, quando serão investigados e trabalhados os dados originários de blogs e redes sociais, tomando-se por base os conceitos presentes no modelo de dados. Nesta direção, o uso de aplicações para fazer o monitoramento, tais como Socialmention (www.socialmention.com) e Topsy (www.topsy.com), possibilitará a coleta, organização e categorização das informações obtidas nas redes sociais, construindo uma proposta para representação de dados e informação na fase de *Discovery*. Já foram identificados, preliminarmente, os seguintes endereços, que irão nortear a delimitação das informações relevantes:

- <https://www.facebook.com/revistaprotecao>
- <https://www.facebook.com/TSTJus?ref=ts>
- <http://nrfacil.com.br/blog/?cat=172>
- <http://atdigital.com.br/direitoprevidenciario/category/acidentes-do-trabalho/>
- <http://seguranca-trabalho-tst.blogspot.com.br>
- <http://blog.pesquisasaude.com/dia-nacional-da-preveno-de-acidentes-trabalho/>

Shiri (2014) convalida a proposta aqui formulada para análise de informação na fase de *Discovery*, quando apresenta facetas e sub-facetadas para análise de grandes volumes de dados. Apoiado em tipos de dados e no contexto dos conjuntos sob análise, além da estrutura dos dados e dos metadados, Shiri propõe estender práticas de organização e representação do conhecimento para trabalhos com Big Data.

O trabalho de análise dos dados e informação originários de redes sociais pode ser melhor estruturado com o apoio de uma Ontologia, pois esta pode guiar a identificação dos objetos, suas características, propriedades e contextos que serão fundamentais no processo de organização (PORTO, BAX, FERREIRA, SILVA; 2012). No entanto, pretende-se com este estudo avaliar a possibilidade de uso de outras formas de representação, que podem apontar caminhos alternativos à investigação, contribuindo para a evolução de métodos de representação na fase de *Discovery* e com a participação de profissionais da informação.

6 CONSIDERAÇÕES FINAIS

O uso das abordagens ligadas ao tema Big Data ainda carece de investigação pela área de Ciência da Informação. O uso dessas abordagens poderá auxiliar na melhoria da oferta de serviços de informação, pois conhecer o ambiente de dados e informação que deve ser originário de diferentes fontes, efetuar a organização de conjunto de dados (categorizá-los?), realizar entrevistas junto aos clientes e/ou usuários e desenvolver os modelos (tanto estruturais quanto matemáticos), contribuirá para o projeto desses serviços.

Para Minelli, Chambers e Dhiraj (2013), o momento que vivemos é especial, pois a contínua redução do custo dos equipamentos, além do uso de novos softwares e ferramentas para apoiar os processos de gestão de dados e informação, têm contribuído para um momento especial no tratamento da informação.

Quando lançamos um olhar na direção das bibliotecas e unidades de informação, é possível perceber que tanto os estoques existentes quanto as demandas por outras fontes têm crescido de forma exponencial, pois os ativos de informação de interesse para os usuários estão armazenados em diferentes bases de dados, usando bancos de dados e plataformas de computação heterogêneas (MELETIOU, KATSIRIKOV; 2009). Adicione-se a isso, a presença cada vez maior de redes sociais e aplicações de colaboração nos serviços das unidades de informação (DE JESUS, DA CUNHA; 2012; SANTOS, DA ROCHA; 2012), ocasionando mudanças no comportamento dos usuários. Em suma, compreender e participar de esforços com o uso de Big Data pode ser valioso para auxiliar na estruturação de unidades de informação.

Espera-se que este movimento de pesquisa sobre o tema Big Data na área da Ciência da Informação, ilumine o caminho a ser trilhado e possibilite que outros pesquisadores interessados possam se engajar nesta discussão, levando os debates sobre este tema para além da tecnologia.

REFERÊNCIAS

BRIETMAN, K. Big Data Seen from the Clouds. 2014. Palestra proferida no 2º. EMC Summer School on Big Data. EMC/NCE/UFRJ, Rio de Janeiro, 12 mai. 2014. Disponível em http://2014.emcbigdataschool.nce.ufrj.br/images/presentations/_Big_Data_Summer_School_Karin.pdf. Acesso em 30 de Maio de 2014.

CAMPOS, M. L. A. **A organização de unidades de conhecimento:** o modelo conceitual como espaço comunicacional para a realização da autoria. 2001. Tese (Doutorado em Ciência da Informação). ECO/IBICT, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2001.

CAMPOS, M. L. M.; CAMPOS, M. L. de A.; CARVALHO, M. G. P. de; LIMA, D. V. M. de. Dados abertos interligados e o espaço do profissional de informação: Uma aplicação no domínio da enfermagem. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO – XIII ENANCIB, 2012, Rio de Janeiro, RJ. **Anais...** Rio de Janeiro: FIOCRUZ, 2012.

CIARINI, A. E. M. Research on Big Data and Opportunities. Palestra proferida no 1º. EMC Summer School on Big Data. EMC/NCE/UFRJ, Rio de Janeiro, Fev. 2013.

DAVENPORT, T. H. **Big Data at Work:** Dispelling the Myths, Uncovering the Opportunities. Boston: Harvard Business Review Press Books. 2014.

DAVENPORT, T. H.; PATIL, D.J. Data Scientist: The Sexiest Job of the 21st Century. **Harvard Business Review** v. 90, n. 10, p. 70–76, Out. 2012.

DE JESUS, D. L.; DA CUNHA, M. B. Produtos e serviços da web 2.0 no setor de referência das bibliotecas. **Perspectivas em Ciência da Informação**, v. 17, n. 1, p. 110-133, 2012.

DUMBILL, E. What is Big Data? In: O'REILLY MEDIA INC. **Big Data Now: current perspectives**. California: O'Reilly Media, 2012. Disponível em: <<http://www.oreilly.com/data/free/files/big-data-now-2012.pdf>>. Acesso em: 03 maio 2014.

EMC. Brazil country brief. The Digital Universe of opportunities. [relatório na internet] 2014. Disponível em: <<http://www.emc.com/collateral/analyst-reports/idc-digital-universe-2014-brazil.pdf>>. Acesso em: 12 mai. 2014.

FERREIRA, A. M. J. F. da C.; SANTANA, R. C.; VIDOTTI, S. A. B. G. Second life: perspectivas para potencializar o acesso a dados públicos. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO – XIII ENANCIB, 2012, Rio de Janeiro, RJ. **Anais ...** Rio de Janeiro: FIOCRUZ, 2012.

FLORISSI, P. Big Data: Challenges and Opportunities. Palestra proferida no 2º. EMC Summer School on Big Data. EMC/NCE/UFRJ, Rio de Janeiro, 13 mai. 2014.

FLORISSI, P. EMC On Big Data. 2012. Disponível em: <<https://www.carecorenational.com/healthcaresummit/powerpoints/PatriciaFlorissiPhD.pdf>>. Acesso em 12 mai. 2014.

FOX, P.; HENDLER, J. Changing the Equation on Scientific Data Visualization. **Science**, v. 331, p. 705, 2011. Disponível em: <http://data2discovery.org/dev/wp-content/uploads/2013/05/Fox-and-Hendler_Visualization_Science-2011-Fox-705-8.pdf>. Acesso em: 15 mai 2014.

GERMANO, E. C. **Modelos de negócios adotados para o uso de dados governamentais abertos:** estudo exploratório de prestadores de serviços na cadeia de valor dos dados governamentais abertos. 2013. Dissertação (Mestrado em Administração) - Faculdade de Economia, Administração e Contabilidade, Universidade de São Paulo, São Paulo, 2013. Disponível em: <<http://www.teses.usp.br/teses/disponiveis/12/12139/tde-10012014-155226/>>. Acesso em: 26 jul 2014.

GOPINATH, M A.; DAS, P. Classification and representation of Knowledge. **Library Science with a Slant to Documentation and Information Studies**, v. 34, n. 2, p. 85-90, 1997.

HEATH T.; BIZER C. **Linked Data:** Evolving the Web into a Global Data Space (1st edition). Synthesis Lectures on the Semantic Web: Theory and Technology. Morgan & Claypool Publishers, 2011.

MARCHAND, D. A.; PEPPARD, J. Why IT Fumbles Analytics. **Harvard Business Review.** Jan-Fev 2013.

MATTOSO, A. Scientific Workflows and Big Data. Palestra proferida no 1º. EMC Summer School on Big Data. EMC/NCE/UFRJ, Rio de Janeiro, Fev. 2013.

MELETIOU, A; KATSIRIKOU, A. Methodology of analysis and interrelation of data about quality indexes of library services by using data-and knowledge-mining techniques. **Library Management**, v. 30, n. 3, p. 138-147. 2009 .

MINELLI, M.; CHAMBERS, M.; DHIRAJ, A. **Big Data, Big Analytics:** Emerging Business Intelligence and Analytic Trends for Today's Businesses. Wiley CIO Series. 2013

OLIVEIRA, A. Data Science and Data Analytics. 2013. Palestra proferida no 1º. EMC Summer School on Big Data. EMC/NCE/UFRJ, Rio de Janeiro, Fev. 2013.

ONUBR. NAÇÕES UNIDAS NO BRASIL. ONU: Dos 7 bilhões de habitantes do mundo 22 de março de 2013. Disponível em <http://www.onu.org.br/onu-dos-7-bilhoes-de-habitantes-do-mundo-6-bi-tem-celulares-mas-25-bi-nao-tem-banheiros/> Acesso em 10 de maio de 2014.

PORTO, F. Big Data in Astronomy: The LIneA-DEXL case . Palestra proferida no 1º. EMC Summer School on Big Data. EMC/NCE/UFRJ, Rio de Janeiro, Fev. 2013.

PORTO, R. M. A. B.; BAX, M. P.; FERREIRA, L. G. da F.; SILVA, G. C. Análise de sentimento sobre veículos em redes sociais. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO – XIII ENANCIB, 2012, Rio de Janeiro, RJ. **Anais ...** Rio de Janeiro: FIOCRUZ, 2012.

RIBEIRO, C. J. S. **Diretrizes para o projeto de portais de informação:** uma proposta interdisciplinar baseada na Análise de Domínio e Arquitetura da Informação. 2008. 298 f. Tese (Doutorado em Ciência da Informação) – Convênio UFF/IBICT, Rio de Janeiro.

RIBEIRO, C. J. S. Projeto de serviços para Gestão Documental com uso de abordagem tecnológica SOA (Arquitetura Orientada a Serviços). **Informação & Sociedade: Estudos**, João Pessoa, v. 22, n. 2, 2012.

RIBEIRO, C. J. S.; ALMEIDA, R. F. Dados Abertos Governamentais (Open Government Data): Instrumento para Exercício de Cidadania pela Sociedade. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO – XII ENANCIB, Brasília, DF. **Anais ...** Brasília: Thesaurus, 2011.

RIBEIRO, C. J. S.; PEREIRA, D. V. El proceso de publicación e los datos gubernamentales abiertos acerca de la clase de la Seguridad Social Brasileña de Vocabulario Controlado del Gobierno Electrónico (VCGE). In: SIMPOSIO LATINOAMERICANO ACCESSO A LA INFORMACIÓN GUBERNAMENTAL, 2014, Cidade do México. **Anais ...** Cidade do México, 2014.

RIBEIRO, C. J. S. Big Data: os novos desafios para o profissional da informação. **Informação & Tecnologia (ITEC)**: João Pessoa/Marília, v. 1, n. 1, p. 96-105, jan./jun., 2014

ROCHA, E.; CHAVES, M. GovMobile: Uma Proposta para Disponibilizar Dados Abertos Georreferenciados para Governo Eletrônico. In: WAIHCWS. 2013. *Anais ...* p. 14-20. Disponível em : <http://ceur-ws.org/Vol-1051/paper2.pdf>. Acesso em 20 de julho de 2014.

RODRIGUES, F. de A. **Mapeamento de tecnologias informacionais sobre dados abertos em saúde pública**: destino de repasses financeiros federais. Dissertação (Mestrado em Ciência da Informação). Universidade Estadual Paulista. Faculdade de Filosofia e Ciências. 2012.

SANTOS, E. L.; DA ROCHA, S. M. O blog como ferramenta de comunicação entre a biblioteca e seus usuários: a experiência da biblioteca Lydio Bandeira de Mello, da Faculdade de Direito da Universidade Federal de Minas Gerais. **Encontros Bibli**, v. 17, n. 33, p. 134-152, 2012.

SANTOS, I. H. R. BigData Research and Developement at Petrobras. Palestra proferida no 2º EMC Summer School on Big Data. EMC/NCE/UFRJ, Rio de Janeiro, 13 mai. de 2014. Disponível em:
http://2014.emcbigdataschool.nce.ufrj.br/images/presentations/Ismael_BigDataTOOL_SummerSchool_v2.pdf. Acesso em: 20 mai. 2014.

SAP Blog. A SAP e a Associação Alemã de Futebol (DFB) transformaram o Big Data em decisões inteligentes para melhorar o desempenho dos jogadores na Copa do Mundo no Brasil. SAP17 de junho de 2014. Disponível em : <http://blogs.sap.com/brazil/big-data/sap-e-associacao-alema-de-futebol-dfb-transformaram-o-big-data-em-decisoes-inteligentes-para-melhorar-o-desempenho-dos-jogadores-na-copa-mundo-brasil-2744>

SARACEVIC, T. The concept of "relevance" in Information Science: an historical review. In: **Introduction to Information Science**. New York: R. R. Bowker, 1970. p.111-154.

SARACEVIC, T. Ciência da informação: origem, evolução e relações. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 1, n. 1, p. 41-62, jan./jun. 1996a.

SARACEVIC, T. Relevance reconsidered 1996. Information Science: Integration in Perspectives. In: INTERNATIONAL CONFERENCE ON THE CONCEPTIONS OF LIBRARY AND INFORMATION SCIENCE, 2 (COLIS, 2). Copenhagen. **Anais ...** Denmark. 1996b.

SATHI, A. **Big Data Analytics**: Disruptive Technologies for Changing the Game. Mc Press. 2013.

SEYMOUR, C. The State of Big Data. **EContentMag.com**. p. 26-27. Jan-Feb 2014.

SHIRI, A. Making Sense of Big Data: A Facet Analysis Approach. **Knowledge Organization (Journal)**, v. 41, n. 5, p. 357-368, 2014.

SUKUMAR, S. R.; FERREL, R. K. 'Big Data' collaboration: Exploring, recording, and sharing enterprise knowledge. **Information Services & Use**, n. 33. 2013.

TAVARES, E. Big Data in Business. Palestra proferida no 2º EMC Summer School on Big Data. EMC/NCE/UFRJ, Rio de Janeiro, 12 mai. 2014. Disponível em:
http://2014.emcbigdataschool.nce.ufrj.br/images/presentations/Apresentacao_Elaine_Tavare s.pdf. Acesso em: 30 mai 2014.

TOFLER, A. **Third Wave**: The classic study for tomorrow. Bantam Books: New York. 1984

WIKIPEDIA. A Enciclopédia Livre. Disponível em:
<http://pt.wikipedia.org/wiki/População_mundial>. Acesso em: 10 mai. 2014.

WURMAN, R. S. **Ansiedade de Informação 2**. São Paulo: Editora de Cultura, 2005.

ONTOLOGIA BASEADA NOS FRBR: PROPOSTA DE APLICAÇÃO EM CATÁLOGOS ONLINE

ONTOLOGY BASED ON FRBR: PROPOSED APPLICATION IN ONLINE CATALOGUE

Rogério Aparecido Sá Ramalho

Zaira Regina Zafalon

Marcos Teruo Ouchi

Resumo: Com o desenvolvimento e a popularização de tecnologias informacionais ampliaram-se e inovaram-se as formas de produção e registro do conhecimento humano de modo a ser possível recorrer aos mais variados suportes para seu registro. Observa-se, também, que, com o aumento exponencial da criação, da transmissão e da transferência de informações em meios digitais, um novo paradigma informacional é apresentado, com destaque ao acesso e não à propriedade. Diante desse cenário, e com o intuito de garantir a recuperação e o (re)uso de informações, (re)definem-se métodos e instrumentos de organização, disseminação e acesso aos registros, no qual destacam-se as ontologias e a catalogação. Neste sentido, apresenta-se a seguinte questão de pesquisa: é possível utilizar as relações existentes entre as entidades dos Requisitos Funcionais dos Registros Bibliográficos como substratos teóricos no desenvolvimento de ontologias, para que seja possível garantir aos usuários pela navegação, o desenvolvimento de tarefas que permitam encontrar, identificar, selecionar e obter acesso aos recursos bibliográficos? Apresenta-se como objetivo geral desenvolver um protótipo de ontologia baseada nas relações previstas entre as entidades dos Requisitos Funcionais dos Registros Bibliográficos (FRBR), que considera, em seu bojo, as tarefas a serem desempenhadas pelos usuários. Define-se, como objetivos específicos, analisar as contribuições teórico-metodológicas da Ciência da Informação no desenvolvimento de ontologias na representação da informação, com suas ferramentas, técnicas, tecnologias e metodologias; e descrever entidades e relacionamentos previstos nos Requisitos Funcionais de Registros Bibliográficos e sua correlação com as tarefas a serem desempenhadas pelos usuários. A pesquisa caracteriza-se com abordagem qualitativa, de natureza aplicada, com objetivos exploratórios e descritivos e recorre-se à pesquisa bibliográfica e documental para a construção do arcabouço teórico. Optou-se, como procedimento metodológico que permite viabilizar a resposta à questão de pesquisa, pelo desenvolvimento de um protótipo que se utiliza das entidades e relacionamentos previstos nos FRBR como substrato teórico para a elaboração da ontologia. Os resultados demonstram viabilidade parcial de se recorrer às entidades dos Requisitos Funcionais de Registros Bibliográficos para fundamentar o desenvolvimento de ontologias. Esse resultado é decorrente do fato de serem encontradas dificuldades conceituais para a modelagem, haja vista os FRBR considerarem que algumas das entidades podem estar previstas em dois grupos. Tal situação faz com que a modelagem seja considerada conceitualmente incorreta, uma vez que não é possível satisfazer aos critérios de entidades pertencerem a diferentes grupos, caracterizando polissemia.

Palavras-chave: Ontologias. Representação da Informação. Catalogação. Requisitos Funcionais para Registros Bibliográficos.

Abstract: With the development and popularization of information technologies have expanded and innovated the forms of production and record of human knowledge in order to be possible to use various media to your registry. Observe also that, with the exponential increase in the creation, transmission and transfer of information in digital media, a new informational paradigm is presented, emphasizing access and not ownership. Given this scenario, and in order

to ensure recovery and (re) use of information, (re) define themselves instruments and methods of organization, dissemination and access to records, in which we highlight the ontologies and cataloging. In this sense, we present the following research question: is possible to use relations existing between the entities of the Functional Requirements for Bibliographic Records as the theoretical substrates applied to development of ontologies, so that you can ensure users for navigation, the development of tasks that allow to find, identify, select and obtain access to library resources? It is presented as general objective to develop a prototype ontology based on relationships between entities provided the Functional Requirements for Bibliographic Records (FRBR), which considers, in its wake, the tasks to be performed by users. Is defined, specific objectives, analyze the theoretical and methodological contributions of Information Science to develop ontologies in information representation, with their tools, techniques, technologies and methodologies; and describe entities and relationships specified in Functional Requirements for Bibliographic Records and its correlation with the tasks to be performed by users. The research is characterized with qualitative approach and applied nature, with exploratory and descriptive goals and resorts to literature and documents to build the theoretical framework research. It was decided, as a methodological procedure to facilitate the response to the research question, the development of a prototype which uses the entities and relationships in FRBR as provided theoretical basis for the elaboration of the ontology. The results demonstrate feasibility of using partial entities of the Functional Requirements of Bibliographic Records to support the development of ontologies. This result is due to the fact that they found it difficult to conceptual modeling, given FRBR consider that some of the entities can be provided in two groups. This situation makes the modeling is considered to be conceptually wrong, since it is not possible to meet the criteria of entities belonging to different groups, characterized polisemy.

Keywords: Ontologies. Representation of Information. Cataloging. Functional Requirements for Bibliographical Records.

1 INTRODUÇÃO

Ao longo dos anos, com o desenvolvimento e a popularização de tecnologias informacionais, ampliaram-se e inovaram-se as formas de produção e registro do conhecimento humano de modo a ser possível recorrer aos mais variados suportes para seu registro, quer sejam analógicos, eletrônicos ou digitais. Observa-se, também, que, com o aumento exponencial da criação, da transmissão e da transferência de informações em meios digitais, um novo paradigma informacional é apresentado, com destaque ao acesso e não à propriedade, sendo possível ignorar barreiras geográficas, culturais e temporais.

Identificam-se, entretanto, problemas advindos da relação intrínseca dessa situação junto ao pressuposto apresentado por Ranganathan (2009), *Poupe o tempo do leitor*, que, por extensão, associa-se a todo e qualquer processo que deva ser envidado de modo a garantir satisfatoriamente a recuperação e o acesso ao documento pelo usuário, e quanto às críticas apresentadas por Bush (1945), de que um registro do conhecimento torna-se útil à Ciência a partir do momento em que se é ampliado, armazenado e, sobretudo, consultado. Se, à época, Bush (1945) apresentava críticas às limitações dos sistemas de organização e representação da informação e à “*artificiality of systems of indexing*”, o que dizer da aplicabilidade de métodos

e de instrumentos do século passado em uma era em que *tablets*, *e-readers*, *smartphones*, bases de conhecimento colaborativo (*wikis*), portais, *blogs*, *microblogs* e redes sociais fazem parte do cotidiano?

Diante desse cenário, e com o intuito de garantir a recuperação e o (re)uso de informações, (re)definem-se métodos e, consequentemente, instrumentos de organização, disseminação e acesso aos registros, no qual destacam-se as ontologias, os catálogos e a catalogação.

Neste sentido, apresenta-se a seguinte questão de pesquisa: é possível utilizar as relações existentes entre as entidades dos Requisitos Funcionais dos Registros Bibliográficos como substratos teóricos no desenvolvimento de ontologias, para que seja possível garantir aos usuários pela navegação, o desenvolvimento de tarefas que permitam encontrar, identificar, selecionar e obter acesso aos recursos bibliográficos?

As ontologias têm favorecido uma aproximação entre os modelos e métodos tradicionalmente utilizados no âmbito da Ciência da Informação com as novas tecnologias semânticas. “[...] as ontologias têm sido apontadas como uma nova possibilidade de representação de recursos informacionais em bibliotecas.” (RAMALHO, 2010, p. 17). As tecnologias, por sua vez, como extensoras das atividades de coleta, processamento, memória e transmissão de dados, contribuem para um sistema de acesso à informação. Svenonius (2000), entretanto, ressalva que a eficácia do sistema é função direta da inteligência adotada para organizá-lo.

Nesta perspectiva, a principal motivação nesta pesquisa advém da contribuição que se espera dar no âmbito da implementação de novos recursos que recorrem à tecnologia para o atingimento da missão fundamental das bibliotecas, tal qual apresentado por Ranganathan (2009, p. 263): “[...] ser um instrumento de educação universal que reúne e difunde livremente todos os recursos de ensino e dissemina o conhecimento com a ajuda delas”.

Apresenta-se, assim, o objetivo geral desse trabalho: desenvolver um protótipo de ontologia baseada nas relações previstas entre as entidades dos Requisitos Funcionais dos Registros Bibliográficos, que considera, em seu bojo, as tarefas a serem desempenhadas pelos usuários. Com vistas ao cumprimento deste objetivo definem-se os seguintes objetivos específicos: a) analisar as contribuições teórico-metodológicas da Ciência da Informação no desenvolvimento de ontologias na representação da informação, com suas ferramentas, técnicas, tecnologias e metodologias; e b) descrever entidades e relacionamentos previstos nos Requisitos Funcionais de Registros Bibliográficos e sua correlação com as tarefas a serem desempenhadas pelos usuários.

Esta pesquisa caracteriza-se com abordagem qualitativa, pois, segundo Silveira e Córdova (2009), apresenta características voltadas à descrição, compreensão e explicação de relações de um determinado fenômeno de modo a considerar o caráter interativo entre os objetivos definidos e as orientações teóricas da pesquisa. Determina-se, também, como pesquisa de natureza aplicada, uma vez que “objetiva gerar conhecimentos para aplicação prática, dirigidos à solução de problemas específicos” (op. cit., p. 35). Quanto aos objetivos, a pesquisa comprehende aspectos exploratórios e descritivos, uma vez que busca tornar o problema mais explícito a partir de hipóteses, o que exige descrição de fatos observados (GIL, 2007; TRIVIÑOS, 1987). Recorre-se a procedimentos de pesquisa bibliográfica e documental para a construção do arcabouço teórico, uma vez que recorre a fontes de informação primárias e secundárias, e a partir do desenvolvimento de um protótipo busca-se analisar a possibilidade da utilização das entidades dos Requisitos Funcionais de Registros Bibliográficos como substratos teóricos para a elaboração de ontologias.

A pesquisa consolida, no tópico dois, as ontologias como instrumento de representação, no tópico três, as entidades e relacionamentos dos Requisitos Funcionais dos Registros Bibliográficos e as tarefas dos usuários, e, no tópico quatro, apresenta um protótipo de ontologia que possibilita a junção dos conceitos relacionados ao desenvolvimento de ontologia com os padrões já sedimentados na área de Catalogação a partir da realização de inferências automáticas baseadas nas entidades e relacionamentos previstos nos Requisitos Funcionais dos Registros Bibliográficos.

2 ONTOLOGIAS COMO INSTRUMENTO DE REPRESENTAÇÃO

No âmbito da área de Ciência da Informação o termo ontologia passou a ser utilizado no final da década de 1990, quando instrumentos de representação de informações despertaram maior interesse dos pesquisadores das mais diversas áreas do conhecimento, impulsionados pela criação do ambiente Web e desenvolvimento de novas tecnologias digitais (SOERGEL, 1999; VICKERY, 1997).

García-Marco (2007) ressalta que o crescente aumento do número de pesquisas relacionadas com ontologias e metadados na área de Ciência da Informação constitui a “ponta do *iceberg*” de um processo de reconfiguração disciplinar, como resultado da integração de diversas Ciências. Deste modo, verifica-se ao longo da última década um maior empenho na consolidação das definições de ontologia e análise das possibilidades de uso como instrumento de representação (SANTOS; CORREA; SILVEIRA, 2013).

Entre as definições de ontologia mais referenciadas na literatura destaca-se a apresentada por Gruber (1993, p. 1) de que “é uma especificação explícita de uma conceitualização”. Contudo, tal definição remete o conceito de ontologia a aspectos filosóficos, favorecendo a geração de conflitos terminológicos, conforme destaca Guarino (1998, p. 5):

No sentido filosófico, podemos nos referir a uma "ontologia" como um sistema particular de categorias que versa sobre uma certa visão do mundo. Desta forma, este sistema não depende de uma linguagem particular: a ontologia de Aristóteles é sempre a mesma, independente da linguagem usada para descrevê-la. Por outro lado, em seu uso mais prevalecente na IA [inteligência artificial], uma ontologia é referida como um artefato de engenharia, constituído de um vocabulário específico usado para descrever uma certa realidade e um conjunto de pressupostos explícitos relacionados com o significado pretendido para as palavras do vocabulário.

Nesta perspectiva, com o intuito de aproximar o conceito de ontologia aos pressupostos da área de Ciência da Informação, principalmente quanto aos processos de representação, tratamento, organização e recuperação de recursos informacionais, recorre-se a Ramalho (2010, p. 107) que define ontologia como

[...] um artefato tecnológico que possibilita representar formalmente as propriedades e relacionamentos de um determinado modelo conceitual, favorecendo a utilização de inferências automáticas nos processos de organização e recuperação de recursos informacionais.

Tal definição caracteriza as ontologias no âmbito da área de Ciência da Informação como instrumentos de nível epistemológico, concebidos a partir de linguagens computacionais com o intuito de favorecer a representação formal dos relacionamentos existentes entre os conceitos de um domínio específico.

Neste contexto, as ontologias figuram como uma nova abordagem computacional de representação, descrição e organização de conteúdos informacionais, favorecendo a aproximação dos modelos clássicos de representação, utilizados tradicionalmente no âmbito da área de Ciência da Informação, com as novas tecnologias semânticas desenvolvidas ao longo dos últimos anos.

As ontologias potencializam a representação dos relacionamentos existentes entre os conceitos que *traduzem* os conteúdos informacionais de modo a possibilitar representações semânticas que não poderiam ser obtidas a partir de descrições textuais, caracterizando-se como um tipo de fonte secundária, instituindo novas formas de acesso e favorecendo maior flexibilidade na recuperação de informações, por meio da realização de inferências. Quanto ao enfoque das representações semânticas, Zafalon (2013) esclarece que “é a semântica que dá

conta dos processos mentais segundo os quais se produz, constitui, comprehende e descreve a representação de um recurso informacional.”

Deste modo, verifica-se que a Ciência da Informação pode contribuir para o desenvolvimento de ontologias a partir do fornecimento de substratos teóricos e metodológicos que possibilitam o reaproveitamento de instrumentos e métodos já sedimentados em seu campo de atuação.

Ao longo dos últimos anos inúmeras metodologias têm sido propostas para o desenvolvimento de ontologias. Dentre elas, destacam-se a de Uschold e King (1995), a de Uschold e Gruninguer (1996), a de Fernández, Gómez-Pérez e Juristo (1997), e o método 101, proposto por Noy e McGuinness (2001). No âmbito da área de Ciência da Informação destacamos o método proposto por Ramalho (2010) para o desenvolvimento de ontologias a teoria do conceito e a teoria da classificação facetada utilizadas na elaboração de tesouros (CAMPOS; GOMES, 2006; GOMES, et al., 1990), bem como a norma ANSI/NISO Z39.19-2005, que apresenta diretrizes para a elaboração de vocabulários controlados.

No escopo desta pesquisa recorre-se, dentre os diversos métodos, modelos e instrumentos adotados na Ciência da Informação, às relações previstas entre as entidades dos Requisitos Funcionais de Registros Bibliográficos (FRBR), estabelecidos pela *International Federation of Library Associations and Institutions* (1998), para o desenvolvimento de um catálogo baseado em ontologia. Esta opção foi feita com base na aproximação cada vez mais notável entre a Ciência da Informação e a Ciência da Computação e pelo fato de que o desenvolvimento de ontologias apresentar-se como um processo não-linear, que requer refinamentos e interações sucessivas até a obtenção de um modelo que se mostre adequado às necessidades propostas. Conforme destacam Zhao e Hu (2007) a utilização de ontologias mostra-se como uma abordagem promissora que possibilita expressar formalmente a semântica de informações estruturadas e semiestruturadas, a fim de apoiar a aquisição, manutenção e acesso de informações.

3 REQUISITOS FUNCIONAIS PARA REGISTROS BIBLIOGRÁFICOS (FRBR) E AS TAREFAS DOS USUÁRIOS

Catalogar recursos informacionais mostra-se como atividade que fundamenta a recuperação e o acesso aos documentos uma vez que é por meio do registro bibliográfico que se torna possível efetivar o processo comunicativo entre unidades de informação e seus usuários; em outras palavras, é por meio dos registros, resultados da catalogação, que os documentos tornam-se recuperáveis e acessíveis em catálogos. Afinal, conforme afirma

Zafalon (2012, p. 46), “As bibliotecas utilizam-se de catálogos, instrumentos disponíveis em diversas formas ou suportes (livros, folhas soltas, fichas ou eletrônicos) com o objetivo de garantir a recuperação dos registros bibliográficos pelo público a que se destina.”

Segundo Santos e Correia (2009), o registro bibliográfico compreende um conjunto de elementos que apresentam funcionalidades de unicidade, instanciamento e armazenamento do item. De acordo com a International Federation of Library Associations and Institutions (2005) o registro bibliográfico é um conjunto de dados, compostos por entidades, que descrevem ou identificam um ou mais obras, expressões, manifestações ou itens.

Em se tratando do universo bibliográfico, a International Federation of Library Associations and Institutions (2009) afirma que as entidades, seus atributos e relacionamentos estão presentes no registro bibliográfico uma vez que “baseia-se, tipicamente, no item enquanto representante da manifestação e pode incluir atributos que pertençam à(s) obra(s) e expressão(ões) nela contidas”. As entidades representam conceitos chaves para uso em registros bibliográficos pois é na catalogação que há sistematização dos atributos e dos relacionamentos característicos das entidades. Zafalon (2011, p. 136) afirma que o processo de representação “visa, por um lado, a individualização de registros do conhecimento, pelos seus diferenciados aspectos físicos, ou de suas manifestações, ou de seus itens; e, por outro, a reunião por meio de suas características de conteúdo, ou de suas expressões ou de suas obras.”

Desse modo, com vistas ao estudo sobre a apresentação de resultados de pesquisa ao usuário estabeleceu-se o modelo conceitual de um conjunto de entidades, seus atributos e relacionamentos, consolidados nos Requisitos Funcionais de Registros Bibliográficos (FRBR) e publicados pela International Federation of Library Associations and Institutions (1998).

Dentre os objetivos traçados para os FRBR está incluso o de “fornecer um quadro estruturado, claramente definido, para relacionar dados registrados em registros bibliográficos às necessidades dos usuários destes registros.” (op. cit.). O modelo apresenta um conjunto de dez entidades, subdivididas, a partir da função que desempenham, do seguinte modo: *obra*, *expressão*, *manifestação* e *item* consolidam o grupo que retrata o produto do esforço intelectual e artístico (o quê); *pessoa* e *entidade coletiva* retratam o grupo responsável pelo conteúdo intelectual ou artístico bem como pela produção e disseminação (quem); *obra*, *expressão*, *manifestação*, *item*, *pessoa*, *entidade coletiva*, *conceito*, *objeto*, *evento* e *lugar* consolidam as entidades que servem como assunto de um *obra* (sobre).

A consolidação de tais entidades, representadas por meio de seus atributos, estabelece relacionamentos dos mais variados tipos. Svenonius (2000) ressalta que relações entre obras incluem relações generalizadas (é uma subclasse de), relações de agregação (é parte de), e

relações associativas (é uma seqüência de, é uma adaptação de, é uma abreviação de, é descrito por). Relações entre nomes de atributos de obra incluem equivalência, hierarquia e relações associativas. A figura abaixo exemplifica a distribuição de entidades e relacionamentos previstos nos FRBR para a obra Dom Casmurro, de Machado de Assis.

FIGURA 1 - Exemplo de relacionamentos e atributos de entidades na obra Dom Casmurro



Fonte: Elaborado pelos autores.

Por meio das relações estabelecidas entre as entidades torna-se possível compreender, segundo International Federation of Library Associations and Institutions (2009), que a funcionalidade de *pesquisa* é característica dos pontos de acesso estabelecidos em um registro bibliográfico e de autoridade, uma vez que delimitam os resultados apresentados, e que, diante da *recuperação* de vários registros que correspondem a um mesmo ponto de acesso, se faz necessária a apresentação dos resultados de uma forma que faça sentido e seja conveniente ao usuário. Assim, de acordo com a *Declaração de Princípios Internacionais de Catalogação*, publicada pela International Federation of Library Associations and Institutions (2009), aos

usuários deve ser possível, por meio da *navegação* entre os registros de um catálogo, *encontrar* documentos que atendam às suas expectativas, de modo a ser *confirmada* a correspondência entre a entidade descrita e a procurada e a *escolha* do recurso a ser *obtido*, independente de onde esteja localizado.

Estabelece-se, portanto, que a eficácia no processo de recuperação e acesso aos recursos informacionais depende tanto da qualidade dos registros bibliográficos, ao indicarem a relação entre entidades, atributos e relacionamentos, quanto da forma de apresentação dos resultados de pesquisa em catálogos, uma vez que esta é uma ação determinante para o sucesso no desenvolvimento de tarefas a serem desempenhadas pelos usuários.

Ressalta-se que, apesar de já se ter conhecimento de algumas iniciativas que consideram o desenvolvimento de catálogos baseados nos FRBR, apresenta-se, no tópico seguinte, um protótipo de ontologia baseado nas entidades e relacionamentos previstos nos FRBR.

4 ONTOLOGIA BASEADA NOS FRBR

Construir uma ontologia compatível tanto com os hábitos da comunidade quanto com as normas que possibilitam usufruir das capacidades oferecidas pelas tecnologias semânticas, é o primeiro passo para que as bibliotecas digitais possam mudar nossa forma de pesquisa, acesso e utilização dos recursos (DABROWSKI; SYNAK; KRUUK, 2009).

Nesta perspectiva, com base no método proposto por Ramalho (2010) propõe-se o desenvolvimento de uma ontologia baseada nas entidades, atributos e relacionamentos previstos nos FRBR, cujo escopo considera as tarefas a serem desempenhadas pelos usuários, e que inicia estudos para aplicação em catálogos online que possam favorecer melhor representação e interoperabilidade de registros bibliográficos, frente a crescente necessidade de compartilhamento de recursos.

De acordo com o método utilizado, após a definição do propósito e escopo da ontologia, realizou-se estudo com o intuito de identificar instrumentos tradicionalmente utilizados no âmbito da Ciência da Informação que pudessem servir de substrato teórico para a modelagem da ontologia. Para que uma ontologia represente de forma fidedigna o conhecimento de um determinado domínio, a definição das classes e conceitos exige cuidado especial na precisão conceitual e identificação de suas relações. Neste sentido, as entidades e relacionamentos previstos nos FRBR serviram de arcabouço teórico para a definição dos componentes estruturais da ontologia.

Conforme destacam Ramalho e Fujita (2011) podem ser considerados como componentes estruturais de uma ontologia as Classes/Subclasses; Propriedades Descritivas; Propriedades Relacionais e Regras/Axiomas. Tais componentes formam a *espinha dorsal* da ontologia, permitindo a representação formal dos elementos e relacionamentos do domínio modelado.

Na perspectiva de reuso de ontologias identificou-se na literatura vocabulários e outras ontologias que possuem relação com a proposta nesta pesquisa. Porém devido ao escopo baseado nas entidades e relacionamentos previstos nos FRBR, recorreu-se ao reaproveitamento de padrões de metadados já sedimentados na área de Catalogação.

Em relação à nomeação dos componentes da ontologia, com o objetivo de facilitar a compreensão e evitar possíveis problemas de interoperabilidade, seguiu-se a convenção de não utilizar acentuação, espaços e caracteres especiais. Definiu-se, também, a utilização de letras maiúsculas no início e em todas as palavras subsequentes contidas no nome das classes e subclasses, notação comumente denominada como *CamelBack*. O mesmo procedimento foi adotado para representar as propriedades relacionais da ontologia, com exceção do uso de letras minúsculas no início de seus nomes, diferenciando-as das classes. Quanto às propriedades descritivas, com o intuito de facilitar a assimilação dos componentes da ontologia, utilizou-se como padrão a denominação prevista nos FRBR, em língua inglesa, sem espaçoamento.

Conforme destacado anteriormente, para que todo o potencial das ontologias possa ser explorado, torna-se necessário sua formalização a partir de linguagens que possam ser processadas por máquinas. Deste modo, após as definições conceituais das etapas anteriores, realizou-se a implementação da ontologia e formalização das regras e axiomas de modo a possibilitar a realização de inferências automáticas.

De acordo com os objetivos da ontologia proposta elegeu-se a linguagem OWL-DL/XML para a implementação, dada a aderência aos padrões do W3C e às características que possibilitam maior decidibilidade lógica/computacional. Como ferramenta de apoio para a modelagem e formalização das regras e axiomas utilizou-se a ferramenta Protégé versão 4.3.0 (Build 304) e o motor de inferência padrão da ferramenta FaCT++.

Após a implementação, realizou-se a avaliação e a documentação da ontologia, utilizando-se dos próprios recursos da ferramenta *Protégé* (a partir da inserção de anotações em seus componentes). Os procedimentos de avaliação, podem ser divididos em duas categorias:

- *Avaliação da hierarquia de classes e componentes da ontologia:* realizada a partir dos critérios da metodologia OntoClean, (re)adequando ontologicamente os componentes estruturais da ontologia até a obtenção do modelo desejado;
- *Avaliação das regras e axiomas:* efetuada por meio do julgamento do resultado de buscas que se aproximem das necessidades informacionais de usuários em potencial, simuladas para avaliar a capacidade representação, recuperação e realização de inferências automáticas de acordo com os objetivos propostos.

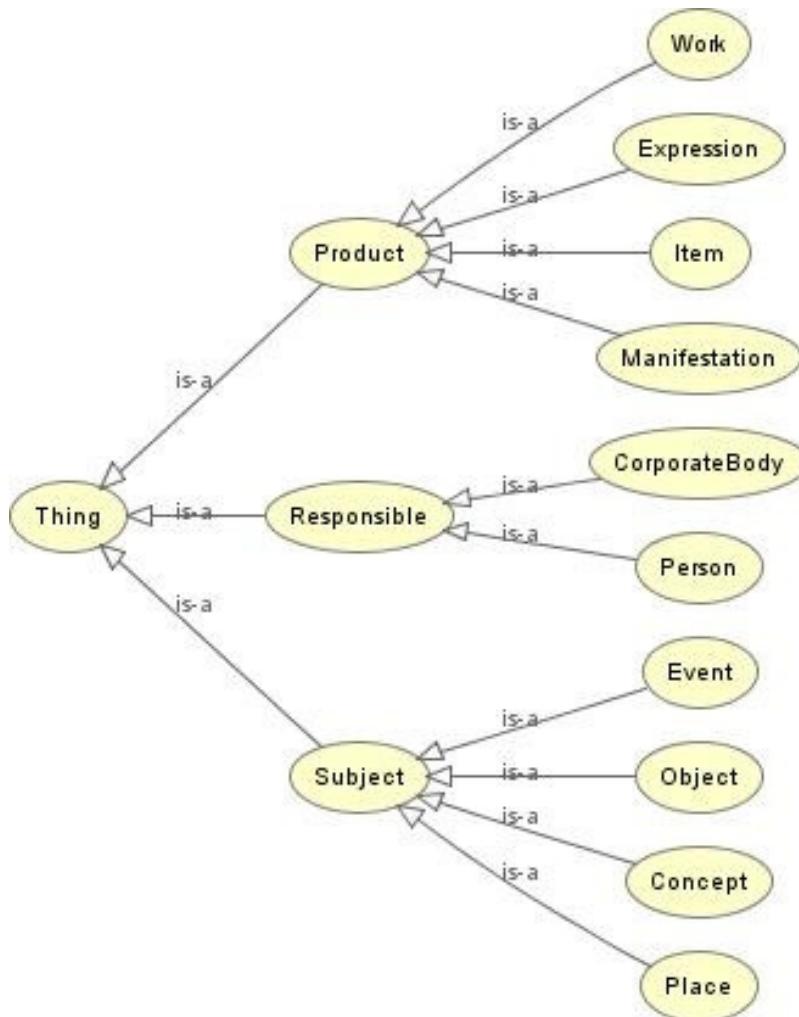
Tal processo foi realizado interativamente e de forma não-linear, por meio de refinamentos sucessivos, para se chegar ao resultado esperado. Nesta perspectiva, apresentam-se, nos próximos tópicos, alguns trechos sobre os componentes estruturais da ontologia baseada nos FRBR.

4.1 Estrutura de Classes

As Classes e Subclasses de uma ontologia são estruturadas por meio de uma taxonomia de conceitos, organizados a partir das características essenciais das instâncias representadas na ontologia e interligadas por relações hierárquicas. A relação de subordinação *is_a* (é_um) é a primitiva de estruturação mais utilizada no âmbito da Ciência da Informação e apresenta-se como uma forma bastante útil para representar a hierarquia de classes de uma ontologia, pois possibilita uma visão geral da ontologia modelada.

Assim, baseando-se nos princípios classificatórios utilizados na construção dos FRBR foram especificados os relacionamentos hierárquicos existentes entre classes da ontologia proposta, de modo que todas as propriedades existentes em uma Classe são *herdadas* por suas Subclasses, conforme apresentado na FIGURA 2, a seguir.

FIGURA 2 - Estrutura de Classes e Subclasses da ontologia



Fonte: Elaborado pelos autores

Conforme pode ser observado na figura 2 são apresentadas as classes que representam as 10 entidades previstas nos FRBR, organizadas em 3 hierarquias que refletem os 3 Grupos dos FRBR, as Subclasses são especializações das Classes, criadas para agrupar instâncias que possuem propriedades que não são comuns a todas as instâncias da classe superior, por este motivo graficamente as Subclasses são representadas como a origem de uma seta única direcionada para a classe superior.

4.2 Propriedades Descritivas

As propriedades descritivas são utilizadas para descrever as características, atributos e/ou qualidades das instâncias da ontologia de modo a possibilitar a atribuição de valores concretos. A utilização de propriedades descritivas aproxima-se do uso dos padrões e instrumentos da Catalogação.

QUADRO 1 – Propriedades descritivas da Classe *Work*

Atributo	Descrição
titleOfWork	Palavra, frase ou grupo de caracteres nomeando a obra. Pode haver um ou mais títulos associados com a obra. Se a obra foi publicada sob títulos diferentes (diferentes na forma, linguagem, etc.), uma agência bibliográfica normalmente seleciona um daqueles títulos como a base de um "título uniforme" para fins de consistência na nomeação, referenciando a obra. Outros títulos em que a obra tem aparecido podem ser tratados como títulos variantes ou em alguns casos, como um título uniforme paralelo. O título de uma obra que faz parte de um trabalho maior pode consistir apenas de um número ou outra designação genérica que é dependente do título da obra maior.
formOfWork	Classe à qual a obra pertence (por exemplo, a novela, jogo, poema, ensaio, biografia, sinfonia, concerto, sonata, mapa, desenho, pintura, fotografia, etc.)
publisherDistributor	Indivíduo, grupo ou organização identificada na manifestação como responsável pela publicação, distribuição, emissão, ou a liberação da manifestação. A manifestação pode estar associada a um ou mais editores ou distribuidores.
Numbering	Designação de volume ou fascículo, etc. e ou data que aparece na manifestação. A numeração pode compreender um componente numérico, alfabético e ou de uma data (por exemplo, volume 1, o número 1 (Janeiro 1971)).

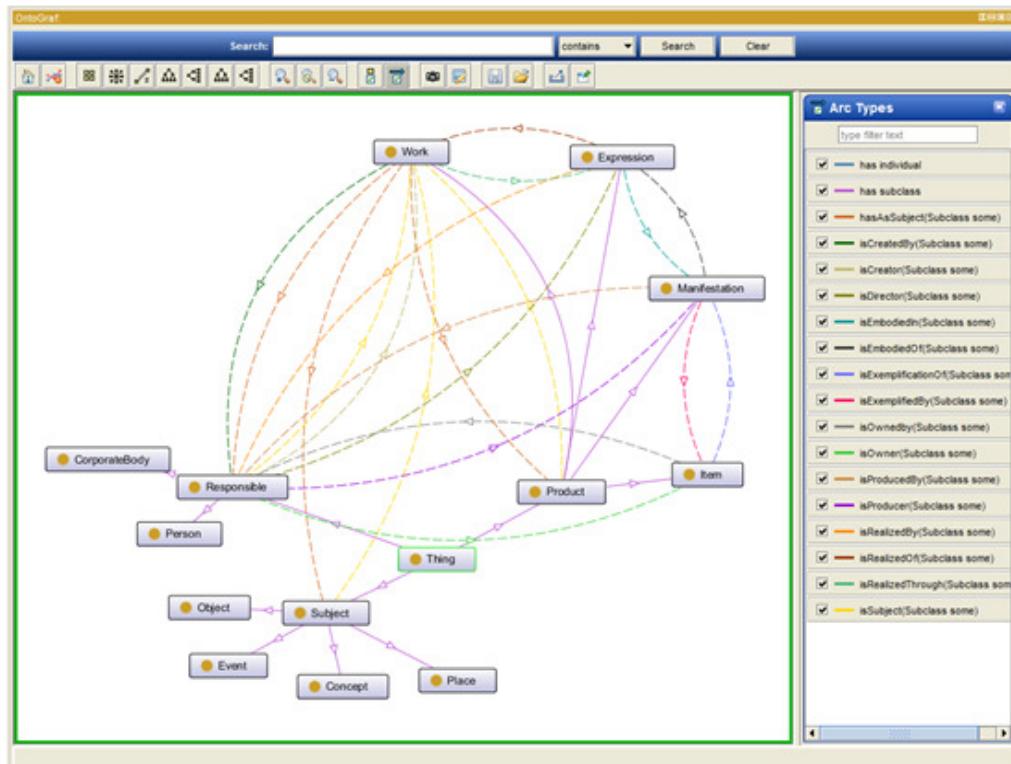
Fonte: Adaptado pelos autores com base nos FRBR (INTERNATIONAL FEDERATION OF LIBRARY ASSOCIATIONS AND INSTITUTIONS, 1998)

Na ontologia desenvolvida as propriedades descritivas representam os atributos das entidades previstas nos FRBR. O quadro 1 apresenta as principais propriedades descritivas da classe *Work* e suas respectivas descrições, por se tratar de um protótipo, com o objetivo de analisar a possibilidade de utilização dos FRBR como substrato para o desenvolvimento de ontologias, não foram representados todos os atributos das entidades.

4.3 Propriedades Relacionais

As Propriedades Relacionais permitem a definição de relacionamentos que vão além das relações hierárquicas presentes na estrutura de Classes e Subclasses. Essas propriedades caracterizam-se como relações ônticas e não-hierárquicas, possibilitando maior liberdade para a representação de relacionamentos entre as instâncias e favorecendo uma representação mais eficiente. As relações ônticas, segundo Dahlberg (1978), podem ser subdivididas em diversos tipos, podendo variar de acordo com o contexto. Nesta perspectiva, com vistas a possibilitar melhor compreensão da ontologia, apresenta-se, na figura 3, a representação gráfica das classes e das propriedades relacionais da ontologia baseada nos FRBR.

FIGURA 3 - Classes e Propriedades Relacionais da ontologia



Fonte: Elaborado pelos autores

Verifica-se que as propriedades relacionais representam os relacionamentos existentes entre objetos, sendo necessário indicar a classe origem, a classe destino e o rótulo do relacionamento, o qual normalmente é definido por meio de verbos (ex.: is, has), ou verbos preposicionados (ex.: IsRealizadBy).

4.4 Regras e Axiomas

As Regras e Axiomas são enunciados lógicos que possibilitam impor condições, restrições para a realização de inferências automáticas. Tais componentes apresentam-se como o principal avanço das ontologias frente aos demais instrumentos de representação.

De acordo com o escopo da ontologia as Regras e Axiomas foram elaborados para atender às tarefas dos usuários em um ambiente que tem por base os FRBR. Para sua definição são utilizadas sentenças descritas em Lógica de Primeira Ordem (LPO), conhecida também como Cálculo de Predicados de Primeira Ordem (CPPO), que possibilitam a definir formalmente a semântica dos formalismos.

QUADRO 2 – Descrição de Axiomas da ontologia

Descrição de Axiomas
Para cada expressão (<i>Expression</i>) pode haver “n” manifestações (<i>Manifestation</i>)
Para cada obra (<i>Work</i>) pode ter 1 ou mais instâncias do grupo do grupo 2 (<i>Responsible</i>) como responsáveis
Para cada obra (<i>Work</i>) pode haver “n” expressões (<i>Expression</i>)
Para cada manifestação (<i>Manifestation</i>) pode haver “n” itens (<i>Item</i>)
Para cada obra (<i>Work</i>) pode ter um ou mais elementos do grupo 3 (Subject) como assunto
Cada manifestação (<i>Manifestation</i>) pode ter um ou mais instâncias do grupo 2 (<i>Responsible</i>) como publicadores

Fonte: Elaborado pelos autores

No QUADRO 2 apresentam-se algumas regras implementadas na ontologia, tais regras consideram aspectos iniciais para a compreensão dos relacionamentos entre as entidades previstas nos FRBR.

5 CONSIDERAÇÕES FINAIS

A partir do estudo realizado foi possível verificar que as ontologias podem contribuir para uma melhor estruturação e representação de informações em catálogos bibliográficos online, possibilitando a realização de inferências automáticas visando facilitar aos usuários a identificação, seleção e acesso de recursos bibliográficos.

Constatou-se a partir do protótipo desenvolvido que a área de Ciência da Informação pode contribuir para o desenvolvimento de ontologias, a partir do aproveitamento de métodos e técnicas para a elaboração de ontologias, tesauros e vocabulários controlados presentes na literatura da área, favorecendo a aplicação conceitos e relacionamentos previstos nos Requisitos Funcionais dos Registros Bibliográficos em catálogos bibliográficos online.

O protótipo de ontologia desenvolvido baseou-se nas relações previstas entre as entidades dos FRBR e mostrou-se como uma alternativa viável para a representação de informações em meios computacionais, contudo é necessário uma análise ontológica

adequada, em especial na modelagem do Grupo 3 (assunto), onde muitas vezes é considerado que entidades do Grupo 1 (produto) e entidades do Grupo 2 (responsável) também fazem parte do Grupo 3, configurando uma modelagem ontológica incorreta. Destaca-se que apenas as entidades originais do Grupo 3 são subclasses da classe “Subject”, estabelecendo uma relação hierárquica de subordinação com esta classe. As entidades do Grupo 1 e Grupo 2 estabelecem relações ônticas com a classe “Subject”, representadas a partir de propriedade relacionais, não estando subordinadas hierarquicamente a esta classe, pois cada um dos Grupos dos FRBR possuem diferentes critérios de identidade, caracterizando polissemia.

Nesta perspectiva, verifica-se que a utilização de ontologias favorece a transposição dos métodos convencionais de representação, contribuindo para a realização de inferências automáticas, como, por exemplo: localizar um termo em diversas fontes de informação, detectar termos genéricos e equivalentes, eliminar ambiguidades, identificar hipônimos, categorizar recursos bibliográficos, entre outras.

Assim, as ontologias possibilitam ir além da representação dos aspectos descritivos e temáticos dos documentos, fornecendo subsídios computacionais que permitem representar os próprios domínios, favorecendo a recuperação de informações contextualizadas. Deste modo, as possibilidades profissionais apresentam condições de ampliação por meio da incorporação de novas potencialidades aos processos de representação, estruturação, disseminação e recuperação de informações em ambientes digitais.

REFERÊNCIAS

ANSI/NISO. **Z39.19 – 2005**: Guidelines for the construction, format, and management of monolingual controlled vocabularies. Bethesda: NISO Press, 2005.

BUSH, V. As we may think. **Atlantic Monthly**, v. 176, n. 1, p. 101-108, 1945. Disponível em: <<http://www.theatlantic.com/magazine/archive/1969/12/as-we-may-think/>>. Acesso em: 10 jun. 2014.

CAMPOS, M. L. A.; GOMES, H. E. Metodologia de elaboração de tesauro conceitual: a categorização como princípio norteador. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 11, n. 3, set/dez. 2006. Disponível em: <<http://www.eci.ufmg.br/pcionline/index.php/pci/article/viewFile/273>>. Acesso em: 20 jun. 2014.

DABROWSKI, M; SYNAK, M.; KRUK, S. R. Bibliographic Ontology. In: KRUK; McDANIEL (Eds.). **Semantic Digital Libraries**. Berlin: Springer, 2009.

DAHLBERG, I. **Ontical structures and universal classification**. Bangalore: Sarada Ranganathan Endowment, 1978.

FÉRNANDEZ, M.; GÓMEZ-PÉREZ, A.; JURISTO, N. **Methontology**: From Ontological Art Towards Ontological Engineering. Palo Alto: Association for the Advancement of

Artificial Intelligence, 1997. AAAI Technical Report SS-97-06. Disponível em: <http://oa.upm.es/5484/1/METHONTOLOGY_.pdf>. Acesso em: 20 jun. 2014.

GARCÍA MARCO, F. J. Ontologías y organización del conocimiento: retos y oportunidades para el profesional de la información. **El profesional de la información**, v. 16, n. 6, p. 541-550, Nov.-Dic. 2007. Disponível em: <<http://www.elprofesionaldelainformacion.com/contenidos/2007/noviembre/01.pdf>>. Acesso em: 10 jan. 2014.

GIL, A. C. **Como elaborar projetos de pesquisa**. 4. ed. São Paulo: Atlas, 2007.

GOMES, H. E. et al. **Manual de elaboração de tesouros monolíngues**. Brasília: CNPq/PNBU, 1990.

GRUBER, T. R. Toward principles for the design of ontologies used for knowledge sharing. In: GUARINO, N.; POLI, R. (Org.). **Formal Ontology in Conceptual Analysis and Knowledge Representation**. [s.l.]: Kluwer Academic Publishers, 1993. Disponível em: <http://pdf.aminer.org/000/912/413/toward_principles_for_the_design_of_ontologies_us_ed_for_knowledge.pdf>. Acesso em: 29 jun. 2014.

GUARINO, N. **Formal Ontology and Information Systems**. [Vilnius]: Vilnius University Faculty of Mathematics and Informatics, 1998. Disponível em: <<http://uosis.mif.vu.lt/~donatas/Vadovavimas/Temos/OntologiskaiTeisingasKoncepcinisModeliavimas/papildoma/Guarino98-Formal%20Ontology%20and%20Information%20Systems.pdf>>. Acesso em: 26 jul. 2014. Amended version of a paper appeared in N. Guarino (ed.), Formal Ontology in Information Systems. Proceedings of FOIS'98, Trento, Italy, 6-8 June 1998. Amsterdam, IOS Press, pp. 3-15.

INTERNATIONAL FEDERATION OF LIBRARY ASSOCIATIONS AND INSTITUTIONS. **Declaração de princípios internacionais de catalogação**. [Haia]: IFLA, 2009. Disponível em: <http://www.ifla.org/files/cataloguing/icp/icp_2009-pt.pdf>. Acesso em: 20 jun. 2011.

INTERNATIONAL FEDERATION OF LIBRARY ASSOCIATIONS AND INSTITUTIONS. **Functional requirements for bibliographic records**: final report. [Haia]: IFLA, 1998. Disponível em: <http://archive.ifla.org/archive/VII/s13/frbr/frbr_2008.pdf>. Acesso em: 20 jun. 2011. Aprovado pelo Standing Committee of the IFLA Section on Cataloguing em setembro de 1997 e alterado e corrigido até fevereiro de 2009.

INTERNATIONAL FEDERATION OF LIBRARY ASSOCIATIONS AND INSTITUTIONS. **Guidelines for Online Public Access Catalogue (OPAC) displays**: final report: may 2005. München: K. G. Saur, 2005. Recommended by the Task Force on Guidelines for OPAC Displays. Approved by the Standing Committee of the IFLA Cataloguing Section.

NOY, N. F.; MCGUINNESS, D. L. **Ontology development 101**: a guide to creating your first ontology. Stanford: Standford University, [2001?]. Disponível em: <<http://www.ksl.stanford.edu/people/dlm/papers/ontology101/ontology101-noy-mcguinness.html>>. Acesso em: 20 fev. 2014.

RAMALHO, R. A. S. **Desenvolvimento e utilização de ontologias em bibliotecas digitais:** uma proposta de aplicação. 2010. 145f. Tese (Doutorado em Ciência da Informação) - Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília, 2012.

RAMALHO, R. A. S.; FUJITA, M. S. L. Aplicabilidad de ontologías en bibliotecas digitales. **Anales de Documentación**, v. 14, n. 1, 2011. Disponível em: <<http://revistas.um.es/analesdoc/article/view/120161>>. Acesso em: 12 mar. 2014.

RANGANATHAN, S. R. **As cinco leis da biblioteconomia.** Brasília: Briquet de Lemos, 2009.

SANTOS, M. T.; CORREA, R. F.; SILVEIRA, M. A. A. Estudos brasileiros sobre ontologia na Ciência da Informação. **Datagramazero**, v. 14, n. 1, 2013. Disponível em: <http://www.dgz.org.br/fev13/Art_05.htm>. Acesso em: 20 jun. 2014.

SANTOS, P. L. V. A. C.; CORRÊA, R. M. R. **Catalogação:** trajetória para um código internacional. Niterói: Intertexto, 2009.

SILVEIRA, D. T.; CÓRDOVA, F. P. A pesquisa científica. In: GERHARDT, T. E.; SILVEIRA, D. T. (Org.). **Métodos de pesquisa.** Porto Alegre: Editora da UFRGS, 2009.

SOERGEL, D. The rise of ontologies or the reinvention of classification. **Journal of the American Society for Information Science**, v. 50, n. 12, p. 1119-1120, Oct. 1999. Disponível em: <<http://www.dsoergel.com/cv/B70.pdf>>. Acesso em: 10 jan. 2014.

SVENONIUS, E. The Intellectual Foundation of Information Organization. Cambridge: MIT Press, 2000.

TRIVIÑOS, A. N. S. **Introdução à pesquisa em ciências sociais:** a pesquisa qualitativa em educação. São Paulo: Atlas, 1987.

USCHOLD, M.; GRUNINGER M. Ontologies: principles, methods and applications. **The Knowledge Engineering Review**, v. 11, n. 2, p. 93-136, 1996.

USCHOLD, M.; KING, M. Towards a Methodology for Building Ontologies. In: IJCAI-95: Workshop on Basic Ontological Issues in Knowledge Sharing, 1995, Montreal, Canada. **Proceedings**. 1995. Disponível em: <http://www1.cs.unicam.it/insegnamenti/reti_2008/Readings/Uschold95.pdf>. Acesso em: 20 jun. 2014.

VICKERY, B. C. Ontologies. **Journal of Information Science**, v. 23, n. 4, p. 277–286, 1997.

ZAFALON, Z. R. Saussure e a representação da informação como um dos modos de comunicação para e com usuários de instituições de patrimônio cultural. In: BOCCATO, V. R.C.; GRACIOSO, L. S. (Org.). **Estudos de linguagem em ciência da informação.** Campinas: Alínea, 2011.

ZAFALON, Z. R. **Scan for MARC:** princípios sintáticos e semânticos de registros bibliográficos aplicados à conversão de dados analógicos para o Formato MARC21 Bibliográfico. 2012. 170f. Tese (Doutorado em Ciência da Informação) - Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília, 2012.

ZAFALON, Z. R. Simbiose entre catalogação e recursos tecnológicos: associações entre Ciência da Informação e Linguística pelo viés dos registros bibliográficos automatizados. In:

ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 14., 2013, Florianópolis. [Anais eletrônicos]. 2013. Disponível em: <<http://enancib2013.ufsc.br/index.php/enancib2013/XIVenancib/paper/download/496/32>>. Acesso em: 20 maio 2014.

ZHAO, Y; HU, C. An Ontology-based Framework for Knowledge Service in Digital Library. In: INTERNATIONAL CONFERENCE ON WIRELESS COMMUNICATIONS, NETWORKING AND MOBILE COMPUTING, 2007, Shangai. **Proceedings...** Palo Alto: IEEE, 2007. p. 5345-5348.

AVALIAÇÃO DO ACESSO E VISUALIZAÇÃO DA INFORMAÇÃO EM REPOSITÓRIOS INSTITUCIONAIS

*EVALUATION OF ACCESS AND VISUALIZATION INFORMATION IN INSTITUTIONAL
REPOSITORIES*

Sandra de Albuquerque Siebra
 Júccia Nathielle do Nascimento Oliveira
 Críssia de Santana Marcelino

Resumo: A ideia do Repositório Institucional como dispositivo de memória acadêmica surgiu da necessidade de garantir a preservação de publicações, favorecer o intercâmbio de informações e para dar visibilidade à produção das instituições, aos seus pesquisadores e órgãos de fomento. Porém, a divulgação da produção científica só se efetiva se a informação for disseminada por uma rede de comunicação eficaz e que propicie fácil acesso, visualização e uso dos diversos artefatos disponibilizados. Neste cenário, essa pesquisa objetivou investigar como os pressupostos teóricos e pragmáticos da Arquitetura, Acessibilidade e Usabilidade da Informação estão sendo utilizados visando facilitar o acesso, busca e visualização de informações em Repositórios Institucionais. Em termos metodológicos esta pesquisa é descritiva e bibliográfica. Além de ser uma investigação experimental, que faz uso do método de estudos de casos múltiplos, onde cada repositório é analisado separadamente e os resultados finais são comparados. Com relação a abordagem de análise dos dados, esta pesquisa se caracteriza como qualitativa e quantitativa. O universo da pesquisa engloba os repositórios institucionais das universidades brasileiras, mas para efetivação da pesquisa tomou-se como amostra um repositório institucional por capital brasileira. O resultado da avaliação mostrou que uma parte das universidades ou não tem repositório institucional ou possui problemas em manter ativo o repositório existente, visto que alguns repositórios passam períodos indisponíveis. Os repositórios ativos apresentam problemas de usabilidade e de interação com o usuário, especialmente no que concerne a funcionalidade de busca, uma das mais importantes nesse tipo de sistema. Além de apresentarem alguns caminhos confusos na navegação até chegar a localização de um documento. Essa pesquisa possibilita uma visão mais global de como os repositórios brasileiros estão promovendo o acesso a informação e a interação com os usuários.

Palavras-chave: Repositório Institucional. Usabilidade. Arquitetura da Informação. Acessibilidade.

Abstract: The idea of the Institutional Repository as academic memory device arose from the need to ensure the preservation of publications, promote the exchange of information and to give visibility to the production of the institutions, their researchers and funding agencies. However, the dissemination of scientific production is only effective if the information is disseminated by an effective communication network that provides easy access, display and use of the various artifacts available. In this scenario, this research aimed to investigate how the theoretical and pragmatic assumptions of Architecture, Accessibility and Usability of information are being used to facilitate access, search and visualization of information in Institutional Repositories. Methodologically this research is descriptive and bibliographical. Besides being an experimental research that makes use of the method of multiple case studies where each repository is analyzed separately and the final results are compared. With respect to data analysis approach, this research is characterized as qualitative and quantitative. The research encompasses institutional repositories of Brazilian universities, but to execute the research was taken as sample an institutional repository by each Brazilian capital. The

evaluation result showed that some of the universities or do not have an institutional repository or has problems maintaining the existing repository active, since they occasionally become unavailable. Assets repositories have usability and user interaction problems, especially regarding search functionality, one of the most important in this type of system. Besides having some confusing navigation paths to reach the location of a document. This research enables a more global view of how the Brazilians repositories are promoting access to information and interaction with its users.

Keywords: Institutional Repository. Usability. Information Architecture. Accessibility.

1 INTRODUÇÃO

Nos últimos anos a produção científica cresceu substancialmente e esse crescimento tem contribuído para o desenvolvimento da ciência e proporcionado a disseminação do conhecimento pelas diversas áreas científicas. Ressalta-se que a comunicação da produção científica, de fato só se efetiva se a informação for disseminada por uma rede de comunicação eficaz, que propicie o acesso democrático e facilidades para visualização e uso dos diversos artefatos produzidos. O acesso livre e democrático traduz a disponibilização pública na *Internet*, de forma a permitir que o usuário possa ler, realizar *download*, fazer cópia, distribuir e imprimir textos científicos completos, sem custo associado.

Porém, o acesso democrático ao conhecimento produzido pelas instituições científicas tem se tornado complexo em função do alto custo de produção e aquisição dos periódicos onde, em geral, o conhecimento produzido é divulgado. Assim, inseridos na dinâmica do movimento de acesso livre à informação, surgiram os repositórios institucionais (RI). Eles reúnem toda a produção científica e/ou acadêmica de uma instituição produzida em formato digital ou digitalizada. Os RI constituem uma inovação no âmbito da comunicação científica e no modo como a informação resultante das atividades acadêmicas e científicas é gerenciada e disponibilizada em rede. É uma iniciativa que constrói as condições necessárias para permitir o acesso livre à produção científica de forma legítima, como também é uma alternativa para guarda e visualização da memória de uma instituição, alterando seu processo de produção, disseminação e uso da informação (BLATTMAN; BOMFÁ, 2006).

Porém, ao se visitar Re却itórios Institucionais diferentes é possível perceber uma diversidade de interfaces de comunicação com os usuários e de identidade visual. Porém, é possível perceber, também, problemas bem semelhantes tais como: uso inadequado de vocabulário, conteúdos redundantes, falta de opções de personalização e problemas de acessibilidade da informação. Assim, neste contexto, esta pesquisa objetivou investigar como os pressupostos teóricos e pragmáticos da Arquitetura, Acessibilidade e Usabilidade da Informação podem contribuir para aprimorar a navegação, acesso, busca e visualização de informações em Re却itórios Institucionais.

Esta pesquisa pode contribuir para que seja facilitado o acesso à informação para a comunidade em geral, de forma democrática e independente de limitações. Promovendo, assim, a divulgação da produção científica das instituições através de seus repositórios institucionais Adicionalmente, esta pesquisa pode colaborar com os pressupostos do acesso livre a informação e do movimento do acesso aberto (MUELLER, 2006)

2 REPOSITÓRIO INSTITUCIONAL

Para Viana e colegas (2005), um repositório digital é uma forma de armazenamento de objetos digitais que tem a capacidade de manter e gerenciar materiais por longos períodos de tempo e prover o acesso apropriado a esses materiais. Em outras palavras, repositórios digitais (RDs) são bases de dados *online* que reúnem, de maneira organizada, a produção científica, em diversos formatos, de uma instituição (repositórios institucionais) ou área temática (repositórios temáticos) (IBICT, 2014).

Dessa forma, os repositórios institucionais, foco desta pesquisa, podem melhorar a comunicação científica e o avanço da pesquisa científica por permitir aos usuários localizarem e recuperarem informações relevantes de forma rápida e fácil. (MEDINA, 2006). Porém, trazem a responsabilidade para a instituição de exercer o controle sobre sua produção intelectual, tornando-a acessível e facilmente recuperável, além de garantir a preservação de seu conteúdo (MEDINA, 2006).

3 USABILIDADE

Segundo a norma ISO/IEC 9126 (2001, p.9), "usabilidade refere-se à capacidade de uma aplicação ser compreendida, aprendida, utilizada e agradável/atraente para o usuário, em condições específicas de utilização". E, ainda defende que usabilidade é a eficiência e satisfação com que um produto permite atingir objetivos específicos de utilizadores específicos em um contexto de utilização específico (ABNT, 2002). Sendo assim, uma característica relevante para qualquer sistema interativo.

Para realizar a avaliação de usabilidade em sistemas interativos, podem ser usados critérios e recomendações tais como: os requisitos não funcionais de usabilidade (PRESSMAN, 2004); as oito regras de Ouro para o aprimoramento da interação de Ben Shneiderman (SHNEIDERMAN; PLAISANT, 2004); os Critérios Ergonômicos de Bastien e Scapin (1993); os Princípios de Diálogo da norma NBR NBR 9241-11:2002 (ABNT, 2002) ou as dez Heurísticas de Nielsen (1993; 1999). Essa última por ser uma das formas de avaliação mais utilizadas e reconhecidas (WINCKLER; PIMENTA, 2002) e ser a adotada para realização desta pesquisa, será descrita a seguir (NIELSEN 1993; 1999):

Visibilidade - o sistema deve informar continuamente ao usuário sobre o status da execução de operações e funcionalidades;

Compatibilidade entre o sistema e o mundo real- as informações devem ser organizadas conforme o modelo mental do usuário;

Controle e liberdade para o usuário- o usuário deve controlar o sistema. Ele deve poder, a qualquer momento, cancelar uma tarefa, desfazer uma operação e retornar ao estado anterior ou retornar a um caminho anterior de navegação;

Consistência e padrões- a mesma operação deve ser apresentada na mesma localização e deve ser formatada/apresentada da mesma maneira em diferentes telas, para facilitar o reconhecimento. Adicionalmente, deve-se adotar um padrão de distribuição de elementos visuais e de nomenclatura;

Prevenção de erros – dar subsídios (usar recursos ou oferecer mensagens) aos usuários para contornar erros cometidos e evitar que novos erros ocorram;

Reconhecimento em lugar de lembrança- o sistema deve mostrar os elementos de diálogo e permitir que o usuário faça suas escolhas, sem a necessidade de lembrar um comando específico;

Flexibilidade e eficiência de uso – é preciso ter formas diversificadas de realizar a mesma operação (ex: por ícone, por opção de *menu*, por teclas de função);

Projeto minimalista e estético - deve-se apresentar exatamente a informação que o usuário precisa no momento, evitando sobrecarga informacional;

Auxiliar os usuários a reconhecer, diagnosticar e recuperar erros – deve ser possível para o usuário entender e resolver o problema que ocasionou o erro e

Ajuda e documentação – o sistema deve oferecer ajuda *on-line* para orientar o usuário na realização de operações.

4 ARQUITETURA DA INFORMAÇÃO

Camargo (2010) defende que a Arquitetura da Informação (AI) é um campo que oferece subsídios teóricos e metodológicos adequados para o desenvolvimento de ambientes informacionais digitais mais funcionais e focados nos usuários.

Em uma abordagem sistêmica a AI é a combinação de quatro subsistemas (QUADRO 1): organização ou estruturação, rotulagem ou representação¹²⁴, busca ou recuperação e navegação SIN, 2009).

QUADRO 1 – Repositórios das Universidades da Amostra

SUBSISTEMAS	DEFINIÇÃO
Sistema de Organização	Define o agrupamento e a categorização de todo o conteúdo informacional.
Sistema de Navegação	Especifica as maneiras de navegar, de se mover pelo espaço informacional e hipertextual. Sua função é indicar ao usuário sua localização e os caminhos que podem ser percorridos para chegar ao seu destino final. Nielsen e Loranger (2007) defendem que um bom sistema de navegação precisa responder a 3 perguntas básicas: <i>Onde estou? Onde estive? Aonde posso ir?</i>
Sistema de Rotulação	Estabelece as formas de representação, de apresentação, da informação definindo signos para cada elemento informativo.
Sistema de Busca	Determina as perguntas que o usuário pode fazer e o conjunto de respostas que irá obter.

Fonte: (MORVILLE; ROSENFELD, 2007, adaptado).

5 ACESSIBILIDADE

Acessibilidade refere-se à transposição de barreiras. Aplicada à *Internet* isto significa a eliminação de quaisquer obstáculos que impossibilitem ou dificultem o acesso das pessoas com deficiência, ao mesmo tempo em que pode beneficiar pessoas sem limitação (ex: idosos). Acessibilidade da informação digital faz referência a garantia de acesso a informação a qualquer tipo de usuário. Torres, Mazzoni e Alves (2002, p.3) afirmam que:

A acessibilidade no espaço digital consiste em tornar disponível ao usuário, de forma autônoma, toda a informação que lhe for franqueável, independentemente de suas características corporais, sem prejuízos quanto ao conteúdo da informação. É forma de democratizar o acesso à informação e de proporcionar a inclusão digital, fazendo com que as informações armazenadas em contextos digitais se torne disponível para toda e qualquer pessoa, independente de limitação.

Porém, na prática, a acessibilidade na *web* e em sistemas interativos ainda é deficiente, prejudicando a inclusão digital e o acesso à informação para as pessoas com deficiência.

Para avaliação de acessibilidade podem ser usados *softwares* denominados validadores automáticos. Validadores são *softwares* que pesquisam o código fonte do *site* que se quer

¹²⁴ Camargo e Vidotti (2003) compreendem como sistemas de representação a aplicação de metadados, vocabulários controlados e tesouros nos ambientes informacionais digitais.

validar e emitem um relatório contendo os erros e avisos de acordo com as Diretrizes de Acessibilidade desenvolvidas pela W3C – World Wide Web Consortium². Uma delas é o WCAG 1 ou Recomendações para a acessibilidade do conteúdo da Web versão 1. Os erros e avisos identificados nos sites são classificados de acordo com o nível de prioridade: 1 – prioridades que devem ser satisfeitas inteiramente, pois impossibilitam o acesso à informação; 2 – prioridades que deveriam ser satisfeitas, pois causam dificuldades a usuários ou grupo deles e 3 – prioridades que poderiam ser satisfeitas, pois não causam desconforto ou alguma dificuldade a alguns usuários ou grupo deles. Os validadores não dispensam a verificação humana e seu foco está mais voltado para questões relacionadas à deficiência visual, mas podem contribuir de maneira significativa para a avaliação de acessibilidade dos conteúdos da internet. Dois exemplos de validadores são o DaSilva (<http://www.dasilva.org.br/>) e o Hera (<http://www.sidar.org/hera/index.php.pt>).

6 METODOLOGIA

Quanto aos objetivos, esta pesquisa tem caráter descritivo. Segundo Gil (2008) as pesquisas descritivas possuem como objetivo a descrição das características de uma população, fenômeno ou de uma experiência. Quanto à fonte de dados pode-se considerar esta pesquisa como bibliográfica. Para Marconi e Lakatos (2007), a pesquisa bibliográfica oferece meios para definir, resolver, não somente problemas já conhecidos, como também explorar novas áreas onde os problemas não se cristalizam suficientemente, tendo como objetivo permitir ao pesquisador um reforço na análise de suas pesquisas ou manipulação de suas informações. A pesquisa envolve o método de estudos de casos múltiplos (GIL, 2008), pois cada repositório foi analisado isoladamente, a fim de ajudar a alcançar os objetivos da pesquisa. Inclusive fazendo o levantamento dos elementos sistêmicos de cada um dos repositórios, traçando dessa forma seu perfil. Considerando as abordagens de análise dos dados, esta pesquisa se caracteriza como quantitativa e qualitativa (GIL, 2008).

O universo da pesquisa são todos os repositórios institucionais das Universidades Federais Brasileiras. E, para efetiva realização da pesquisa, foram selecionados, de forma aleatória, apenas um repositório de Universidade por cada capital brasileira. Assim, como amostra, foram selecionados 27 (vinte e sete) repositórios institucionais de universidades brasileiras (QUADRO 2). A decisão por avaliar os repositórios das Universidades Federais se deu por estas serem instituições mantidas com dinheiro público, sendo assim direito da sociedade ter acesso à produção científica e acadêmica delas.

QUADRO 2 – Repositórios das Universidades da Amostra

REGIÃO	INSTITUIÇÃO (SIGLA)	ENDEREÇO DO RI
NORDESTE	UFAL	http://www.repositorio.ufal.br
	UFBA	http://www.repositorio.ufba.br/ri
	UFC	http://www.repositorio.ufc.br:8080/ri/
	UFMA	http://www.repositorio.ufma.br:8080/jspui/
	UFRN	http://repositorio.ufrn.br:8080/jspui/
NORTE	UFPE	http://www.repositorio.ufpe.br/
	UFPB	http://rei.biblioteca.ufpb.br/jspui/
	UFS	http://www.repositorio.biblioteca.ufs.br/
	UFAC	http://repositorios.ufac.br:8080/repositorio/
	UFPA	http://www.repositorio.ufpa.br/jspui/
CENTRO OESTE	UFMS	http://repositorio.cbc.ufms.br:8080/jspui/
SUDESTE	UFMG	https://dspaceprod02.grude.ufmg.br/dspace/ http://www.bibliotecadigital.ufmg.br/dspace/
	UFES	http://repositorio.ufes.br/
	UFSC	http://repositorio.ufsc.br/
SUL	UFPR	http://dspace.c3sl.ufpr.br/dspace/
	UFRGS	http://www.lume.ufrgs.br/

Fonte: Os autores, 2014.

Para alcançar os objetivos deste trabalho, em cada repositório foi primeiro realizada a avaliação de usabilidade segundo as Heurísticas de Nielsen (NIELSEN, 1993;1999). Depois, foram avaliados os subsistemas da arquitetura da informação, com base nos seguintes questionamentos:

Sistema de Organização

- 1) As informações disponibilizadas no repositório estão categorizadas ou classificadas em grupos (assunto, tema, tipo, ordem alfabética)?
- 2) O repositório apresenta documentos que são classificados em mais de uma categoria de classificação?
- 3) O site apresenta uma boa organização geral? (agradável de se ver e se usar)

Sistema de Navegação

- 1) O sistema de navegação responde as três perguntas básicas: Onde estive? Onde estou? Onde posso ir? (NIELSEN; LORANGER, 2007).
- 2) A navegação é simples (não se usam muitos cliques para se alcançar uma funcionalidade)?
- 4) São oferecidos recursos de internacionalização (ex: mudança de idioma)?

Sistemas de Rotulagem

- 1) A simbologia utilizada nos ícones representam adequadamente as funcionalidades oferecidas?
- 2) As nomenclaturas (vocabulário) e símbolos utilizados estão adequados ao público alvo?

Sistema de Busca

- 1) O sistema oferece busca? Se sim: que recursos são oferecidos na página de resultados (Ex: sugestão de palavras, paginação, ordenação dos resultados)
- 2) Existe busca avançada? Nela é feito uso de operadores booleanos?

Finalmente, para a avaliação de acessibilidade foram utilizados 2 validadores de acessibilidade: o DASilva e o HERA. Para isso, acessou-se o referido *site* e, posteriormente, digitou-se o endereço do *URL* de cada site a ser avaliado. Usou-se o padrão de avaliação WCAG 1.0. Adicionalmente, foi analisado, visualmente pelas pesquisadoras, se o site possuía algum recurso dos seguintes recursos relacionado à acessibilidade: aumento/diminuição do tamanho da fonte, possibilidade de mudança de cores e modificação do contraste.

As avaliações foram realizadas no período de fevereiro a junho de 2014, por três pesquisadores isoladamente, depois os pesquisadores se reuniram para discutir os resultados obtidos.

7 RESULTADOS E DISCUSSÕES

Para organização da apresentação dos resultados, eles foram subdivididos em 4 subseções: a primeira tratando de observações gerais sobre os repositórios da amostra e as seguintes abordando, em sequencia, as avaliações de usabilidade, arquitetura da informação e acessibilidade.

3.1 Avaliação Geral dos Re却itórios

A partir do levantamento dos repositórios institucionais (RIs) da amostra feito no QUADRO 2, ao acessar cada repositório, no período da avaliação, foi possível perceber que das 9 Unidades da Federação do Nordeste, apenas 1 não possui RI: a Universidade Federal do Piauí (UFPI). Na região Norte, das 7 instituições existentes, apenas 2 possuem repositórios: a Universidade Federal do Acre (UFAC) e a Universidade Federal do Pará (UFPA) e na região Centro-oeste apenas a Universidade Federal do Mato Grosso do Sul (UFMS). Na região Sudeste, das 4 IES, apenas a Universidade Federal de Minas Gerais (UFMG) e a Universidade Federal do Espírito Santo (UFES) tem RIs e na região Sul todas as IES, Universidade Federal

do Paraná, Universidade Federal de Santa Catarina e a Universidade Federal do Rio Grande do Sul (UFPR, UFSC e UFRGS), possuem RI.

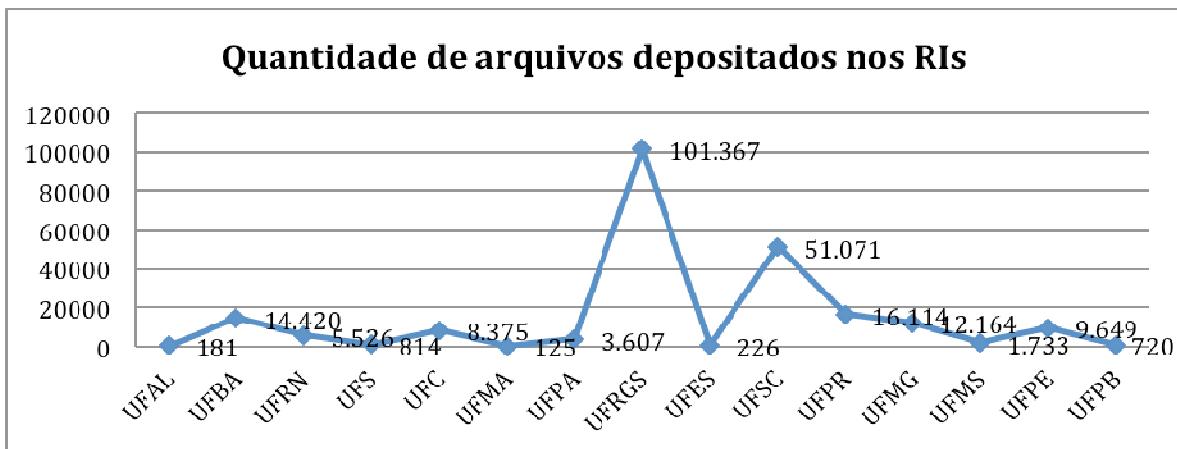
Vale ressaltar que os repositórios da Universidade Federal de Pernambuco (UFPE) e da Universidade Federal da Paraíba (UFPB) não estavam disponíveis durante o período da avaliação. Os dados referentes a esses dois repositórios foram acrescentados em período posterior (setembro/2014). Além disso, durante a maior parte do desenvolvimento da avaliação, verificou-se instabilidades em alguns repositórios: o RI da UFMA estava no ar, mas não estava funcionando, voltando a funcionar apenas a partir do início de março/2014. O RI da UFAC tornou-se indisponível (saiu do ar) e não mais se tornou disponível durante todo o período de avaliação ou da produção deste artigo. O RI da UFS esteve no ar até o mês de fevereiro/2014, depois saiu do ar e ficou inacessível, voltando a entrar no ar na segunda quinzena de abril/2014. O RI da UFMS apresentou instabilidades diárias, acompanhadas durante o período da avaliação e, em junho/2014, houve vários dias que o RI não estava no ar.

Para fazer um reconhecimento dos repositórios, como funcionam e como estão organizados, foi feito, inicialmente, o acesso a cada repositório, simuladas buscas, além de ser feita navegação pela diversidade de acervos disponíveis. Também se buscou informação sobre o repositório e a equipe responsável por ele que, em sua maioria, era multidisciplinar. O GRÁFICO 1 apresenta a quantidade total de arquivos depositados nos Repositórios Institucionais analisados, onde se destacam o repositório da UFRGS e da UFSC.

Dentre os arquivos depositados nos repositórios destacam-se trabalhos apresentados em eventos/congressos, livros e capítulos de livros, artigos, Trabalhos de Conclusão de Curso (TCC), Dissertações e Teses, sendo os três últimos tipos os mais frequentemente encontrados.

Todos os repositórios foram criados fazendo uso do *software DSPACE* e, talvez por causa disso, possuem design muito similar e apresentam todos praticamente os mesmos problemas, como poderá ser visualizado nas subseções a seguir.

GRÁFICO 1 – Quantidade de Arquivos dos RIs das Universidades da Amostra



Fonte: (OPENDOAR, @2006-2014, adaptado)

Legenda: Dados levantados em 13/09/2014

3.1 Avaliação de Usabilidade

Com os perfis dos repositórios traçados, cada repositório foi visitado e avaliado segundo as Heurísticas de Nielsen, destacando pontos fortes e fracos que os repositórios possuíam relacionados a cada heurística. Os resultados da avaliação de usabilidade nos repositórios pode ser visualizada no QUADRO 3. Na avaliação foi gerado um quadro para cada repositório. Porém, aqui, por questões de espaço e pelo fato de que os resultados das avaliações terem sido bem parecidos na maioria dos repositórios, a avaliação de usabilidade foi sumarizada em um único quadro.

QUADRO 3 – Sumarização da Avaliação de Usabilidade nos Repositórios da Amostra

Nº	CRITÉRIO	PONTO(S) FORTE(S)	PONTO(S) FRACO(S)
1	Visibilidade do estado do sistema	Carregamento da página inicial é rápido. As opções do que pode ser feito estão visíveis.	Em vários repositórios não é dado <i>feedback</i> quando alguma operação é realizada. Ex: busca.
2	Mapeamento entre o sistema e o mundo real	A nomenclatura utilizada é clara.	Nenhum dos menus ou submenus possuem equivalentes iconográficos.
3	Liberdade e controle do usuário	A qualquer momento o usuário consegue voltar para a página inicial, como também escolher outra opção de busca, uma vez que os <i>menus</i> se apresentam sempre visíveis. Nos resultados da busca, o usuário tem liberdade para optar pela exibição dos resultados: por relevância, data ou título.	Na hora de fazer o <i>download</i> do arquivo encontrado é necessário primeiro abrir o documento, tornando o processo de <i>download</i> mais demorado. Na maioria dos repositórios, não há como navegar entre os documentos existentes em uma coleção/acervo. Se começa navegando ao escolher o centro/acervo, depois, obrigatoriamente, uma busca deve ser utilizada.
4	Consistência e padrões	A maioria das coleções/acervos são organizados, em geral, por seu centro/departamento de	Quando o usuário opta por uma busca que não seja pelas “Comunidades e Coleções” os RIs apresentam o caminho até o

Nº	CRITÉRIO	PONTO(S) FORTE(S)	PONTO(S) FRACO(S)
		origem.	documento (“migalhas de pão”) diferente do que foi percorrido pelo usuário, confundido e retardando a navegação. A maioria dos RIs quando são visualizados em um idioma que não o português, há opções de <i>menu</i> que desaparecem e nem todo o texto dos <i>menus</i> é traduzido, ficando uma mistura de idiomas confusa. Talvez por fazerem uso de tradução automática.
5	Prevenção de erros		Na busca por data não é indicado o formato do ano. Na grande maioria dos RIs, quando um documento é cadastrado, não é verificado se o autor já existe na base de dados. Isso ocasiona que o mesmo autor possa ser cadastrado diversas vezes, devido a digitação diferente do seu nome (ex: uso de abreviatura, espaçamento, ponto). Isso tornará difícil e imprecisa a busca por autor.
6	Reconhecer em vez de relembrar	A estrutura de <i>menus</i> facilita o reconhecimento das opções disponíveis para navegação.	A utilização de ícones ajudaria o usuário a reconhecer o lugar, facilitando a navegação sempre que ele voltasse ao ambiente.
7	Flexibilidade e eficiência de uso	Os documentos podem ser localizados pela busca ou pelas opções de visualização disponíveis (por data de publicação, autor, título ou assunto).	Não há a possibilidade de utilização de teclas de atalho. Não é possível navegar pelos repositórios fazendo uso exclusivo do teclado. A manipulação da tecla TAB não está bem configurada. Não é possível navegar pelos documentos de um acervo/centro específico, sem passar por uma busca.
8	<i>Design</i> estético e minimalista	Na maioria, o <i>design</i> é minimalista e bastante similar.	O <i>design</i> da maioria dos repositórios é pouco atraente (uso do <i>design</i> padrão da ferramenta DSPACE).
9	Supporte para o usuário reconhecer, diagnosticar e recuperar erros	Não foi possível avaliar esse item.	
10	Ajuda e documentação	Alguns poucos repositórios disponibilizam arquivos com informações para utilização.	Na maioria dos repositórios não existe ajuda disponível para o usuário sobre como fazer uso do repositório. O que há disponível é a ajuda padrão do DSpace e informações sobre esse software e apenas no idioma inglês.

Fonte: Os autores, 2014.

Para ilustrar alguns pontos, destacamos alguns exemplos de problemas encontrados. Com relação a heurística 2 – mapeamento entre o sistema e o mundo real, para alcançá-la é preciso que o repositório esteja adequado ao seu contexto de uso e ao seu público alvo. Neste

sentido, o repositório da UFPR acaba por não atender a esta heurística porque está no idioma inglês, apenas com links e algumas poucas informações em português, acabando por mesclar os idiomas, tornando-se confuso para seu público alvo.

Com reação à heurística 4 – consistência e padrões, em alguns RIs, como no da UFRN (FIGURA 1), não há padronização no formato da data na apresentação dos resultados da busca.

FIGURA 1 – Exemplo de falta de consistência na apresentação da data de publicação dos documentos no RI da UFRN

Pré-visualizar	Data de Publicação	Título	Autor(es)
	30-Nov- 2	Gestão judiciária: a escola empreendedora e o magistrado na formação de estratégia como um processo visionário para melhoria dos serviços judiciais	Fernandes, Walteres Verônica Saldanha
	30-Nov- 2	Paisagens arquitetônicas da cidade de João Pessoa/PB e seus significados: o perceptivo dos moradores	Bandeira, Sânia Érika Alves de Cárdenas
	30-Nov- 2	Alterações nos genes da E-caderina e β-catenina em adenoma pleomórfico e carcinoma adenóide cístico: estudo molecular e imuno-histoquímico	Cavalcante, Roberta Barroso
	30-Nov- 2	Ánalise da capacidade funcional e da distribuição regional da ventilação pulmonar em pacientes com doença de chagas	Oliveira, Georges Willeneuve de Sousa
	14-Jan- 12	fdafdas	-
	29-Mar- 212	Literatura e educação: o bullying nos contos de fada, uma discussão possível	Medeiros, Lívia Cristina Cortez Lula de
	21-Abr-1934	A hermenêutica brasileira	Gabriel, Silva; Debora, Nagazaki
	1985	Formação do educador no curso de pedagogia de Caicó-RN: reprodução ou transformação social	Araújo, Marta Maria de
	11-Apr-1985	Análise de linhas de fita não-homogêneas com camadas dielétricas anisotrópicas	Albuquerque, Maria Rosa Medeiros Lins de

Fonte: (UFRN, [2010])

Legenda: Dados levantados em Maio de 2014

Com relação à heurística 5 – prevenção de erros, na grande maioria dos RIs, como por exemplo nos da UFRN, UFC e UFS (FIGURA 2), não é evitado um cadastro repetido de autor, ocasionando duplicações diversas.

FIGURA 2 – Exemplos de problema com replicação do nome de um mesmo autor nos RIs da UFRN e da UFC

Alves, Janaína da Silva	Albuquerque Junior, José Alberto Alves de
Alves, José Antônio Barros Leal Reis	ALBUQUERQUE, Kelly Moreira de
Alves, Jósemery Araújo	ALBUQUERQUE, Lindomar
Alves Jr, C.	Albuquerque, Luiz Botelho
Alves Jr, Clodomiro	ALBUQUERQUE, Luiz Botelho
Alves Jr., Clodomiro	ALBUQUERQUE, Luiz Botelho de
Alves jr., Clodomiro	ALBUQUERQUE, Maria Elenir Botelho
ALVES JUNIOR, Clodomiro	Albuquerque, Maria Gílrene de Freitas
Apresentar os registros de 101 até 180 de 5842 < anterior próximo >	Albuquerque, Mariana Pires de Carvalho e

Fonte: (UFRN, [2010]; UFC, [2011])

Legenda: Dados levantados em Maio de 2014

Referente à heurística 8 – design estético e minimalista, os RIs com design mais aprimorado são os da UFRGS e o da UFMG, pois possuem uma tentativa de se diferenciarem dos demais.

3.2 Avaliação de Arquitetura da Informação (AI)

A organização geral dos repositórios analisados é praticamente a mesma, assim como a ordem dos *menus*, com poucas variações. Talvez porque todos utilizem o *software DSPACE* e as pessoas responsáveis pelos repositórios tenham tido as mesmas orientações para criação deles, fornecida pelo IBICT (Instituto Brasileiro de Informação em Ciência e Tecnologia). Os três repositórios que apresentam uma personalização maior e uma mudança na disposição dos elementos na tela são os da UFPA, UFRGS, UFMG e da UFMA. O restante da avaliação da AI foi feita com base nos seus sistemas.

No tocante ao subsistema de organização temos que os repositórios organizam seu conteúdo em acervos nomeados, na maioria, com o nome dos centros, departamentos, escolas, faculdades ou unidades organizacionais da instituição (ex: pró-reitorias) e a disposição destes é em ordem alfabética. O *menu* lateral da maioria dos repositórios analisados, às vezes, se tornam confusos, pois tanto incluem *links* que levam a informação, como incluem formas de visualização dos acervos (ex: Por Assunto, Autor, Tipos de Documento, etc).

Todos os repositórios possuem algum tipo de replicação de documento em categorias diferentes. Por exemplo, no acervo de um centro encontra-se a tese de um professor e na BDTD também há a mesma tese. O que não representa um problema, pois é uma forma de tornar o documento mais fácil de localizar. Não foi possível verificar se são apenas dois *links* (duas formas de acesso) para o mesmo documento ou se o documento está realmente duplicado dentro do repositório. Em alguns repositórios há documentos replicados em mais de dois acervos.

Uma observação pertinente é sobre a organização dos arquivos pertencentes a UFMG. A instituição tem dois repositórios, um intitulado ‘Biblioteca Digital (BD)’ onde estão depositadas Teses, Dissertações, Monografias e trabalhos de Especializações; e outro site denominado ‘Repositório Digital (RD)’, onde estão outros tipos de documentos (ex: artigos), o que foge a estrutura e organização dos demais RIs.

Com relação ao subsistema de navegação, todos os repositórios disponibilizam o recurso de navegação chamado “migalhas de pão” (*links* que mostram o caminho percorrido pelo usuário). Todavia o caminho registrado nas migalhas de pão, curiosamente, não corresponde ao percorrido pelo usuário, especialmente se for feito uso da busca. Em geral, ao se chegar a um documento, a migalha de pão não apresenta o caminho percorrido para chegar ao documento, mas apenas a comunidade do qual o documento faz parte. Em geral, nos repositórios, o usuário leva de 3 a 5 cliques para alcançar um documento qualquer dentro dos

acervos, pela navegação convencional, a partir de um critério selecionado (ex: por autor, data, assunto ou título).

Existe nos repositórios da UFRGS e da UFBA uma navegação por tipo de documento que não existem nos demais repositórios. Porém, os tipos cadastrados no RI da UFBA são duplicados ou confusos (vide FIGURA 3). Por exemplo, fica-se na dúvida sobre qual é a diferença entre o tipo “Livro” e “Produção Bibliográfica: Livros” e isso pode provocar cadastramentos duplicados ou inadequados. Também verifica-se (FIGURA 3) um número elevado de cadastros no tipo documental “Outros” (849 registros), porém, ao verificar o conteúdo desse item, a grande maioria é de projetos PIBIC (Programa Institucional de Bolsas de Iniciação Científica) o que poderia indicar a necessidade de criação dessa categoria específica. Já o RI da UFRGS dispõe 36 tipos de documentos, mas não apresenta redundâncias na descrição e representação de seus tipos.

FIGURA 3 – Navegação por tipo de documento no repositório da UFBA

Navegando por Tipo de documento

The screenshot shows a search interface with a navigation bar at the top. The search bar has placeholder text "Ir para: 0-9 A B C D E F G H I J K L M N O P Q R S T U V W X Y Z ou entre com as primeiras letras:" followed by a search button. Below the search bar are buttons for "Ordem:" (set to "Ascendente"), "Resultados/Página" (set to 20), and "Atualizar". A message below the buttons says "Mostrando resultados 1 a 20 de 54" and a "Próximo >" link. The main content area lists document types with their counts:

- [ALUNOS QUEREM CURSO \[1\]](#)
- [Artigo de Evento \[20\]](#)
- [Artigo de Revista \[24\]](#)
- [Livro \[117\]](#)
- [Outros \[849\]](#)
- [Produção bibliográfica: Apresentação de trabalho e palestras \[4\]](#)
- [Produção bibliográfica: Apresentação de trabalho e palestras \[10\]](#)
- [Produção bibliográfica: Artigos aceitos para publicação \[1\]](#)
- [Produção bibliográfica: Artigos aceitos para publicação \[1\]](#)
- [Produção bibliográfica: Artigos aceitos para publicação \[12\]](#)
- [Produção bibliográfica: Artigos completos publicados em periódicos \[2118\]](#)
- [Produção bibliográfica: Artigos completos publicados em periódicos \[85\]](#)
- [Produção bibliográfica: Artigos completos publicados em periódicos \[4043\]](#)
- [Produção bibliográfica: Artigos completos publicados em periódicos \[1\]](#)
- [Produção bibliográfica: Artigos completos publicados em periódicosRevista de Ciências Médicas](#)
- [Produção bibliográfica: Capítulos de livros \[62\]](#)
- [Produção bibliográfica: Livros \[137\]](#)

Fonte: (UFBA, [2010])

Outa característica peculiar é que dentre os 15 Ris analisados, apenas o da UFRGS possui mapa do *site*, ajudando o usuário a ter uma visão geral das funcionalidades disponíveis. A UFRGS também dispõe do *link* ‘AJUDA’ onda há instruções de como navegar e fazer pesquisas no site, dicas para combinar termos e realizar uma revocação mais precisa. O *link* ‘AJUDA’ dispõe de ilustrações mostrando o passo a passo de como as buscas devem ser realizadas. (como mostra figura abaixo);o repositório da UFPA oferece um manual de Pesquisa que ajuda o usuário a definir a melhor forma possível de busca e o *link* para

perguntas frequentes. Em vários outros repositórios, como por exemplo os da UFBA, UFC, UFMS e UFES, a ajuda oferecida não corresponde a expectativa, visto que o usuário é direcionado para a página de ajuda do *software DSPACE* (em inglês) E, finalmente, o *link* de ajuda da UFRN não funciona, ao se clicar, nada acontece.

Quanto a questões de internacionalização, os repositórios da UFMA, UFMS, UFSC, UFMG, UFC e UFS e UFPE não oferecem qualquer opção nesse sentido. Porém, os demais que oferecem a opção de tradução do site para outros idiomas como inglês, espanhol e francês, não fazem essa tradução a contento (ex: partes do site continuam sem tradução ou os idiomas estrangeiros ficam misturados), como no repositório da UFES, UFBA e UFRN. Outro fato curioso é que na página traduzida para outro idioma, somem da tela diversos itens/recursos. Divergentemente dos demais, o RI da UFPR apresenta todas as funcionalidades em inglês, como idioma base, disponibilizando seus conteúdos em português. O *site* possibilita a tradução para o português, porém a tradução é confusa e incoerente, apresentando termos misturados. O que, inclusive, foge ao contexto do usuário, visto que é um repositório brasileiro.

No tocante ao subsistema de rotulação, praticamente não são utilizados ícones e o vocabulário dos menus dos repositórios é simples. Todos os repositórios consultados disponibilizam a busca convencional apenas com um único campo e na página de resultados oferecem os recursos de paginação e ordenação. Com exceção dos repositórios da UFAL, UFPE e UFBA, os outros repositórios apresentam também a busca avançada com o uso de operadores booleanos, com telas muito similares (FIGURA 4).

Finalmente, no repositório da UFBA há uma opção no lado direito da tela chamada “Busca Facetada” (vide FIGURA 5 – lado esquerdo). Porém, ela é bem confusa, inclusive seus botões de navegação ficam ocultos no meio do texto, não deixando claro como navegar entre as opções. No repositório da UFAL há o mesmo tipo de informações a direita, mas chamada de REFINAR (vide FIGURA 5 – lado direito), os botões estão um pouco mais destacados, ainda assim a navegação pelas opções não é clara para os usuários.

FIGURA 5 – Menu da direita dos repositórios da UFBA e da UFAL

Busca Facetada	RI-UFAL
<p>Autor</p> <p>ICI-UFBA - Projeto "Elementos... (1009) Barreto, Maurício Lima (155) ICI-UFBA – Projeto "Elementos... (141) Palm, Jairnilson Silva (94) Andrade, Jailson B. de (70) Carvalho, Edgar M. (55) Rangel, Maria do Carmo (48) Melo, Ailton (46) ICI-UFBA - Projeto "Elementos... (45) Barral, Aldina (43) próximo ></p> <p>Assunto</p> <p>Educação (223) Brazil (138) Saude publica (135) Bahia (113) - - - - -</p>	<p>Refinar</p> <p>Autor</p> <p>Acioli, Catarine Gonçalves (1) Albuquerque, Antonia Adriana ... (1) Alvim, Tatiana Araújo (1) Andrade, Victor Hugo Oliveira de (1) Araújo Neto, Renato Américo de (1) Araújo, Ana Luiza Nogueira de (1) Assis, Alexandre Manguéira Li... (1) Barros, Karina Nakai de Carvalho (1) Barros, Marília Gracelidá do... (1) Bezerra, Rodrigo José Rodrigues (1)</p> <p>próximo ></p> <p>Assunto</p> <p>DIREITO (53) Direitos fundamentais (7) - - - - -</p>
Repositório da UFBA	Repositório da UFAL

Fonte: (UFBA, [2010]; UFAL, [2013])

3.3 Avaliação de Acessibilidade

No QUADRO 4 é possível visualizar um resumo da avaliação de acessibilidade realizada nos repositórios usando os avaliadores DaSilva e HERA. Foram utilizados as recomendações WCAG 1.0 e os erros estão classificados em termos de prioridades 1 (P1), 2 (P2) e 3 (P3).

A partir QUADRO 4 é possível constatar que todos os *websites* pesquisados possuem erros que podem dificultar a utilização dos repositórios por usuários com deficiência. A avaliação apresenta resultados similares entre os sites, porém destacam-se os RIs da UFS, UFMS e UFRGS com mais erros, de acordo com a ferramenta validadora DaSilva. Entretanto, esses mesmos *sites* permaneceram na média de erros dos outros *sites*, de acordo com a ferramenta validadora HERA. O repositório da UFBA não pôde ser avaliado pela ferramenta DaSilva, pois o endereço do RI não foi reconhecido pela ferramenta.

QUADRO 4 – Resultados da Avaliação de Acessibilidade

Repositórios	DaSilva			HERA		
	ERROS			ERROS		
	P1	P2	P3	P1	P2	P3
UFAL	11	1	3	0	4	4
UFBA	--	--	--	2	6	4
UFC	11	3	3	2	6	4
UFMA	7	2	3	1	5	3
UFRN	10	4	3	2	6	4
UFPE	7	2	3	1	5	3
UFPB	11	3	3	2	6	4
UFS	18	4	3	0	6	4
UFPA	6	11	2	3	6	4
UFMS	16	3	3	1	4	4
UFMG	11	2	2	2	6	3
UFES	11	3	3	2	6	4
UFSC	4	3	3	1	4	3
UFPR	1	3	3	1	4	3
UFRGS	14	8	2	3	8	4

Fonte: Os autores, 2014

Os itens de prioridade 1 merecem serem vistos com mais cautela e precisam ser atendidos. Nesta prioridade, alguns dos erros encontrados foram: página não possui um equivalente textual para cada imagem apresentada. Isso implica a não tradução e, portanto a não assimilação do teor da imagem por usuários que tenham deficiência visual. Outro problema é que os repositórios não asseguram que todas as informações veiculadas em cor estejam também disponíveis sem cor, para não prejudicar a visualização da informação por pessoas daltônicas. Um outro item revela que há heterogeneidade ao longo das páginas, recomendando usar padrões de desenhos, agrupar itens no menu de forma coerente, como também é recomendada a especificação de cada abreviatura ou sigla.

Outro problema considerado grave é a não identificação do idioma da página para utilização pelos softwares leitores de tela. Isso se torna ainda mais complicado, visto que ao se usar o recurso de visualizar a página em outro idioma, a tradução é feita parcialmente, mesclando idiomas, o que dificultaria a leitura do conteúdo do site por leitores de tela. Constatou-se, também, através de análise manual que a maioria dos sites não apresentavam qualquer ferramenta para facilitar a navegação de usuários com necessidades especiais, tal como a chamada barra de acessibilidade, que ofereceria opções para mudança do contraste, do padrão de cores e possibilidade de aumento ou diminuição do tamanho da fonte. Apenas no site do RI da UFRGS foi encontrada isoladamente a opção de aumento da fonte.

Trabalhar para promover acessibilidade da informação é poder atender um direito de todo e qualquer cidadão. Os repositórios institucionais como meio de divulgação da produção de instituições que utilizam dinheiro público, deveriam primar por atendê-la. Desta forma não estaria restringindo seu número de usuários e corroboraria com a ideia de democracia informacional.

8 CONSIDERAÇÕES FINAIS

Esta pesquisa propôs a avaliação do acesso, busca e visualização de informações nos Repositórios Institucionais com base nos princípios de Arquitetura, Acessibilidade, Usabilidade da Informação. Os resultados obtidos trazem indícios de que é preciso repensar os modelos de interação dos repositórios institucionais das universidades federais brasileiras, devido aos diversos problemas encontrados. Uma das possíveis causas da quantidade e similaridade de problemas é a falta de experiência com o *software* utilizado para a criação dos repositórios e a utilização de esquemas padrões presentes nele. Observou-se, também, problemas de manutenção nesses ambientes e que muitos chegam a ficar semanas indisponíveis (fora do ar). O que se nota é um cuidado apenas para criar esses ambientes, de forma que eles existam, mas não é evidenciada uma preocupação com a interação com a comunidade ou correção de qualquer problema existente, mesmo os mais explícitos.

Como uma continuidade dessa pesquisa, pretende-se realizar testes de usabilidade com usuários reais, de grupos com perfis diversos, nos repositórios. Além de realizar o levantamento das dificuldades, necessidades e anseios desses usuários na utilização dos repositórios. Com isso, espera-se cruzar as informações da avaliação aqui realizada (mais técnica), com os resultados da avaliação feita com os usuários, seguindo o caminho dos chamados estudos híbridos da informação (COSTA; RAMALHO, 2010).

Ainda há muito que se explorar sobre o tema proposto e são muitos os desafios a serem enfrentados. Porém, se tem o cunho social como motivador principal, visto que pesquisas desse tipo podem contribuir com o acesso livre a informação, uma vez que os repositórios institucionais podem ser um vetor para democratização e acesso ao conhecimento científico produzidos neste país.

REFERÊNCIAS

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. **NBR 9241-11:** Requisitos Ergonômicos para Trabalho de Escritórios com Computadores: Parte 11: Orientações sobre Usabilidade. Rio de Janeiro, 2002. 21p.

BASTIEN, C.; SCAPIN, D. **Ergonomic criteria for the evaluation of human-computer interfaces**. Roquencourt: INRIA, 1993. 83p.

BLATTMANN, U. ; BOMFÁ, C. R. Z. Gestão de conteúdos em bibliotecas digitais: acesso aberto de periódicos científicos eletrônicos. **Revista Brasileira de Biblioteconomia e Documentação**, São Paulo, v. 2, n. 1, p. 41-56, 2006. Disponível em: <<http://www.ced.ufsc.br/~ursula/papers/RBBD-2006-16.pdf>> . Acesso em: 05 maio 2014.

CAMARGO, L. S. A.; VIDOTTI, S. A B. G. Análise de elementos de arquitetura da informação em repositórios institucionais digitais: um enfoque ao acesso. In: **CONGRESSO BRASILEIRO DE CIÊNCIAS DA COMUNICAÇÃO**, 26., 2003, Belo Horizonte: INTERCOM – Sociedade Brasileira de Estudos Interdisciplinares da Comunicação, Belo Horizonte: ENDOCOM, p. 1-12, 2003.

CAMARGO, L. S. de A. de. **Metodologia de Desenvolvimento de Ambientes Informacionais Digitais a partir dos Princípios da Arquitetura da Informação**. 2010. Tese (Doutorado em Ciência da Informação) – Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília, 2010, 287p.

COSTA, L. F. da; RAMALHO, F. A. Novas perspectivas dos estudos de satisfação de usuários. **Encontros Bibli: Revista eletrônico de biblioteconomia e ciência da informação**, Florianópolis, v. 15, n. 30, p. 57-73, 2010.

GIL, A. C. **Como elaborar projetos de pesquisa**. 5. ed. São Paulo: Atlas, 2008.

IBICT - Repositórios Digitais. Instituto Brasileiro de Informação em Ciência e Tecnologia. Disponível em: http://www.ibict.br/informacao-para-ciencia-tecnologia-e-inovacao%20/repositorios-digitais_. Acesso em: 23 jul. 2014.

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION. **ISO/IEC 9126-1: Software engineering: product quality: Part 1: Quality Model**. Geneve, 2001.

LAKATOS, E. M.; MARCONI, M. de A.. **Fundamentos de metodologia científica**. 5. ed. São Paulo: Atlas, 2007. 315p.

MEDINA, A. L. **Beneficios de los repositorios institucionales para la comunidad universitaria**. Mi+d – Acceso Abierto a La Infomación Científica. 31 agosto, 2006. Disponível em: <http://www.madrimasd.org/blogs/openaccess/2006/08/31/39195> . Acesso em: 15 jul. 2014.

MORVILLE, P.; ROSENFIELD,L. **Information Architecture for the World Wide Web**. 3. ed. Califórnia: O'reilly media, 2007.

MUELLER, S. P. M. A comunicação científica e o movimento de acesso livre ao conhecimento. **Ciência da Informação**, v. 35, n. 2, p.27-38, maio/ago. 2006.

NIELSEN, J. **Designing Web Usability**. California: New Riders, 1999. 419p.

NIELSEN, J. **Usability Engineering**. San Francisco: Morgan Kaufmann, 1993. 362p.

NIELSEN, J.; LORANGER, H. **Usabilidade na Web: Projetando Websites com Qualidade**. 2. Ed. Campus, 2007.

OPENDOAR. Directory of Open Access Repositories: Brazil. Nottingham: University of Nottingham, @2006-2014. Disponível em: <<http://www.opendoar.org/index.html>>. Acesso em: 14 set. 2014.

PRESSMAN, R. S. **Engenharia de Software.** São Paulo: Editora Makron Books, 2004.

REPOSITÓRIO institucional. In: GLOSSÁRIO. DSpace. **Repositórios digitais.** Brasília: Instituto Brasileiro de Informação em Ciência e Tecnologia, 2007. Disponível em: <http://dspace.ibict.br/index.php?option=com_content&task=view&id=43&Itemid=77>. Acesso em: 20 jun. 2014.

SHNEIDERMAN,B.; PLAISANT, C. Designing the User Interface: Strategies for Effective Human - Computer Interaction. Boston: Addison Wesley, 2004.

TORRES, E. F.; MAZZONI, A. A.; ALVES, J. B. da M. A acessibilidade à informação no espaço digital. **Ciência da Informação**, Brasília, v. 31, n. 3, p. 83-91, set./dez. 2002.

UNIVERSIDADE FEDERAL DA BAHIA (UFBA). **Repositório Institucional.** Salvador: UFBA, [2010]. Disponível em: <<https://repositorio.ufba.br/ri/>>. Acesso em: 15 set. 2014.

UNIVERSIDADE FEDERAL DE ALAGOAS (UFAL). Repositório Institucional. Maceió: UFAL, [2013]. Disponível em: <<http://www.repositorio.ufal.br>>. Acesso em: 15 set. 2014.

UNIVERSIDADE FEDERAL DO CEARÁ (UFC). **Repositório Institucional UFC.** Fortaleza: UFC, [2011]. Disponível em: <<http://www.repositorio.ufc.br>>. Acesso em: 15 set. 2014.

UNIVERSIDADE FEDERAL DO RIO GRANDE DO NORTE (UFRN). **Repositório Institucional.** Natal: UFRN, [2010]. Disponível em: <<http://repositorio.ufrn.br:8080/jspui/>>. Acesso em: 15 set. 2014.

VIANA, C. L. M; MÁRDERO ARELLANO, M. A; SHINTAKU, M. Repositórios institucionais em ciência e tecnologia: uma experiência de customização do DSpace. In: SIMPÓSIO DE BIBLIOTECAS DIGITAIS, 3., 2005, São Paulo. **Anais eletrônicos...** São Paulo: CRUESP, p. 1-27, 2005.

VIDOTTI, S. A. B. G.; CUSIN, C. A. Inclusão digital via acessibilidade web. **Liinc em revista**, Rio de Janeiro, v. 5, n. 1, p. 45-65, mar. 2009.

WINCKLER, M. A.; PIMENTA, M. S. Avaliação de Usabilidade de Sites Web. In: **Escola de Informática da SBC SUL** (ERI 2002) ed. Porto Alegre: Sociedade Brasileira de Computação (SBC), 2002, v. 1, p. 85-137.

Modalidade da apresentação: Pôster

INFLUÊNCIA DA INOVATIVIDADE E FINANCIAMENTO NO RESULTADO EM START-UPS

INFLUENCE OF INNOVATIVENESS AND FINANCING ON OUTCOME AT WEB START-UPS

Anselmo Battisti

Carlos O. Quandt

Resumo: A inovação constante (inovatividade), e os resultados, são fatores crucial para a sobrevivência das organizações em nossa sociedade pós-moderna. Os novos canais mediadores de comunicação são explorados comercialmente por empresas que nascem em um ambiente de extrema incerteza. Estas empresas que utilizam a Internet para oferecer seus serviços são chamadas *web start-ups*. A obtenção de capital por elas para financiar suas atividades não é uma tarefa simples. Os mecanismos tradicionais de financiamento exigem contratos cujos termos e garantias raramente podem ser atendidos por estas organizações, desta forma, novos meios de financiamento foram desenvolvidos para dar suporte as mesmas. É neste contexto que este trabalho está inserido e, o seu objetivo é investigar qual a relação existente entre a inovatividade, o financiamento e o resultado nas *web start-ups*. Este estudo pode ser caracterizado como correlacional-causal com corte transversal. A unidade de análise são as *web start-ups* e seus investidores. Serão utilizados como instrumento de coleta de dados questionários e entrevistas semiestruturadas. A análise dos dados será realizada utilizando métodos mistos de análise onde serão combinadas as técnicas de análise de conteúdo e análise fatorial. Os resultados deste estudo poderão ser utilizados pelas *web start-ups* e financiadores para aumentar o entendimento sobre os fatores que influenciam o resultado nas *web start-ups*.

Palavras-chave: Inovatividade. Financiamento. Resultado. *Web Start-up*.

Abstract: The constant innovation (innovativeness), and the results are crucial factors for the survival of organizations in our postmodern society. The new mediators communication channels are exploited commercially by companies that are born in an environment of extreme uncertainty. These companies use the Internet to offer their services are called *web start-ups*. Traditional funding mechanisms require contracts whose terms and warranties can't be contemplated by these organizations, thus new means of financing have been developed to support the same. Your goal is to investigate the relationship between innovativeness, funding and outcome in *web start-ups*. This study can be characterized as causal-correlational cross-sectional. The unit of analysis is the *web start-ups* and their investors. Will be used as a tool for data collection questionnaires and semi-structured interviews. Data analysis will be performed used mixed methods of analysis which will be combined techniques of content analysis and factor analysis. The results of this study could be used by start-ups and web funders to increase understanding of the factors influencing the outcome in *web start-ups*.

Keywords: Innovativeness; Financing; Incomes; *Web Start-up*.

1 INTRODUÇÃO

A criação de tecnologias como: transistor; microprocessador e as redes de computadores, transformou o modo como a sociedade se organiza (SCHNEIDER;

CASTELLS, 1997). A ordem social que foi apresentada ao mundo após a segunda revolução industrial está sendo modificada. A esta nova ordem social onde a informação desempenha papel central, principalmente com o surgimento da Internet, podemos chamar de sociedade da informação (PINHO, 2011).

Novos empreendimentos nos setores de tecnologia da informação (TI) possuem grande potencial econômico e de transformação social. As *web start-ups* como são comumente chamadas as novas empresas deste setor têm recebido fomento dos governos de diversos países (SCHUBERT; HAUSLER, 2001). Isso ocorre por diversos motivos, alguns deles são: geram empregos com altos salários; trazem receita para o governo; possuem baixo impacto ambiental. Mesmo atuando em um setor onde a informação e o conhecimento são altamente relevantes, o capital financeiro não deixa de ser elemento importante para o surgimento e consolidação destas organizações.

As *web start-ups* criam serviços que são consumidos em um ambiente virtual mediado por um *web site*. Este web site em geral possui público com nível de maturidade tecnológica bastante heterogêneo. Como forma de facilitar o consumo da informação disponibilizada o mesmo deve possuir requisitos básicos de usabilidade e acessibilidade (VIDOTTI; SANCHES, 2004; WURMAN, 2005).

O conhecimento é os processos são elementos fundamentais para a inovação. Dentro das diversas fontes de vantagem competitiva de uma organização o conhecimento tácito é uma das mais duradouras, porém, a formalização do conhecimento tácito em um conhecimento tangível ou também chamado de conhecimento explícito não é uma tarefa simples (ICHIJO; NONAKA, 2007). Entretanto, este processo é importante nas *web start-ups* para que a mesma crie ativos tangíveis, um exemplo desta formalização do conhecimento é a geração de patentes. Um dos fatores que influenciam na obtenção de capital é a tangibilidade dos ativos de uma organização (CASSAR, 2014; CONTI; THURSBY; THURSBY, 2013).

2. FUNDAMENTAÇÃO TEÓRICA EMPÍRICA

2.1 Inovatividade

A inovação é o resultado da mobilização de um conjunto de recursos, comportamentos e atividades que possibilitam o desenvolvimento de novos produtos, processos e sistemas (QUANDT; FERRARESI; BEZERRA, 2013). Portanto, o seu surgimento está associado à convergência de vários fatores complexos e dinâmicos em um contexto organizacional específico. Sem a existência da inovação constante a proposta de valor da organização pode

eventualmente ser copiada por seus concorrentes, deixando assim a competição limitada ao aspecto preço (KAPLAN; NORTON, 2004b).

A inovatividade é a capacidade de inovar de forma contínua e duradoura (QUANDT; FERRARESI; BEZERRA, 2013). (PORTER; KETELS, 2003) apresentam evidências da relação entre a rede de relacionamento de uma organização com seus aspectos inovadores. Por outro lado, (QUANDT; FERRARESI; FREGA, 2012) também encontraram relação positiva entre o desempenho inovador e aspectos como aprendizagem; valorização do comportamento empreendedor; esforços para a interiorização do conhecimento. Desta forma, podemos entender que o fenômeno inovatividade pode ser influenciado tanto por fatores internos como por fatores externos a organização.

2.2 Web Start-ups

Uma *start-up* é uma organização criada para desenvolver novos produtos e serviços sob condições de extrema incerteza (RIES, 2011), cujo modelo de negócio é repetível e escalável (GRANDO, 2012) e evolui por cinco fases independentes, sendo elas: cliente, produto, equipe, modelo de negócio e financiamento, além disso, ela ainda deve estar em um estágio de investimento (MARMER et al., 2012).

As *web start-ups* devem possuir um modelo de negócio escalável e repetível (DRAPER, 2012). Escalável significa que o custo por cliente diminui quanto maior for o número de clientes. O termo repetível significa que o mesmo serviço é consumido por todos sem a necessidade de personalização. A não necessidade de customização pra um cliente específico é o que fortalece a existência das premissas de usabilidade e da acessibilidade nos produtos oferecidos por uma *web start-up*.

2.3 O Financiamento em Start-ups

Uma organização é formata pelos seus recursos tangíveis e intangíveis (KAPLAN; NORTON, 2004a). A construção de uma *web start-up* demanda recursos destas duas naturezas. Neste trabalho serão considerados recursos tangível o capital utilizado advindo de algum tipo de financiamento externo ou próprio dos empreendedores. Os recursos intangíveis representam tudo aquilo que advém de forma implícita ao financiamento ou seja, o conhecimento agregado, influência e *network* gerado em função da entrada de um investidor na *web start-up*.

A fase embrionária de uma *start-up* é denominada *bootstrap*. Assim como nos negócios tradicionais, existem diversas formas de se financiar uma *start-up*. A forma dominante de financiamento externo em *start-ups* são as empresas de capital de risco (*venture*

capital) (BAUM; CALABRESE; SILVERMAN, 2000). Em se tratando especificamente de *web start-ups* outras formas de financiamento também são utilizadas, algumas delas são: capital ou patrimônio próprio; amigos, família; empréstimo em banco; financiamento estatal; incentivos; financiamento por cliente; *crowdfunding*; emprego alternativo (GRANDO, 2012).

O tipo de financiamento demandado varia de acordo com a fase da organização pois, ao longo do tempo as necessidades do montante de capital modificam-se (GREGORY et al., 2005; SPINA, 2012). O uso de recursos internos da organização também deve ser classificado como financiamento, (PEREIRA; GOMES, 2014) apontam que o uso deste tipo de capital torna um ativo antes disponível em um ativo imobilizado ao longo de determinado período de tempo, essa imobilização impede o uso no aproveitamento de possíveis oportunidades que possam surgir.

O custo do financiamento está fortemente relacionado com a assimetria das informações possuídas pelo financiado e financiador. Quanto mais arriscada a operação aparenta ser, maior será a taxa que o financiador irá exigir pelo empréstimo do capital. No outro extremo deste *continuum*, o solicitante por sua vez possui informações que poderiam diminuir esta assimetria, Porém, tal como apresentado por (SHANNON; WEAVER, 1964) existem problemas semânticos e de eficácia que dificultam a comunicação e, apesar da transmissão dos dados ser efetiva estes problemas impedem que o mesmo seja convertido em informação a fim de atingir o seu objetivo primário de comunicar.

2.4 Resultado em Start-ups

No processo de monitoramento da performance de uma organização os indicadores são de fato o elemento mais crítico (CALDEIRA, 2012). A escolha dos mesmos deve englobar tanto os indicadores financeiros, contábeis e os não contábeis. Ainda não existe um conjunto consagrado de indicadores de desempenho para o trabalho intelectual (DRUCKER, 1995), e, por consequência indicadores de resultado consistentes em *web start-ups*.

A identificação do conjunto principal de indicadores de desempenho organizacional é uma tarefa bastante complexa. (NEELY, 1999) apresenta duas possibilidades para esta situação. A primeira delas é que os indicadores que fazem sentido para a organização evoluem ao longo do tempo. A segunda é que nem sempre é óbvio identificar o que necessita ser mensurado. Desta forma a construção destes indicadores deve passar por uma análise individual de cada organização, porém, neste trabalho um conjunto de indicadores padronizados será utilizado como mecanismos para a mensuração do resultado. Estes

indicadores contemplarão as perspectivas cliente; financeira; processo crítico e de aprendizagem organizacional.

3. MÉTODO

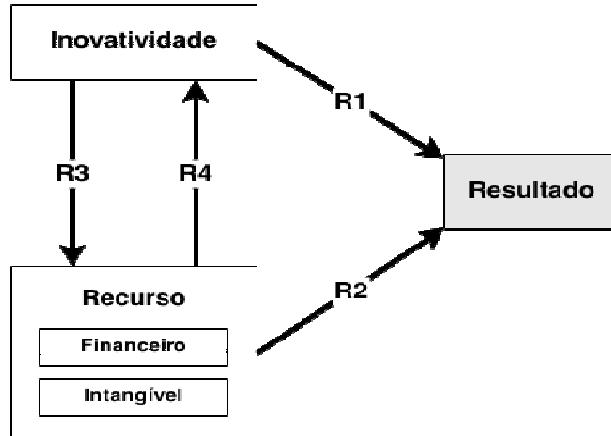
Segundo a taxonomia utilizada para a classificação dos tipos de pesquisa proposta por (SAMPLIERI; COLLADO; LUCIO, 2013) este estudo pode ser caracterizado como sendo, transversal e correlacional-causal / descritivo. Esta classificação se justifica como transversal pois ele busca retratar o fenômeno em um ponto temporal específico. O estudo é correlacional-causal pois busca verificar a existência da relação de causa e efeito entre os construtos, como pode ser visto na Figura 1, e descritivo pois apresentará a descrição de como um fenômeno ocorre.

A unidade de análise deste estudo será composta por duas entidades. A primeira delas são as *web start-ups* e a segunda os investidores que aplicam capital nessas organizações. A população pode ser definida como todas as *web start-ups* brasileiras. Serão admitidas *web start-ups* que estejam em qualquer nível de desenvolvimento.

Esta pesquisa será realizada utilizando métodos mistos de análise. Serão empregados métodos quantitativos e qualitativos, e, sem predominância entre as abordagens. O método misto foi escolhido para aumentar a validade interna e externa da pesquisa, pois sua aplicação permite uma maior triangulação dos dados (SAMPLIERI; COLLADO; LUCIO, 2013).

A relação existente entre os construtos estudados pode ser vista na Figura 1. As relações R1 e R2 serão estudadas individualmente e concomitantemente. Para sua análise serão utilizados métodos quantitativos. A relação R3 e R4 será estudada tanto de forma quantitativa como qualitativa.

FIGURA 1 – Esquema analítico dos construtos analisados



Fonte: Desenvolvido pelo autor

4. CONSIDERAÇÕES PARCIAIS

Este estudo busca analisar a existência da relação entre os construtos inovatividade, financiamento e resultado em *web start-ups*. Ao final da pesquisa espera-se identificar o nível da correlação existente entre estes construtos, bem como, apresentar uma explicação sobre como ocorre o processo de financiamento nestas organizações. Este trabalho justifica-se pois:

- a) É uma associação inédita entre os construtos;
- b) Os resultados obtidos neste estudo poderão auxiliar os demais pesquisadores do campo pois as *web start-ups* como as definidas neste trabalho são um fenômeno relativamente recente e com poucos estudos realizados;
- c) Os empreendedores poderão utilizar os resultados desta pesquisa como elemento norteador sobre suas escolhas de financiamento e, analisar se o esforço necessário para construir *web start-up* inovativa realmente se converte em bons resultados.

REFERÊNCIAS

- BAUM, J. A. C.; CALABRESE, T.; SILVERMAN, B. S. Don't go it alone: alliance network composition and startups' performance in Canadian biotechnology. **Strategic Management Journal**, v. 21, n. 3, p. 267–294, mar. 2000.
- CALDEIRA, J. **100 Indicadores da Gestão Key Performance Indicators**. 1. ed. Lisboa: Actual, 2012.
- CASSAR, G. Industry and startup experience on entrepreneur forecast performance in new firms. **Journal of Business Venturing**, v. 29, n. 1, p. 137–151, jan. 2014.
- CONTI, A.; THURSBY, J.; THURSBY, M. Patents as Signals for Startup Financing. **The Journal of Industrial Economics**, v. LXI, n. 3, p. 592–622, 2013.
- DRUCKER, P. The information executives truly need. **Harvard Business Review**, v. 73, n. 1, p. 54–62, 1995.
- GRANDO, N. **Empreendedorismo Inovador Como criar start-ups de tecnologia no Brasil**. 1. ed. São Paulo / SP: Ed. Evora, 2012. p. 582
- GREGORY, B. T. et al. An Empirical Investigation of the Growth Cycle Theory of Small Firm Financing. **Journal of Small Business Management**, v. 43, n. 4, p. 382–392, out. 2005.
- ICHIJO, K.; NONAKA, I. **Knowledge Creation and Management: New Challenges for Managers**. 1. ed. [s.l.] Oxford University Press, USA, 2007. v. 2011p. 336
- KAPLAN, R.; NORTON, D. Measuring the strategic readiness of intangible assets. **Harvard Business Review**, v. 82, n. 2, p. 52–63., 2004a.
- KAPLAN, R. S.; NORTON, D. P. **Strategy Maps: Converting Intangible Assets Into Tangible Outcomes**. [s.l.] Harvard Business Press, 2004b. p. 454
- MARMER, M. et al. **Startup Genome Report Extra on Premature Scaling**. Disponível em: <<http://bit.ly/1oRf25b>>. Acesso em: 15 set. 2014.

- NEELY, A. The performance measurement revolution: why now and what next? **International Journal of Operations & Production Management**, v. 19, n. 2, p. 205–228, 1999.
- PEREIRA, L.; GOMES, P. **Relação com Investidores Da Pequena Empresa ao Mercado de Capital**. 1. ed. Rio de Janeiro: IBRI – CVM, 2014.
- PINHO, J. A. G. Sociedade da Informação, Capitalismo e Sociedade Civil : Reflexões Sobre Política, Internet e Democracia na Realidade Brasileira. **Revista de Administração de Empresas**, v. 51, n. 1, p. 98–106, 2011.
- PORTER, M.; KETELS, C. **UK Competitiveness: moving to the next stage**. [s.l: s.n.]. Disponível em: <<http://www.bis.gov.uk/files/file14771.pdf>>. Acesso em: 23 abr. 2014.
- QUANDT, C. O.; FERRARESI, A. A.; BEZERRA, C. A. **10 dimensões da inovatividade e seus impactos no desempenho inovador**Encontro da Associação dos Programas de Pós-Graduação em Administração. **Anais...**Rio de Janeiro / RJ: 2013
- QUANDT, C. O.; FERRARESI, A. A.; FREGA, J. R. **Gestão de Ideias e Inovação em Grandes Empresas do Sul do Brasil**Encontro da Associação dos Programas de Pós-Graduação em Administração. **Anais...**Rio de Janeiro / RJ: 2012
- RIES, E. **The Lean Startup**. New York, New York, USA: Crown Business, 2011. p. 336
- SAMPIERI, R. H.; COLLADO, C. F.; LUCIO, P. B. **Metodologia da Pesquisa**. São Paulo / SP: McGraw Hill, 2013.
- SCHNEIDER, P. A.; CASTELLS, M. The Information Age: Economy, Society and Culture Volume I: The Rise of the Network Society. **Journal of Marketing**, v. 61, n. 4, out. 1997.
- SCHUBERT, P.; HAUSLER, U. **E-government meets e-business: a portal site for startup companies in Switzerland** System Sciences, 2001. Proceedings of the 34th Annual Hawaii International Conference on. **Anais...**Hawaii: 2001
- SHANNON, C.; WEAVER, W. **The mathematical theory of communication**. 1. ed. Urbana / Illinois: The University of Illinois Press, 1964.
- SPINA, C. A. **Investidor Anjo Guia Para Empreendedores e Investidores**. São Paulo / SP: nVersos, 2012.
- VIDOTTI, S.; SANCHES, S. **Arquitetura da Informação em web sites** Simpósio Internacional de Bibliotecas Digitais. **Anais...**2004
- WURMAN, R. S. **Ansiedade de Informação 2**. São Paulo / SP: Editora de Cultura, 2005.

TECNOLOGIAS PARA APLICAÇÃO DA WEB SEMÂNTICA NAS BIBLIOTECAS E ARQUIVOS

TECHNOLOGIES FOR APPLICATION OF SEMANTIC WEB IN LIBRARIES AND ARCHIVES

Marcelo Tomita
Maria Elisabete Catarino

Resumo: o projeto *Web Semântica* requer o emprego de várias tecnologias para a organização e estruturação dos conteúdos da *Web*, imprescindíveis para possibilitar a localização e recuperação da informação nos ambientes das unidades de informação. Esta pesquisa pretende relacionar e conceituar as tecnologias utilizadas na *Web Semântica* que podem contribuir para o desenvolvimento da Organização da Informação, dentro da Representação Descritiva e Temática, possibilitando, aos utilizadores das unidades, uma qualidade maior nas buscas de conteúdos e resultados mais significativos. Para a execução da proposta metodológica, será empregada a abordagem qualitativa, que permite uma visão minuciosa e melhor entendimento da *Web Semântica*. Os estudos bibliográficos e documentais englobarão as técnicas adotadas, agregadas à metodologia de estudo de caso, esta última apropriada a uma percepção mais direta das tecnologias que abrangem o projeto, sendo a biblioteca virtual Europeana, o estudo de caso selecionado. Com este trabalho, tem-se a expectativa de preparar um material conciso e bastante para descrever, de maneira sintética, as possibilidades de aplicação da *Web Semântica* nas unidades de informação.

Palavras-chave: Internet. *Web Semântica*. Unidades de Informação. Organização da Informação.

Abstract: the Semantic Web project requires the use of various technologies for the organization and structuring of web content, essential to enable the location and retrieval of information in environments of information units. This research aims to relate and conceptualize the technologies used in the Semantic Web that can contribute to the development of Information Organization, within the Descriptive Representation and Theme, thus allowing unit users a higher quality in searches of content and more meaningful results. For implementation of the methodological procedure, a qualitative approach will be adopted based on the assumption that it enables a thorough and better understanding of the Semantic Web vision. Bibliographic and documentary studies will use techniques, adding to ones already employed in the case study so to allow for a more direct perception of the technologies that comprise the project, with the virtual library Europeana, the selected case study. This work is expected to prepare a concise material able to succinctly describe the application possibilities of the Semantic Web in the units of information.

Keywords: Internet. Semantic Web. Information Units. Organization of Information.

1 INTRODUÇÃO

É nítido o fato de que a *Web* se desenvolve de forma vertical, os conteúdos armazenados e disponíveis são incalculáveis, e isso se dá em virtude de que nunca, na humanidade, se produziu como hoje, e as estatísticas atuais não são capazes de dimensionar, com exatidão, o tamanho real da *Web*. A Internet se tornará imprescindível para armazenar todo o conhecimento exigido nas relações profissionais e pessoais; será como uma extensão

dos nossos pensamentos, pois é praticamente impossível armazenar, organizar e recuperar todas as informações de que se precisa.

Quando um *website* é acessado, os seres humanos são capazes de identificar, por meio de uma rápida visualização, dezenas de elementos nesta página, distinguir seus componentes, compreender as informações, obter várias interpretações sobre esse conteúdo. No entanto, um computador ou *Gadget*¹²⁵, mesmo com seus algoritmos mais recentes e inovadores, ainda é incapaz de realizar, de forma consistente, essa interpretação.

O *World Wide Web Consortium*(W3C) tem sido a principal entidade que fomenta a padronização da *Web* ao desenvolver protocolos e diretrizes para que as recomendações possam estar ao alcance das pessoas que produzem e disponibilizam conteúdos na *Web*. Dentre as recomendações para a organização estrutural da *Web*, destaca-se o projeto da *Web Semântica* que permitirá a organização e padronização da informação na *Web*, que por sua vez, possibilitará sua precisa localização e recuperação por parte do usuário.

Para aperfeiçoar essa *Web*, são diversos os recursos tecnológicos recomendados: a linguagem de marcação para estruturação e descrição de dados, *eXtensible Markup Language* (XML); a linguagem para construção de ontologias, *Web Ontology Language* (OWL); o modelo *Resource Description Framework* (RDF) que permite a estruturação dos dados e a interoperabilidade entre sistemas; a linguagem de buscas *SPARQL Protocol and RDF Query Language* (SPARQL), criada especificamente para a localização de dados na *Web Semântica*, entre outros (WORLD WIDE WEB CONSORTIUM, 2011a).

Por outro lado, na Ciência da Informação (CI) tem-se a Organização da Informação (OI) com suas tradicionais práticas para organização e estruturação das informações, realizadas por intermédio da Representação Descritiva e Temática nas unidades de informação¹²⁶. Existem diversos tipos de estoques informacionais, dentre eles, os Arquivos e Bibliotecas, sendo eles o foco do presente estudo, nos quais também se observa a problemática do volume de informações armazenadas e organizadas para serem recuperadas, principalmente na *Web*.

¹²⁵ “*Gadget*: um dispositivo pequeno, quase sempre mecânico ou eletrônico, com um uso prático, muitas vezes visto como uma novidade.” (DICIONÁRIO MERRIAN-WEBSTER, 2014, tradução nossa).

¹²⁶ Unidade de Informação: entidade encarregada de adquirir, processar, armazenar e disseminar informações com o objetivo de satisfazer as necessidades do usuário. Em muitos casos, é sinônimo de Biblioteca, agência de informação, [...] (CUNHA; CAVALCANTI, 2008, p.370).

A proposta deste trabalho, considerando todo esse contexto, é apresentar um estudo sobre as tecnologias necessárias para a aplicação da *Web Semântica* nas Bibliotecas e Arquivos. Para que o objetivo proposto possa ser atingido, pretende-se realizar uma pesquisa bibliográfica e estudo de caso, que possibilitem contemplar as tecnologias que envolvem a *Web Semântica* e as técnicas de tratamento de informação por parte dos bibliotecários e arquivistas. Esta contextualização permitirá identificar, de forma pontual, quais tecnologias e técnicas serão necessárias para que a *Web Semântica* seja aplicada.

2 INTERNET E A WEB SEMÂNTICA

A internet é uma complexa rede de computadores e dispositivos interligados pelo planeta. Inicialmente, foi concebida sem qualquer organização estrutural, elaborada como uma rede para simples compartilhamento de informação, a qual cresceu de maneira descontrolada, como mencionam Souza e Alvarenga (2004, p.133).

Embora tenha sido projetada para possibilitar o fácil acesso, intercâmbio e a recuperação de informações, a Web foi implementada de forma descentralizada e quase anárquica; cresceu de maneira exponencial e caótica e se apresenta hoje como um imenso repositório de documentos que deixa muito a desejar quando precisamos recuperar aquilo de que temos necessidade.

Robredo (2010, p.14), em um artigo denominado “Ciência da informação e Web semântica: Linhas convergentes ou linhas paralelas?”, refere-se à história da *Web* em “Web 1.0, Web 2.0 e Web 3.0”. A primeira “[...] caracterizava-se por suas páginas estáticas, comportando textos, imagens e links que somente podiam ser alterados pelo Webmaster.” A segunda, *Web 2.0*, caracterizava-se pelo aperfeiçoamento dos serviços de compartilhamento e comunicação, havendo uma “proliferação de redes comunitárias e sociais, hospedagem de serviços e aplicações, compartilhamento de vídeos, wikis e blogs.” A terceira, *Web 3.0*, nomeada como *Web Semântica*, é definida pelos autores Tim Berners Lee, Hendler; Lassila (2001, p.1, tradução nossa) como:

[...] uma extensão da atual, na qual é dada à informação um significado bem definido, o que permite às pessoas e aos computadores trabalharem em cooperação. [...] Num futuro próximo, esses desenvolvimentos deverão conduzir a significativas novas funcionalidades à medida que as máquinas irão se tornando bem mais capazes de processar e —entender os dados que, por enquanto, se limitam somente a apresentar.

Na visão da Inteligência artificial, “A *Web Semântica* é um conjunto de iniciativas para promover uma *Web* futura, cujas páginas serão organizadas, estruturadas e codificadas para que os computadores sejam capazes de fazer inferências e raciocinar seu conteúdo” (CODINA; ROVIRA, 2006, p.2).

O projeto da *Web Semântica* demanda o emprego de muitas tecnologias para a organização e estruturação do conteúdo da *Web*, necessários para facilitar a localização e recuperação da informação. Souza e Alvarenga (2004, p.134) entendem que:

Para atingir tal propósito, é necessária uma padronização de tecnologias, de linguagens e de metadados descritivos, de forma que todos os usuários da *Web* obedecam a determinadas regras comuns e compartilhadas sobre como armazenar dados e descrever a informação armazenada e que esta possa ser “consumida” por outros usuários humanos ou não, de maneira automática e não ambígua. Com a existência da infra-estrutura tecnológica comum da Internet, o primeiro passo para este objetivo está sendo a criação de padrões para descrição de dados e de uma linguagem que permita a construção e codificação de significados compartilhados.

Breitman (2010) realiza uma analogia na maneira como os conteúdos da *Web* estão organizados atualmente com uma biblioteca, no qual os livros estariam todos misturados, sem qualquer organização de assunto. Dessa forma se houvesse a necessidade de localizar um livro com o assunto protocolo TCP/IP, a busca seria relacionada à área de redes. No entanto, se a palavra chave for somente a palavra rede, o resultado da busca poderia trazer livros com assuntos referentes a redes telefônicas, redes de transmissão elétricas e redes de artesanato nordestino.

3 ORGANIZAÇÃO DA INFORMAÇÃO

A CI está focada em pesquisas relacionadas à origem, coleção, organização, armazenamento, recuperação, interpretação, transmissão, transformação e utilização da informação (BORKO, 1968). O foco desta pesquisa na Organização da Informação (OI), subárea da CI está na Representação Descritiva e Temática. A Representação Descritiva realiza a Catalogação, para identificação e descrição dos documentos, utilizando-se de códigos, normas e padrões de metadados. A Representação Temática retrata os conteúdos e auxilia na recuperação da informação, empregando a indexação para extrair os termos mais representativos de documentos e referenciá-los para melhor recuperação. Assim também o faz por meio dos teshauros com seus vocabulários controlados dentro de um domínio específico, as taxonomias que permitem criar estruturas classificatórias para organizar e reunir documentos de forma lógica e as ontologias que representam modelos de dados, abstraindo as especificações dos objetos.

As unidades de informações que adquirem, armazenam e organizam diversos materiais de mídias distintas, sendo virtuais ou físicos, têm se ajustado às mudanças de comportamentos dos usuários frente às novas tecnologias. Essas mudanças não envolvem somente a manipulação dos objetos que deixaram de ser concretos, mas nas técnicas de tratamento das

informações. No entanto, com a ratificação da *Web Semântica*, os processos tradicionais de catalogação, indexação e classificação ainda são efetivos e eficazes para a realidade contemporânea.

4 METODOLOGIA

Para a realização deste estudo, será empregada a abordagem qualitativa, uma vez que não haverá a necessidade de medir ou enumerar eventos na análise dos dados. A abordagem qualitativa possibilita obter os dados descritivos mediante contato direto e interativo do pesquisador com a situação do objeto em estudo (NEVES, 1996), permitindo uma visão minuciosa e melhor compreensão do ambiente da *Web Semântica*, desde a concepção até a sua aplicação.

Com o propósito de compreender os detalhes que possibilitem o funcionamento da *Web Semântica* adotaram-se as pesquisas de análise exploratória, que “[...] tem como objetivo proporcionar maior familiaridade com o problema, com vistas a torná-lo mais explícito” (GIL, 2002, p.41) e a descritiva do tema, “[...] que tem como objetivo primordial a descrição das características de determinada população ou fenômeno” (GIL, 2002, p.42).

As técnicas necessárias para o desenvolvimento do estudo são a pesquisa bibliográfica, documental e estudo de caso. A pesquisa bibliográfica foi realizada em periódicos científicos e bases de dados, resultando em contato com artigos, dissertações, teses e livros. A pesquisa documental buscou informações nas diretrizes e recomendações do W3C. O estudo de caso possibilita a observação direta dos acontecimentos que estão sendo estudados e “[...] conta com o diferenciador de lidar com uma ampla variedade de evidências – documentos, artefatos, entrevistas e observações” (YIN, 2005, p.26).

Na técnica de estudo de caso de acordo com Yin (2005, p.40-41), “[...] os projetos de pesquisa de estudo de caso ainda não foram esquematizados [e] [...] cada tipo de pesquisa empírica possui um projeto de pesquisa implícito”. Entende-se que não há uma sistematização que deve ser seguida no desenvolvimento da pesquisa, as etapas podem ser redefinidas e ajustadas durante o progresso do estudo. No entanto, para que haja uma parametrização na realização do plano de pesquisa serão adotadas as etapas a seguir, conforme Nascimento *et al*(2013) sugerem: Preparação Teórico Metodológica, Seleção do Caso, Coleta de Dados e Análise dos Dados.

O estudo de caso permitirá compreender como se dá na prática a adoção das tecnologias necessárias para a aplicação da *Web Semântica*. O caso selecionado foi a

Europeana, um projeto da *Europeana Foundation*, que com apoio dos países da comunidade europeia organiza conteúdos de domínio público.

A Coleta de Dados consistirá em levantar e documentar as aplicações da *Web Semântica* no *site* da *Europeana*, empregando fontes de evidências, conforme sugere Yin (2010). Ainda segundo o autor supracitado, é relevante utilizar “[...] duas ou mais fontes, convergindo sobre os mesmos fatos ou descobertas [...]” (2010, p. 124), visto que tal ação possibilita contemplar os tópicos exigidos. Para este estudo de caso, alinhado com as ideias de Yin (2010, p.129) serão utilizadas cinco fontes de evidências: Documentação, Registro em arquivo, Entrevistas, Observação Direta, Artefatos Digitais. Por meio desta coleta de dados, pretende-se recolher dados que possibilitem pormenorizar os detalhes das tecnologias envolvidas na produção do *website* biblioteca virtual *Europeana*.

Na fase de Análise de dados, será elaborada uma avaliação das tecnologias-empregadas para detectar quais delas serão necessárias à aplicação nas atividades das Bibliotecas e Arquivos. As prováveis Categorias de Análise ou Análise Categorial: Metadados, Linguagem de marcação, Linguagem de busca, Ontologias e Vocabulários e Método de varredura de dados.

5 CONSIDERAÇÕES FINAIS

Com este estudo, pretende-se apresentar, de forma sintética, as possibilidades de aplicação da *Web Semântica* nas unidades de informação, produzindo um material que possibilite, aos envolvidos na preparação e distribuição de conteúdos para a *Web*, uma visão clara das tecnologias envolvidas para a padronização da *Web*.

O presente trabalho, em andamento, aspira por contribuir para novas pesquisas, tais como prova de conceito para resultados das recomendações da *Web Semântica* e fortalecimento da formação do profissional da informação.

Para possibilitar uma percepção prática de todo o ambiente necessário para aplicação das tecnologias da *Web Semântica* em Arquivos e Bibliotecas, será realizado um estudo de caso com o *site* da biblioteca virtual *Europeana*, o qual atende as recomendações dos protocolos e diretrizes da tecnologia estudada.

REFERÊNCIAS

BERNERS-LEE, T; LASSILA, O; HENDLER, J. The Semantic Web. *Scientific American*, 2001.p.34-43.

BORKO, H. **Information science: what is it?** American Documentation, Jan. 1968. Disponível em: <<http://pt.scribd.com/doc/533107/Borko-H-v-19-n-1-p-35-1968>> . Acesso em: 8 abr. 2014.

BREITMAN, Karin Koogan. **Web Semântica:** A Internet do Futuro. Rio de Janeiro: LTC, 2010. 190p.

CODINA, L.; ROVINA, C. La Web Semántica. In: TRAMULLAS, J.(Org.).**Tendencias em Documentacion Digital.** Guijón: Trea, 2006. Cap.1, p.9-54.

DICIONÁRIO MERRIAN-WEBSTER. **Gadget.** Disponível em: <<http://www.merriam-webster.com/dictionary/gadget>>. Acesso em: 17 jun. 2014

GIL, A. C. **Como elaborar projetos de pesquisa.** 4.ed. São Paulo: atlas, 2002. 176p.

NASCIMENTO, A. V. et al. O uso do método Estudo de Caso na Ciência da Informação no Brasil.InCID:. **Revista de Ciência da Informação e Documentação**, v. 4, p.49-69, 2013.

NEVES, J.L. Pesquisa Qualitativa – Características, Usos e Possibilidades. **Caderno de pesquisas em Administração**, São Paulo, V1, N 3, 2 Sem. 1996. Disponível em: <<http://www.ead.fea.usp.br/cad-pesq/arquivos/c03-art06.pdf>>. Acesso em: 08 fev. 2014.

ROBREDO, J. Ciência da informação e Web semântica: Linhas convergentes ou linhas paralelas? In: ROBREDO, J.; BRÄSCHER, M. (Org.). **Passeios pelo Bosque da Informação:**. Brasília: IBICT, 2010. p.2-40. Disponível em: <http://repositorio.unb.br/bitstream/10482/7952/6/CAPITULO_CienciaInformacaoWebSemantica.pdf>. Acesso em: 20 abr. 2014.

SOUZA, R. R.; ALVARENGA, L. A web semântica e suas contribuições para a ciência da informação. **Ciência da Informação**, Brasília, v. 33, n. 1, p.132-141, jan./abr. 2004.

UNIDADE DE INFORMAÇÃO. In: CUNHA, M. B. da; CAVALCANTI, C R. O. **Dicionário de Biblioteconomia e Arquivologia**. Briquete de Lemos. Brasília. 2008. 451p.

WORLD WIDE WEB CONSORTIUM. **Web Semântica.** 2011a. Disponível em:<<http://www.w3c.br/Padroes/WebSemantica>>. Acesso em: 31 jan. 2014.

YIN, R. K. **Estudo de Caso:** Planejamento e Métodos. 3. ed. Porto Alegre. Bookman. 2005. 212p.

YIN, R. K. **Estudo de Caso:** Planejamento e Métodos. 4. ed. Porto Alegre. Bookman. 2010. 248p.

FERRAMENTA DE SOFTWARE LIVRE PARA CONSTRUÇÃO DE BASES DE DADOS REFERENCIAIS DE ARTIGOS CIENTÍFICOS DIGITAIS

FREE SOFTWARE TOOL FOR THE CONSTRUCTION OF REFERENTIAL DATABASE FOR DIGITAL SCIENTIFIC ARTICLES

Flávio Ribeiro Córdula
Wagner Junqueira de Araújo

Resumo: Com a especialização de conteúdos nas bases de dados, a Internet se mostra fonte de pesquisa para estudantes, professores e pesquisadores. Existem várias bases de dados que referenciam, de forma exclusiva, itens bibliográficos, como livros, artigos científicos, dissertações, teses etc. O portal de Periódicos Capes, o SciELO e a Brapci são exemplos de bases de dados referenciais. O objetivo geral desta pesquisa foi estudar requisitos para o desenvolvimento de um *software livre* para criação de bases referenciais que auxilie a disseminação de informação científica para a comunidade científica do Brasil. Esta pesquisa conta com o apoio do CNPQ e foi desenvolvida em nível de mestrado. Trabalha com dois tópicos estudados pelos pesquisadores da Ciência da Informação, as bases de dados e os mecanismos de recuperação da informação em suporte digital. Como metodologia utilizou-se de revisão de literatura e linguagens de programação. Como resultado, foi desenvolvido um *software livre* que foca na construção de bases referenciais para recuperação de artigos científicos digitais por meio da disponibilização de *links*. O *software* foi denominado "Base de Dados Referencial de Artigos Científicos" (BADRAC), está hospedado no endereço eletrônico <http://badrac.herokuapp.com> e permite que qualquer interessado construa uma base de dados referencial. Por fim, deve-se considerar que o *software* disponibilizado está na versão 1.0.0 beta, e necessita de mais testes e da contribuição dos usuários e dos pesquisadores da área da Ciência da Informação.

Palavras-chave: Base de Dados Referencial. Recuperação da Informação Digital. Periódico Científico Digital.

Abstract: With the specialization of content in databases the Internet proves resource for students, teachers and researchers. There are several databases that reference, exclusively, bibliographic items, such as books, journal articles, dissertations, theses etc. The portal de Periódicos Capes, SciELO and Brapci are examples of referential databases. The main objective of this research was to study requirements for the development of a free software that helps to building referential databases and for dissemination of scientific information to the scientific community in Brazil. This research has been supported by CNPq and was developed at master's degree level. It works with two topics studied by information science researchers, databases and mechanisms for information retrieval in digital format. The methodology used was literature review and programming languages. As a result, we developed a free software that focuses on building referential databases for the retrieval of digital scientific articles by providing links. The software was called "Base de Dados Referencial de Artigos Científicos" (BADRAC), is hosted at the website <http://badrac.herokuapp.com> and allows any interested party to build referential databases. Finally, one must consider that the software is available in its 1.0.0 beta version and needs more testing and input from users and researchers in the field of Information Science.

Keywords: Referential Database. Digital Information Retrieval. Digital Scientific Journal.

1 INTRODUÇÃO

Considerando que a evolução e o desenvolvimento da sociedade fundamentam-se na disponibilização, no acesso, no uso e, eventualmente, no descarte da informação, é necessário refletir sobre o papel e as aplicações da tecnologia para a comunicação e a sua disseminação na sociedade. Essa comunicação pode ser realizada por meio de diferentes canais – dentre os quais se destaca a comunicação científica – e alcançar públicos distintos, como a comunidade científica, alvo ao qual se destina esse tipo de comunicação.

Assim, a comunicação científica é parte inerente do desenvolvimento da ciência e está fundamentada na informação científica, gerando conhecimento, que por sua vez é divulgado, sobretudo, por periódicos científicos (GUEDES, 1998).

O periódico científico, por sua vez, desenvolve a disseminação da produção científica e tem um papel fundamental no meio acadêmico, promovendo avanços e destacando autores e editores. Na década de 1990, com a chegada da informação eletrônica e o uso da Internet, ocorreu a grande ruptura no modo de editoração e disseminação de informações, principalmente da produção científica (FACHIN, 2002). Com a Internet, o formato digital passou a promover a disseminação de conhecimento científico em larga escala e permitir uma divulgação mais eficiente de suas publicações.

Porém, devido a grande quantidade de informação armazenada na Internet, o que se observa é que a heterogeneidade das informações é um problema para que uma recuperação de informação relevante na *Web* seja alcançada. Essa questão, entretanto, pode ser suavizada com a organização especializada das bases de dados.

Com a especialização de conteúdos nas bases de dados, a Internet se mostra fonte de pesquisa para estudantes, professores e pesquisadores. Existem várias bases de dados que referenciam, de forma exclusiva, itens bibliográficos, como livros, artigos científicos, dissertações, teses etc. O portal de Periódicos Capes, o SciELO e a Brapci são exemplos de bases de dados referenciais.

É fato que a organização especializada das bases de dados reduz o escopo das informações passíveis de busca. Diferentemente de uma pesquisa geral na *Web*, onde vários tipos de bases de dados, que referenciam os mais diversos tipos de informação, são acionadas, as bases de dados que referenciam apenas itens bibliográficos aumentam a relevância dos resultados das buscas.

Em acréscimo a funcionalidade proporcionada pelas principais bases de dados referenciais em que os usuários usufruem dos serviços de busca de artigos previamente cadastrados, um *software* livre para a construção de base de dados referencial de artigos

científicos proporciona a estudantes, professores e pesquisadores a possibilidade de criação de suas próprias bases de dados, auxiliando na disseminação da informação científica.

O objetivo geral desta pesquisa foi estudar requisitos para o desenvolvimento de um *software livre* para construção de bases de dados referencias que auxilie a disseminação de informação científica para a comunidade científica do Brasil. Esta pesquisa conta com o apoio do CNPQ e trabalha com dois tópicos estudados pelos pesquisadores da Ciência da Informação, que são as bases de dados e o mecanismos de recuperação da informação em suporte digital.

2 BASE DE DADOS

As bases de dados provocaram um grande impacto no uso de computadores. Elas representam um papel crítico em todas as áreas de conhecimento em que o computador e suas tecnologias específicas são utilizados: nas áreas afins de Negócio e *Business*, nas Engenharias, na Matemática e na Física, na Medicina, na Educação, na Ciência da Informação, entre outras áreas.

A definição de base de dados é vista por Heemann (1997, p. 2) como “um arquivo ou um conjunto de arquivos computacionais no qual são armazenados dados, permitindo a recuperação e atualização de informações”. Já para Rowley (2002, p. 125), o termo base de dados é conceituado como “uma coleção geral e integrada de dados junto com a descrição deles, gerenciada de forma a atender a diferentes necessidades de seus usuários”.

Zaninelli e Catarino (2004) afirmam que o meio eletrônico é uma forma rápida de armazenar, transmitir, compartilhar e recuperar documentos e informações. As autoras ainda expõem que:

Um recurso muito utilizado e eficaz para a disponibilização e troca de informações entre organizações públicas e privadas, é o banco de dados e, consequentemente, as suas bases de dados que ampliam significativamente a qualidade das buscas informacionais, visto que essas bases apresentam diversificados pontos de acesso à informação. (ZANINELLI; CATARINO, 2004, p. 2).

Existem vários tipos e nomenclaturas que diferenciam as bases de dados, como por exemplo: referencial, de fonte, de texto completo, numérica, multimídia, de diretórios, transacional (ROWLEY, 1994; FEATHER, STURGES, 2003; REITZ, 2014).

Classificam-se como referencias e de fontes aquelas disponíveis por meio de um serviço de busca (ROWLEY, 1994; CUNHA, 2001). As bases referenciais remetem o usuário à outra fonte – base de dados de fonte – que pode conter dados digitais originais e, em muitos casos, de texto completo (ROWLEY, 1994).

Feather e Sturges (2003) asseveram que as bases de dados bibliográficas contêm referências a publicações, incluindo periódicos, artigos, anais de congressos, relatórios governamentais, publicações legais, patentes, livros, etc. Diante desta afirmação, percebe-se que uma base de dados bibliográfica é uma base de dados referencial de escopo reduzido, especializado em itens bibliográficos.

3 RECUPERAÇÃO DA INFORMAÇÃO DIGITAL

Os Sistemas de Recuperação de Informação (SRIs) das bibliotecas de terracota na Babilônia, de pergaminho em Pérgamo e de papiro em Alexandria passaram por transformações significativas até chegar aos modernos sistemas com bases de dados (ARAÚJO, 1995). Apesar desse longo caminho percorrido, foi apenas em 1951 que Calvin Mooers, baseado em estudos sobre informação, criou o termo *Information Retrieval* (Recuperação da Informação). Segundo o autor “A Recuperação de Informação trata dos aspectos intelectuais da descrição da informação e sua especificação para busca, e também de qualquer sistema, técnicas ou máquinas que são empregadas para realizar esta operação.” (MOOERS, 1951, p. 21, tradução nossa).

Os sistemas de recuperação de informação viabilizam o acesso aos documentos, a partir da execução de tarefas de representação da informação. Tal processo se materializa pela representação e, também, pela recuperação das informações representadas e dos documentos armazenados com a finalidade de atender às necessidades dos usuários (SOUZA, 2006).

A recuperação de documentos e informações em bases de dados é um assunto amplamente discutido pelos profissionais da Ciência da Informação (FERNEDA, 2003; BRANSKI, 2004; CHU, 2010). Na conjuntura da Ciência da Informação, a recuperação de informação (RI) é uma operação por meio da qual são selecionados documentos a partir de um acervo, um repositório ou uma base de dados. Segundo Ferneda (2003, p. 14), “há ainda autores que conceituam a recuperação de informação de forma muito mais ampla, ao subordinar à mesma o tratamento da informação (catalogação, indexação, classificação)”. Pode-se dizer que a RI tem sido utilizada para designar a busca de literatura (LANCASTER, WARNER, 1993; FERNEDA, 2003) em bases de dados bibliográficas.

Atualmente, a recuperação de informação segue o conceito do *two steps information retrieval* ou recuperação de informação em duas etapas. Gonzalez de Gomez (2004, p. 56) explica esses dois *steps*, ao afirmar que “o primeiro *step* consistiria em um dispositivo de representação que descreve, sumariza e codifica as fontes primárias de informação para facilitar sua busca seletiva e sua localização efetiva” e que o segundo *step*, seria a localização

e disponibilização das fontes primárias escolhidas, nelas mesmas, dando acesso a seu texto completo.

4 DESENVOLVIMENTO, RESULTADO E CONSIDERAÇÕES

Como resultado desta pesquisa em nível de mestrado, foi desenvolvido um *software* livre que foca na recuperação de artigos científicos digitais por meio da disponibilização de *links*, dependendo do termo de busca pesquisado, e remete o usuário à fonte externa, que contém os arquivos digitais disponíveis para *download*. O *software* foi denominado “Base de Dados Referencial de Artigos Científicos” (BADRAC) e está hospedado no endereço eletrônico <http://badrac.herokuapp.com>.

Este está dividido em duas interfaces: a primeira, disponível apenas para o gestor da base, é formada pelo *menu* administrativo, que tem como função principal o cadastro das referências dos artigos. A segunda, disponível para qualquer usuário, é formada pelo mecanismo de recuperação de informação ou de busca.

Para o desenvolvimento da ferramenta, foi feito um estudo prévio no código fonte e base de dados dos *software* Open Journal System (OJS) e DSpace. Esses dois sistemas são exemplos de *software* livre que auxiliam na produção e disseminação da informação científica. Embora ambos possuam características de bases de dados de fonte, e não de referência, a lógica usada no desenvolvimento do mecanismo de busca e no processo de representação da informação é similar.

Outra fonte de estudo usada para o desenvolvimento da pesquisa foi o padrão de metadados Dublin Core, que auxiliou a criação do Modelo de Entidade-Relacionamento (MER) para a descrição e representação lógica das informações na base de dados. Os metadados auxiliam a construção de uma base de dados por meio da descrição da realidade, representando logicamente as informações armazenadas nas bases de dados.

O MER tem por base a descrição de algo do mundo real realizada por intermédio de um conjunto de objetos chamados “entidades” e pelo conjunto de “relacionamentos” entre esses objetos. Entidade é uma coisa (física ou conceitual) do mundo real. Além disso, as entidades possuem “atributos”, que são propriedades particulares que as descrevem. Os valores assumidos por cada atributo descrevem as entidades e formam o conjunto de dados armazenados na base de dados (ELMASRI; NAVATHE, 2005). Esses valores são as representações da informação, que podem ser a informação de fato, ou o fornecimento de pontos de acesso ao texto, ao documento, a informação completa.

Este *software* foi desenvolvido sob a plataforma Linux Ubuntu 12.4 LTS por meio da linguagem de programação Ruby on Rails versão 4.0.2 e sistema gerenciador de banco de dados PostgreSQL versão 9.1.13. Ele pode ser acessado em qualquer computador ou dispositivo móvel, independente do sistema operacional, desde que o faça por meio dos navegadores Google Chrome (v. 34.0 ou superior) ou Mozilla Firefox (v. 31.0 ou superior).

O resultado deste trabalho se caracteriza como *software* livre, pois o código fonte está hospedado no GitHub – um repositório que tem como função essencial o armazenamento e o compartilhamento de códigos abertos.

A partir da utilização da Ferramenta pesquisadores, grupos de pesquisa ou interessados podem criar bases de dados que referenciem apenas artigos de uma linha ou assunto específico ou, analogamente, um programa de pós-graduação, independentemente da sua área de atuação, pode subsidiar estudos e propostas ao referenciar a produção de artigos daquela área em particular. Ou seja, sendo um *software* livre, pode ser manipulado tanto por uma universidade/faculdade quanto por um docente ou discente; é livre para que qualquer um que tenha interesse em criar uma base de dados referencial personalizada às suas necessidades usufrua de seus benefícios.

O cadastro de referências de artigos é feito no *menu* administrativo e necessita seguir uma ordem lógica e comprehensível daquilo que está sendo armazenado. Primeiro, pelo menos um periódico necessita ser cadastrado e, a partir de então, edições podem ser a ele vinculadas. Um periódico pode ter várias edições, mas uma edição pode pertencer a apenas um único periódico. Depois de realizado os cadastros dos periódicos e suas respectivas edições, as referências dos artigos podem ser armazenadas. Um artigo pertence a uma edição que está, por sua vez, vinculada a apenas um periódico.

Com o artigo cadastrado, autores e palavras-chave podem ser vinculados. Um artigo pode ter sido escrito por diversos autores e um autor por ter escrito mais de um artigo. Essa mesma lógica é usada para o atributo palavra-chave. Assim, a descrição de algo do mundo real, nesse caso, artigos científicos, está de acordo com a percepção do usuário, evitando possíveis erros de busca provenientes de uma tarefa cadastral mal executada.

A inclusão, atualização e remoção de referências a artigos científicos devem ser realizadas regularmente. Esse processo cadastral pode ser realizado a qualquer momento. Tão logo uma nova edição de um periódico de interesse for lançada, a inclusão dos artigos deve ser feita.

Por fim, deve-se considerar que o *software* está na versão 1.0.0 beta, e necessita de mais testes e da contribuição dos usuários e dos pesquisadores da área da Ciência da Informação.

REFERÊNCIAS

- ARAÚJO, V. M. R. H. de. Sistemas de Informação: nova abordagem teórico-conceitual. **Ciência da Informação**, Brasília: v. 24, n. 1, 1995. Disponível em: <<http://dici.ibict.br/archive/00000141/01/Ci%5B1%5D.Inf-2004-577.pdf>>. Acesso em: 05 jul. 2013.
- BRANSKI, Regina Meyer. Recuperação de informações na Web. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 9, n. 1, p. 70-87, jan./jun. 2004. Disponível em: <<http://www.brapci.ufpr.br/documento.php?dd0=0000003315&dd1=e4719>>. Acesso em: 09 nov. 2013.
- CHU, Heting. Information Representation and Retrieval in the Digital Age. New Jersey: ASIS& T, 2010.
- CUNHA, Murilo Bastos da. **Para saber mais**: fontes de informação em ciência e tecnologia. Brasília, DF: Briquet de Lemos/Livros, 2001. 168 p.
- ELMASRI Ramez; NAVATHE Shamkant B. **Sistemas de Banco de Dados**. 4. ed. São Paulo: Pearson, 2005.
- FACHIN, G. R. B. **Modelo de avaliação para periódicos científicos on-line**: proposta de indicadores bibliográficos e telemáticos. 2002. 210 f. Dissertação (Mestrado em Engenharia de Produção) – Universidade Federal de Santa Catarina, Santa Catarina, 2002. Disponível em: <<https://repositorio.ufsc.br/bitstream/handle/123456789/83088/185438.pdf?sequence=1>>. Acesso em: 11 fev. 2014.
- FEATHER, J. STURGES, P. International Encyclopedia of Information and Library Science. Londres. ed. 2, 2003.
- FERNEDA, E. **Recuperação da informação**: análise sobre a contribuição da Ciência da Computação para a Ciência da Informação. 2003. 137 f. Tese (Doutorado em Ciências da Comunicação) – Escola de Comunicação e Artes da Universidade de São Paulo, São Paulo, 2003. Disponível em: <<http://www.teses.usp.br/teses/disponiveis/27/27143/tde-15032004-130230/pt-br.php>>. Acesso em: 19 fev. 2013.
- GONZALEZ DE GOMEZ, M.N. **Novas fronteiras tecnológicas das ações de informação**: questões e abordagens. Ciência da Informação, v. 33, n. 1, p.55-67, jan./abr.2004.
- GUEDES, Maria das Graças Targino Moreira. **Comunicação científica**: o artigo de periódico nas atividades de ensino e pesquisa do docente universitário brasileiro na pós-graduação. 1998. 387 f. Tese (Doutorado em Ciência da Informação) – Departamento de Ciência da Informação e Documentação, Universidade de Brasília, Brasília, 1998.
- HEEMANN, V. M. **Avaliação ergonômica de interfaces de bases de dados por meio de Checklist especializado**. 1997. Dissertação (Mestrado em Engenharia da Produção) – Universidade Federal de Santa Catarina, Florianópolis, 1997.

LANCASTER, F. W. e WARNER, A. J. **Information Retrieval Today**. Information Resources Press, 1993.

MOOERS, C. (1951). **Zatocoding applied to mechanical organization of knowledge**. American Documentation, v. 2, n. 1, p.20-32.

REITZ, Joan M. **ODLIS** – Online Dictionary for Library and Information Science: Libraries Unlimited, 2014. Disponível em: <<http://www.abc-clio.com/ODLIS/searchODLIS.aspx>>. Acesso em: 13 mar. 2014.

ROWLEY, Jennifer. **Informática para bibliotecas**. Tradução Antonio Agenor. Brasília: Brinquet de Lemos/Livros, 1994, 307 p.

ROWLEY, Jennifer. **A biblioteca eletrônica**. Tradução Antonio Agenor. Brasília: Brinquet de Lemos/Livros, 2002, 399 p.

SOUZA, R. R. **Sistemas de Recuperação de Informações e Mecanismos de Busca na Web**: panorama atual e tendências. Perspectivas da Ciência da Informação, Belo Horizonte, v. 11, n. 2, p. 161-173, maio./ago. 2006. Disponível em: <<http://www.scielo.br/pdf/pci/v11n2/v11n2a02.pdf>>. Acesso em: 05 jul. 2013.

ZANINELLI, T. B.; CATARINO, M. E. Recuperação de informação no banco de dados da Embrapa soja: estudo do ainfo. **Inf. Inf.**, Londrina, v. 9, n. 1/2, jan. / dez. 2004.

MODELO PARA O DESCARTE SEGURO DA INFORMAÇÃO EM SUPORTE DIGITAL

Silvio Lucas da Silva
Wagner Junqueira Araújo

Resumo: Ao se observar uma das etapas do ciclo de vida da informação como o descarte, percebe-se que este procedimento em suporte digital difere do descarte de informações em suporte físico. Verificou-se na literatura uma lacuna sobre este tema na área da Ciência da Informação, pois ao se descartar uma informação em suporte digital, é preciso considerar o tipo de mídia na qual a informação está registrada e sua possível reutilização. O descarte deve considerar os requisitos de gestão de segurança da informação, de tal forma que impossibilite a recuperação posterior dessa informação. Esta pesquisa teve como objetivo elaborar uma proposta de modelo para o descarte seguro da informação em suporte digital, considerando que alguns procedimentos de descarte inviabilizam o uso posterior da mídia informática. No caso dos sistemas gerenciadores de banco de dados (SGBDs) que são utilizados em sistemas de informação de muitas organizações, por exemplo, os equipamentos não podem ser destruídos nem os sistemas podem ser suspensos por longos períodos de tempo para aplicação de uma técnica de sobrescrita. Esta pesquisa foi desenvolvida como um estudo de caso, descritiva, com coleta de dados implementada de forma empírica realizada por anotações de testes de laboratório. Como resultado é apresentado uma proposta de modelo que leva em conta as necessidades das organizações fundamentada na literatura e nos resultados dos testes de laboratório.

Palavras-chave: Gestão da informação e do conhecimento. Tecnologia da informação e comunicação. Gestão da segurança da informação. Documento digital.

Abstract: By observing one of the stages of the information life cycle such as the disposal, we found that this procedure in digital support differs from the disposal of information on physical support. A gap was noticed in the literature on this topic in the field of information science, because when it is necessary to discard some information in digital support, we need to consider the type of media on which the information is recorded and its possible reuse. Disposal should consider the requirements of information security management in such a way that precludes the subsequent recovery of this piece of information. This research aimed to develop a model proposal for the secure discard of information in digital support taking into account that some disposal procedures make unviable the subsequent use of the media. In the case of databases management system (DBMSs) that are used in information systems of many organizations, for example, the pieces of equipment cannot be destroyed systems and the systems cannot be suspended for long periods of time for applying an overwriting technique. This research was conducted as a case study, descriptive, whose data collection was implemented empirically performed by notes of laboratory tests. As a result, it is presented a model proposal that considers the organizations needs based on the literature and on the results of laboratory testing.

Keywords: Information and knowledge management. Information and communication technology. Information security management. Digital Document.

1 INTRODUÇÃO

A sociedade sempre passa por diferentes transformações em todas as áreas do conhecimento. Nestas transformações, a informação assume um papel de destaque não só para as pessoas, mas também para as organizações.

Dentro deste contexto, as tecnologias da informação e comunicação (TICs) contribuíram para o desenvolvimento da chamada “sociedade da informação”, através das mais diversas formas como, por exemplo, dispositivos de hardware e sistemas de software para disseminação, armazenamento, e recuperação da informação. Conforme ressaltam Capurro e Hjørland (2007, p. 149), “é o surgimento da tecnologia da informação e seus impactos globais que caracterizam a nossa sociedade como uma sociedade da informação”.

Sendo a informação definida como um conhecimento inscrito de forma impressa ou digital em um suporte (LE COADIC, 2004), ela necessita seguir os preceitos do ciclo de vida da informação, como caracteriza Beal (2008), através de sete etapas distintas: identificação das necessidades e dos requisitos, obtenção, tratamento, distribuição, uso, armazenamento e descarte.

Ao se observar uma das etapas do ciclo de vida da informação, “o descarte”, percebe-se que este (em suporte digital) difere do descarte de informações em suporte físico. Descartar a informação em suporte físico (papel) pode ser realizado através de processos mecânicos/manuais como fragmentação e/ou incineração, enquanto a informação em suporte digital carece de metodologias próprias. Deve-se ter o cuidado para evitar que a informação gravada em mídia digital seja recuperada, posteriormente, por terceiros não comprometendo assim sua confidencialidade.

Foi verificado que existe uma lacuna sobre esse tema no campo da Ciência da Informação (CI), tendo esta pesquisa o objetivo de elaborar uma proposta de modelo para o descarte seguro da informação em suporte digital, sendo desenvolvida em nível de mestrado.

2 O DESCARTE DE INFORMAÇÕES EM DISPOSITIVOS DE ARMAZENAMENTO DIGITAL

A informação em suporte digital encontra-se registrada em diversos dispositivos informáticos, por exemplo: discos rígidos (*Hard Disk Drives – HDDs*), *pen drives*, cartões de memória, CDs, DVDs, etc., denominados de dispositivos de armazenamento. Estes dispositivos são constituídos por um material específico no qual a informação fica efetivamente registrada: mídias magnéticas ou mídias em estado sólido. O Quadro 1 exemplifica o tipo de mídia utilizada nos dispositivos de armazenamento.

QUADRO 11 – Dispositivo informático X Tipo de mídia utilizada

Dispositivo informático	Mídia utilizada para armazenamento de informações	Exemplo de utilização da tecnologia
Disco Rígido (<i>Hard Disk – HDD</i>)	Magnética	Computadores, <i>notebooks</i> , <i>storages</i>
<i>Pen Drive</i>	Memórias FLASH	Armazenamento e transporte de

		documentos
Cartões de Memória (internos ou externos)	Memórias FLASH	Câmeras fotográficas digitais, <i>tablets, smartphones</i>
<i>Solid State Drive (SSD)</i>	Memórias FLASH	<i>Notebooks compactos</i>
Fitas Magnéticas (LTO, DAT)	Magnética	Cópias de segurança (<i>backup</i>)
<i>Compact Disks (CDs), Digital Versatile Disks (DVDs), BDs (Blu-Ray Disks) - regraváveis ou não</i>	Óptica	Cópias de segurança (<i>backup</i>), CDs de áudio / filmes / jogos para computador / vídeo games, dentre outros.

Fonte: Elaborado pelos autores.

Com o objetivo de organizar as informações registradas, os dispositivos de armazenamento utilizam-se de tecnologias para gerenciar a gravação das informações, denominadas sistemas de arquivos como, por exemplo, FAT, NTFS (criados pela Microsoft), EXT2, EXT3, EXT4 e ReiserFS (pertencentes aos sistemas operacionais UNIX e LINUX), além das tecnologias de armazenamento utilizadas em *storages*¹²⁷, como *Storage Area Network (SAN)*, entre outras.

Estas se utilizam de uma tabela (um índice), gravada em uma área determinada da mídia, que permite ao sistema operacional recuperar as informações disponíveis, seguindo um mecanismo de endereçamento específico de cada tecnologia. Ao apagar informações, os sistemas de arquivos removem a informação presente neste índice, ou seja, apagam somente a informação de onde o arquivo está localizado fisicamente na mídia e não a informação em si. Consequentemente, a informação gravada continua intacta, sendo passível de recuperação.

Farmer e Venema (2007, p. 131) afirmam que “uma grande quantidade de informações excluídas podem ser recuperadas [...], mesmo quando essas informações foram excluídas há muito tempo”. Quando se trata da gestão documentos digitais, tal característica não está de acordo com as recomendações feitas pelo Conselho Nacional de Arquivos (CONARQ) para a eliminação da informação:

[...] a eliminação deve ser realizada de forma a impossibilitar a recuperação posterior de qualquer informação confidencial contida nos documentos eliminados, como, por exemplo, dados de identificação pessoal ou assinatura; todas as cópias dos documentos eliminados, inclusive cópias de segurança e cópias de preservação, independentemente do suporte, deve ser destruídas. (CONARQ, 2011, p.30).

Semelhante às questões ligadas ao espaço utilizado por documentos físicos em um arquivo, o espaço liberado por um documento digital após o seu descarte pode ser ocupado por novos documentos. Consequentemente, descartar a informação digital, além de satisfazer

¹²⁷ *Storages* são dispositivos que permitem o armazenamento de grandes volumes de informações de forma descentralizada.

à confidencialidade (ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS, 2006), libera espaço nos dispositivos de armazenamento.

Ao descartar-se a informação em suporte digital, é preciso considerar o tipo de mídia na qual a informação está registrada e sua possível reutilização. O descarte deve considerar os requisitos de gestão de segurança da informação, de tal forma que impossibilite a recuperação posterior dessa informação (seja esta recuperação de forma intencional ou não).

De acordo com o *National Institute of Standards and Technology* (2012, p. 39, tradução nossa), o termo sanitização é utilizado para descrever um processo capaz de “[...] tornar o acesso aos dados presentes na mídia inviáveis para um dado nível de esforço”, sendo este termo utilizado para identificar diferentes técnicas responsáveis por remover, de forma segura, informações existentes nos diversos dispositivos de armazenamento, a exemplo de discos rígidos, CDs, DVDs, cartões de memória, *pen drives*, *smartphones* e *tablets*, dentre outros. Nos casos em que a mídia será reutilizada, aplicam-se métodos que se sobrescrevem às informações, a exemplo da técnica DoD 5220.22M, que consiste em sobreescrivendo a informação três vezes (UNITED STATES DEPARTMENT OF DEFENSE, 1995) e técnica criada por Peter Guttman, que consiste em sobreescrivendo a informação trinta e cinco vezes (DIESBURG; WANG, 2010) através de aplicativos como o File Shredder¹²⁸. Para os casos onde a mídia não será reutilizada, aplicam-se as técnicas de fusão, fragmentação, lixamento, pulverização, banho de ácido e desmagnetização, que podem variar de acordo com o tipo de mídia a ser descartada ou necessidade da organização que está descartando estas informações.

3 DESENVOLVIMENTO E PROCEDIMENTOS METODOLÓGICOS

Esta pesquisa foi desenvolvida como um estudo de caso, descritiva, com coleta de dados implementada de forma empírica realizada por caderno de anotações em laboratório, tendo como foco identificar procedimentos que viabilizem o descarte seguro da informação gravada em um disco rígido, utilizando-se de aplicativos específicos para tal fim.

Para o procedimento prático, utilizou-se de uma máquina virtual com o sistema operacional *Microsoft Windows 7 Thin PC* (versão de avaliação), com a configuração de 7 GBytes de disco rígido e utilizando-se do sistema de arquivos *New Technology File System* (NTFS). A configuração dispunha de 2 GBytes de memória RAM, processador Intel Core i5 64 bits e os softwares *WinHEX 16.7*¹²⁹ e o aplicativo gratuito *Recuva*¹³⁰ versão 1.51.1063, que

¹²⁸ File Shredder. Disponível em: <<http://www.fileshredder.org>>. Acesso em: 15 nov. 2013.

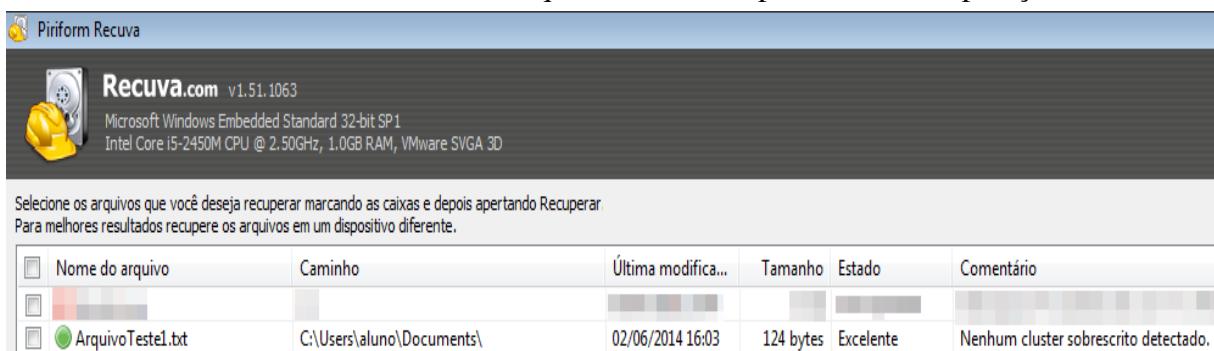
¹²⁹ WinHEX. Disponível em: <<http://www.winhex.com/>>. Acesso em: 28 abr. 2014

¹³⁰ Recuva. Disponível em <<https://www.piriform.com/recuva>>. Acesso em: 1 mai. 2014.

permite a recuperação de arquivos e pastas removidas previamente. Optou-se pela escolha das ferramentas *WinHEX* e *Recuva* por serem ferramentas indicadas na literatura. (BUCHNAN-WOLLASTON *et al.*, 2014; SCHWAMM, 2014; SHAVERS e ZIMMERMAN, 2014). Como sistema gerenciador de banco de dados (SGBD), optou-se pelo aplicativo Mysql¹³¹ versão 5.6.17 também de uso gratuito e por possuir versões de avaliação ou de uso livre (*opensource*).

A coleta de dados aconteceu em maio de 2014. A primeira parte do procedimento consistiu em se gravar um documento texto no disco e na remoção (apagamento) desse (“ArquivoTeste1.txt”) com a utilização do conjunto de teclas “SHIFT + DELETE” do computador e sua posterior procura no sistema de arquivos do disco rígido com a ferramenta *Recuva*. Verificou-se que mesmo com o apagamento do documento, a informação continuava presente no disco rígido do computador, conforme indicado na Figura 1, sendo passível de recuperação posterior.

FIGURA 1 - Documento “ArquivoTeste1.txt” passível de recuperação



Fonte: Dados da pesquisa.

Na segunda etapa do procedimento, buscou-se verificar se as informações armazenadas em um sistema gerenciador de banco de dados (SGBD) são removidas completamente de uma tabela após o uso do comando “delete”, presente nos SGBDs. Para a execução do procedimento, foi criado um banco de dados de nome “teste” e a tabela “trabalhos_academicos”. Nesta tabela foram inseridos cinco registros com informações e teses e dissertações da Universidade Federal da Paraíba (UFPB), com o objetivo de se ter informações para o procedimento.

Após a inserção das informações na tabela, removeu-se o primeiro registro utilizando-se o comando “delete” do SGBD e, posteriormente, foi utilizado o aplicativo WinHEX para

¹³¹ Mysql. Disponível em <<http://www.mysql.com>>. Acesso em: 5 mai. 2014.

verificar se a informação foi removida do disco rígido utilizado na simulação. A Figura 2 demonstra que a informação continua presente no dispositivo, mesmo após a sua exclusão.

FIGURA 2 – Informações excluídas em um SGBD: sem criptografia e criptografada

75 6D 00 02 00 0B 00 00 5A 81 69 1F 37 20 00 00 00 07 06 86 00 00 01 39 74 65 74 75 72 61 20 64 61 63 61 6F 20 61 70 6C 6C 65 69 74 6F 72 65 73 6F 6B 53 49 4C 56 41 2C	infimum supremumZ i 7 ÿí 9 pArquitetura d a Informacao apl icada a leitores de e-bookSILVA,	75 6D 00 02 00 0B 00 00 5A 81 69 1F 40 20 00 00 00 07 0C 08 00 00 01 CD 7B EC 89 86 37 DE 81 27 1A 0C 75 C1 65 1A 27 0E E6 02 0F 40 84 4A 36 55 CE 0E 66 0E 48 D2 8B D4 2C 20 4D 61 72 69 61 20	infimum supremumZ i @ ÿí pEÖNS{i /í < Ga ÿcs uÁe áT+ I.Íæ @ J6U í-ó* \$ÓI f HÓU IIISILVA, Maria
---	---	--	---

Fonte: Dados da pesquisa.

Portanto, foram evidenciados que mesmo depois de emitidos os comandos usuais para eliminação de um arquivo digital, seja em uma tabela de banco de dados ou em sistema de arquivo, as informações gravadas no disco rígido continuam presentes conforme indicado na literatura por Farmer e Venema (2007), o que justifica o desenvolvimento de procedimentos que possibilitem o descarte dessas informações.

Conforme verificado, a exclusão de determinado registro ou documento armazenado nos SGBDs e no NTFS não resultou na remoção segura da informação ali contida. Portanto, para as informações registradas em bancos de dados (SGBDs), optou-se por criptografar¹³² as informações que foram descartadas com uma senha aleatória, impossibilitando sua recuperação posterior, pois elas tornaram-se incompreensíveis após a realização do procedimento, de acordo com o resultado demonstrado na FIG. 2.

4 CONSIDERAÇÕES E PROPOSTA DE MODELO PARA O DESCARTE SEGURO

Diversas verificações são necessárias para que o descarte ocorra de forma segura, a exemplo da reutilização da mídia informática e a forma como as informações estão registradas em documentos digitais ou em registros em um banco de dados. Alguns procedimentos de descarte inviabilizam o uso posterior da mídia informática. No caso dos SGBDs que são utilizados em sistemas de informação de muitas organizações, por exemplo, onde os equipamentos não podem ser destruídos nem os sistemas podem ser suspensos por longos períodos de tempo para aplicação de uma técnica de sobreescrita, os registros precisam ser criados no formato binário¹³³ e, para documentos armazenados no sistema de arquivos, em

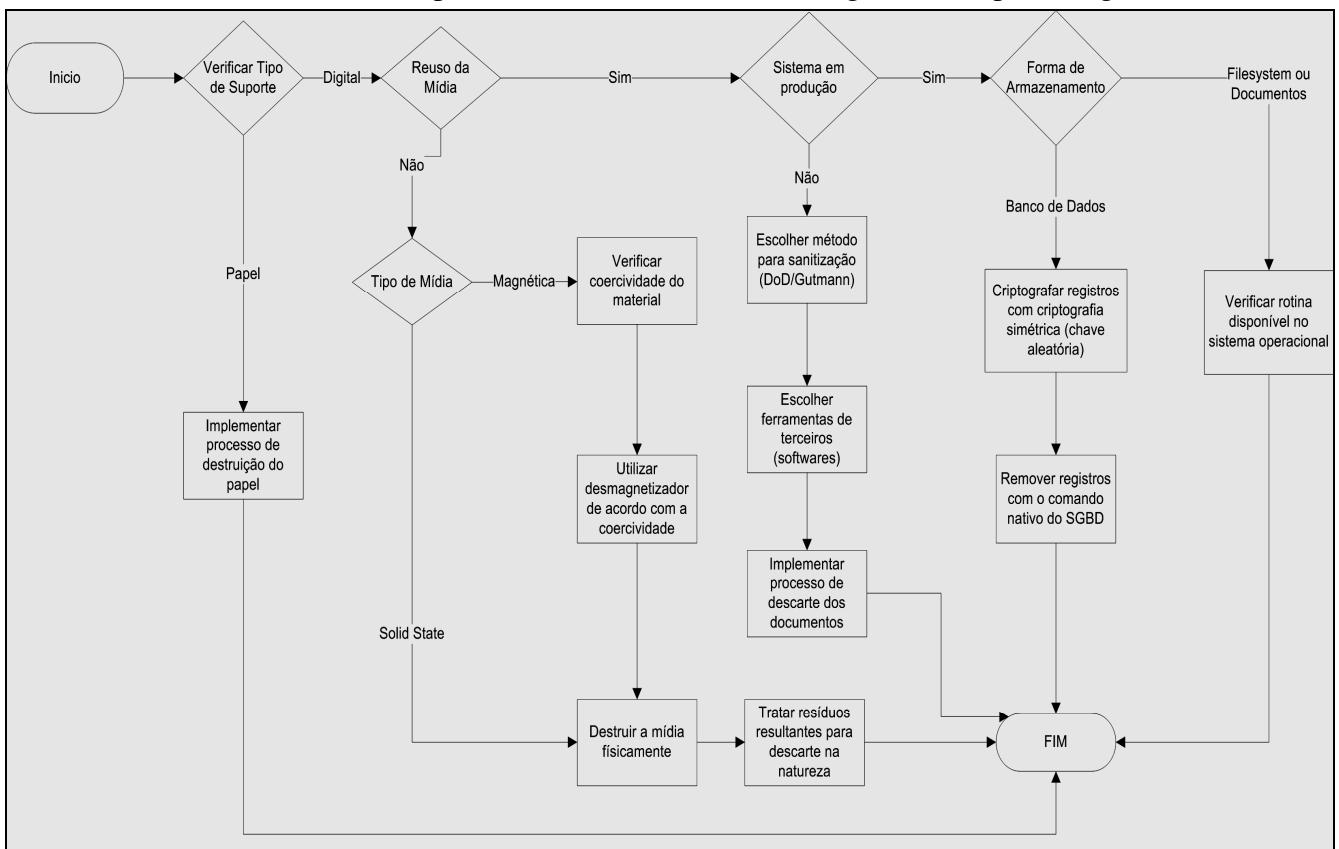
¹³² O processo de criptografia consiste em transformar a informação deixando-a ilegível através de uma chave (senha), onde apenas o detentor dessa chave é capaz de decifrá-la, trazendo novamente a informação para a sua forma original.

¹³³ Os campos em formato binário permitem que dados em linguagem de máquina sejam registrados neles.

sistemas em produção, é necessário um estudo mais detalhado, pois a rotina para a exclusão segura varia de acordo com o sistema operacional utilizado.

Portanto, ao descartar informações em suporte digital, é necessário considerar se a mídia será reaproveitada ou não. O modelo ilustrado na FIG. 3 apresenta uma proposta de modelo para descarte seguro da informação em suporte digital.

FIGURA 3 – Proposta de Modelo de Descarte Seguro em Suporte Digital



Fonte: Elaborado pelos autores.

Neste trabalho, os níveis de classificação da informação não foram avaliados, considerando-se apenas que a informação, independentemente de sua classificação, precisa ser eliminada de forma segura após o fim de sua utilidade. Ademais, a remoção segura em *storages* também não foi considerada por esses serem mais complexos e necessitarem, assim, de um estudo mais aprofundado do seu funcionamento.

Um dispositivo que não foi descartado corretamente pode conter informações que comprometem a reputação de uma organização. O descarte seguro da informação em suporte digital é uma temática nova e que precisa ser levada em consideração não somente nas organizações, mas em toda a sociedade. Não são raros os casos de informações recuperadas em dispositivos que foram descartados sem que a devida importância fosse dada às informações ali contidas.

REFERÊNCIAS

- ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS – ABNT. **ABNT NBR ISO/IEC 27001**: Tecnologia da informação — Técnicas de segurança — Sistemas de gestão de segurança da informação — Requisitos. Rio de Janeiro-RJ: ABNT, 2006.
- BEAL, Adriana. **Segurança da Informação**: princípios e melhores práticas para a proteção dos ativos de informação nas organizações. São Paulo: Atlas, 2008.
- BUCHANAN-WOLLASTON *et al.* A comparison of forensic toolkits and mass market data recovery applications. In: NINTH ANNUAL IFIP WG 11.9 INTERNATIONAL CONFERENCE ON DIGITAL FORENSICS, 9., 2013, Florida, USA. **Anais eletrônicos...** Florida: National Center for Forensic Science Orlando, 2013. Disponível em: <<http://eprints.gla.ac.uk/71698/>>. Acesso em: 13 jul. 2014.
- CAPURRO, Rafael; HJØRLAND, Birger. O conceito de informação. **Perspectivas em Ciência da Informação**, v. 12, n. 1, p. 148-207, 2007. Disponível em: <<http://portaldeperiodicos.eci.ufmg.br/index.php/pci/article/view/54/47>> Acesso em: 24 nov. 2013.
- CONSELHO NACIONAL DE ARQUIVOS. **e-ARQ Brasil**: Modelo de Requisitos para Sistemas Informatizados de Gestão Arquivístico de Documentos. Rio de Janeiro, 2011. Disponível em: <<http://www.documentoseletronicos.arquivonacional.gov.br/media/e-arq-brasil-2011-corrigido.pdf>>. Acesso em: 16 jul. 2014.
- DIESBURG, Sarah; WANG, Andy. A survey of confidential data storage and deletion methods. **ACM Computing Surveys**, v. 43, n. 1, art. 2, 2010. Disponível em: <<http://dl.acm.org/citation.cfm?id=1824797>>. Acesso em: 18 jul. 2014.
- FARMER, Dan; VENEMA, Wietse. **Perícia Forense Computacional**: Teoria aplicada a prática. São Paulo: Pearson Prentice Hall, 2007.
- LE COADIC, Yves-François. **A ciência da Informação**. 2. ed. Brasília: Briquet de Lemos/Livros, 2004.
- NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY. **Guidelines for Media Sanitization**: Recommendations of the National Institute of Standards and Technology Special Publication Draft 800-88r1. Maryland: U.S Government Printing Office, 2012. Disponível em: <http://csrc.nist.gov/publications/drafts/800-88-rev1/sp800_88_r1_draft.pdf>. Acesso em: 15 set. 2013.
- SHAVERS, Brett; ZIMMERMAN, Eric. **X-Ways Forensics Practitioner's Guide**. Maryland: Elsevier, 2014.
- SCHWAMM, Riqui. **Effectiveness of the factory reset on a mobile device**. 2014. 67f. Dissertação (Mestrado em Ciência da Computação) – Department of Computer Science, Naval Postgraduate School, Monterey, U.S., 2014.
- UNITED STATES DEPARTMENT OF DEFENSE. **5220.22-M**: National Industrial Security Program Operating Manual. Washington: U.S. Government Printing Office, 1995.

ACESSIBILIDADE WEB NA PERSPECTIVA DOS SISTEMAS DE RECUPERAÇÃO DA INFORMAÇÃO

WEB ACCESSIBILITY IN VIEW OF INFORMATION RETRIEVAL SYSTEMS

Lucinéia Souza Maia

Resumo: A Ciência da Informação busca compreender as demandas informacionais dos indivíduos em suas atividades e contextos através da mediação da informação. Assim, as informações são organizadas para que possam ser acessadas pelos usuários e, a partir de uma seleção feita por eles, ser apropriada. O acesso à informação anula a distância entre o usuário e a informação por meio de um canal, físico ou virtual, que disponibiliza a informação para o usuário através de suportes, tais como os Sistemas de Recuperação da Informação. Contudo, problemas de acessibilidade nesses sistemas são fatores de exclusão de usuários, principalmente aqueles com algum tipo de deficiência. Nesse cenário, em sistemas Web, tais problemas afetam a navegação, o entendimento, as operações e a interação do usuário com a informação. Neste sentido, o objetivo deste trabalho é explorar a literatura a fim de verificar os esforços das pesquisas em Ciência da Informação, no âmbito dos Sistemas de Recuperação da Informação, no que diz respeito à acessibilidade Web.

Palavras-chave: Acessibilidade Web. Sistemas de Recuperação da Informação. Mediação da Informação.

Abstract: Information Science seeks to understand the informational demands of individuals in their activities and contexts through the mediation of information. Thus, the information is organized so that they can be accessed by users and, from a selection made by them to be appropriate. Access to information nullifies the distance between the user and the information through a channel, physical or virtual, which provides information to the user through media such as Information Retrieval Systems. However, accessibility issues in these systems are deleting users, particularly those with a disability factors. In this scenario, Web systems, such problems affect navigation, understanding, operations and user interaction with information. In this sense, the objective of this study is to explore the literature in order to verify the efforts of research in Information Science, under the Information Retrieval Systems, with regard to Web accessibility.

Keywords: Web Accessibility. Information Retrieval Systems. Mediation of Information.

1 INTRODUÇÃO

A mediação da informação engloba a organização, o acesso e a transferência da informação de forma a satisfazer aos indivíduos em suas demandas informacionais em suas atividades e contextos (SMIT, 2009). Pode-se dizer, sumariamente, que a organização da informação é um processo, o acesso é a interação do usuário com a informação anteriormente organizada e, a transferência é a apropriação da informação pelo usuário.

No entanto, problemas relacionados à acessibilidade podem acarretar em barreiras que impeçam o acesso às informações pelas pessoas com deficiência, o que afeta a relação organização-acesso-transferência que são imprescindíveis para a mediação da informação.

De acordo com o W3C/WAI (2005), a acessibilidade diz respeito à diminuição ou à eliminação de barreiras que impeçam o acesso a algo. No contexto Web, estas barreiras configuram-se em obstáculos que dificultam a percepção, o entendimento, a navegação e a interação com o conteúdo pelo usuário (HARPER; YESILADA, 2008).

Neste artigo, o acesso à informação pelo usuário será abordado sob a perspectiva dos Sistemas de Recuperação da Informação (SRI). Esses sistemas extraem informações a partir de uma base documental de acordo com a comparação entre as necessidades de informação dos usuários com os documentos existentes na base de dados. A utilização de SRI é uma importante estratégia de mediação da informação, assim é necessário que estes sistemas sejam desenvolvidos de forma a contemplar todos os usuários. Neste contexto, o objetivo deste trabalho é explorar a literatura a fim de verificar os esforços das pesquisas em Ciência da Informação, no âmbito dos SRI, no que diz respeito à acessibilidade Web.

2 RECUPERAÇÃO DA INFORMAÇÃO

De acordo com Lopez-Yepes (1991) a informação pode ser contextualizada sob a perspectiva de um processo comunicativo que abrange as necessidades de informação do usuário, os canais de informação e sua tipologia, e o conceito de transferência de informação.

As necessidades de informação estão fortemente relacionadas aos usuários da informação. Segundo Le Coadic (1996, p.39), “O conhecimento da necessidade de informação permite compreender por que as pessoas se envolvem num processo de busca de informação”. De acordo com Souza (2006) e Vera (2011), esta busca de informação pode ser classificada em modelos de recuperação (estratégia de pergunta ou *querying*) e modelos de navegação (estratégia de exploração ou *browsing*).

No ambiente Web, nos modelos de navegação o usuário percorre os conceitos para encontrar a informação desejada. Por outro lado, nos modelos de recuperação, o usuário introduz palavras-chave ou termos relacionados que representam seus questionamentos, o sistema busca nas bases de dados o conteúdo e oferece ao usuário uma lista com os resultados pertinentes (VERA, 2011). Essa estratégia é suportada pelos SRI, que será abordado na próxima seção.

2.1 Sistemas de Recuperação da Informação

Um Sistema de Recuperação da Informação (SRI) lida com a representação e a recuperação de informações. Os SRI possuem informações que remetem a documentos, com exceção das bases de dados cadastrais e estatísticas.

No contexto das bibliotecas, os OPAC (*Online Public Access Catalog* ou Catálogos de Acesso Público *Online*) são um tipo de SRI com características próprias (CHU, 2003). Segundo Marchionini (1992), os OPAC são parte de sistemas integrados que permitem o acesso remoto a dezenas de catálogos de bibliotecas e outros serviços de informação.

De acordo com Falgueras e Codina (2005), os SRI (incluindo os OPAC) têm os seguintes componentes: a necessidade de informação, também denominada “pergunta” do usuário; o documento; as representações, tanto do documento quanto da necessidade de informação e; uma função de comparação entre a necessidade de informação e a representação dos documentos. Todos esses componentes devem ser contemplados nas interfaces do SRI. Assim, essas interfaces geralmente contêm o formulário de consulta, a lista de resultados e a visualização do documento completo.

2.1.1 Os usuários dos Sistemas de Recuperação da Informação

Segundo Ferreira (1995), a busca da informação é parte de um processo de tomada de decisão e de solução de problemas pelo indivíduo. Nesse contexto, o usuário busca a informação com algum objetivo, ou seja, a busca pela informação em si não é um ato isolado, ela é parte de algum outro processo. Desta forma, é importante conhecer os usuários para intermediar coerentemente as informações por eles demandadas, compreendendo os contextos de uso dos sistemas por eles.

Os contextos de uso dos sistemas podem ajudar a nortear a classificação dos usuários, assim é possível criar grupos de usuários cuja demanda e comportamentos em relação à informação são parecidos entre si (GUINCHAT; MENOU, 1994).

Outrossim, é fundamental compreender que alguns usuários podem ter características específicas que devem ser consideradas, tais como deficiências. As deficiências ou limitações físicas ou intelectuais dos usuários interferem na forma como eles têm acesso à informação. Os usuários com deficiência física e/ou motora, apresentam dificuldade em usar dispositivos apontadores como o mouse, usar o teclado ou operar programas que requerem uma resposta em período de tempo específico. Usuários com deficiência visual têm problemas com informação mostrada na tela, com o uso de dispositivos apontadores que requerem coordenação dos olhos-mãos (como o mouse) e problemas relacionados ao contraste de cores. Já os usuários com deficiência auditiva deparam-se com barreiras na detecção de sons, na distinção informações com barulho de plano de fundo e quando a aplicação exige entrada de voz, pois muitas pessoas com problemas de audição também apresentam dificuldades na fala. Por fim, os usuários com deficiência cognitiva encontram barreiras quando o *site* exige

reconhecimento e retenção de informação, compreensão, envolvimento, identificação, escolha e implementação de soluções (KAVCIC, 2005; HARPER; YESILADA, 2008).

Todos esses problemas, típicos do ambiente Web, muitas vezes são replicados em SRI disponíveis nesse contexto.

Os problemas de acessibilidade afetam não só usuários com deficiência, mas usuários sem deficiência, porém com limitações temporárias, idosos e usuários iniciantes. Além disso, fatores relacionados ao *hardware*, *software* e ambiente do usuário, também podem interferir no acesso à informação.

3 ACESSIBILIDADE WEB

A acessibilidade *Web* refere-se à leis, regulamentações e recomendações que buscam garantir o acesso ao conteúdo *Web* por todos os usuários.

No Brasil, a garantia da acessibilidade foi regulamentada por meio do Decreto-Lei 5.296, em dezembro de 2004 (BRASIL, 2004). E, em janeiro de 2005 foi criado o e-MAG (Modelo de Acessibilidade do Governo Eletrônico) como uma recomendação para padronizar a acessibilidade dos sites e portais do governo eletrônico (BRASIL, 2005).

Mundialmente, a principal recomendação de acessibilidade *Web* existente é o WCAG (*Web Content Accessibility Guidelines*) (W3C/WAI, 2008), criado pelo W3C/WAI (*World Wide Web Consortium/Web Accessibility Initiative*) em 1999.

As recomendações de acessibilidade têm como objetivo encorajar os projetistas a desenvolverem sites conforme especificações, possibilitando a conformidade com tecnologias assistivas¹³⁴ para que usuários com deficiência possam interagir com o conteúdo *Web* (PACIELLO, 2000; THATCHER et al., 2006; HARPER; YESILADA, 2008). Estas recomendações também são utilizadas para embasar as avaliações de acessibilidade *Web*.

As avaliações de acessibilidade *Web* podem ser realizadas em uma página *Web*, em uma coleção de páginas, em processos específicos de uma aplicação *Web* ou em todo o site. Para isso diferentes abordagens técnicas podem ser utilizadas, tais como: avaliação com ferramentas automatizadas, avaliação de especialistas em acessibilidade e testes com usuário (ABOU-ZAHRA, 2008).

¹³⁴ Ferramentas que dão suporte ao acesso de pessoas com deficiência na Internet, como por exemplo, os leitores de tela.

4 ABORDAGENS DE ACESSIBILIDADE EM CIÊNCIA DA INFORMAÇÃO

Esta Seção apresenta um levantamento sobre as abordagens de acessibilidade no contexto da Ciência da Informação (no escopo da mediação da informação), com foco nos SRI, a fim de compreender como a acessibilidade Web dos SRI está posicionada nas pesquisas que envolvem a Ciência da Informação. A pesquisa baseou-se em uma amostra de 26 artigos. Esses artigos foram recuperados a partir do Portal de Periódicos CAPES/MEC¹³⁵ e do Google Acadêmico¹³⁶. Uma pré-seleção foi realizada através da leitura do título e *abstract* dos trabalhos retornados, nessa fase, os trabalhos que tratavam de acessibilidade Web superficialmente foram desconsiderados. Após, os trabalhos pré-selecionados foram lidos na íntegra e, novamente selecionados de acordo com a relevância.

Para a interpretação dos dados, os trabalhos foram categorizados em: trabalhos abordando especificamente acessibilidade Web (53,85%) e trabalhos abordando acessibilidade no geral (46,15%). Outrossim, foram categorizados outros três grandes grupos: bibliotecas (69,23%), Ciência da Informação (19,23%) e SRI (11,54%). Sabe-se que a biblioteca e o SRI estão inseridos na Ciência da Informação, contudo, para fins desse trabalho, resolveu-se separá-los.

O QUADRO 1 apresenta os trabalhos relacionados à bibliotecas e acessibilidade Web (34,61%). Os trabalhos foram agrupados e apresentados em ordem cronológica de publicação. Como é possível observar, alguns trabalhos abrangem mais de um grupo.

QUADRO 1 – Trabalhos relacionados à bibliotecas e acessibilidade Web

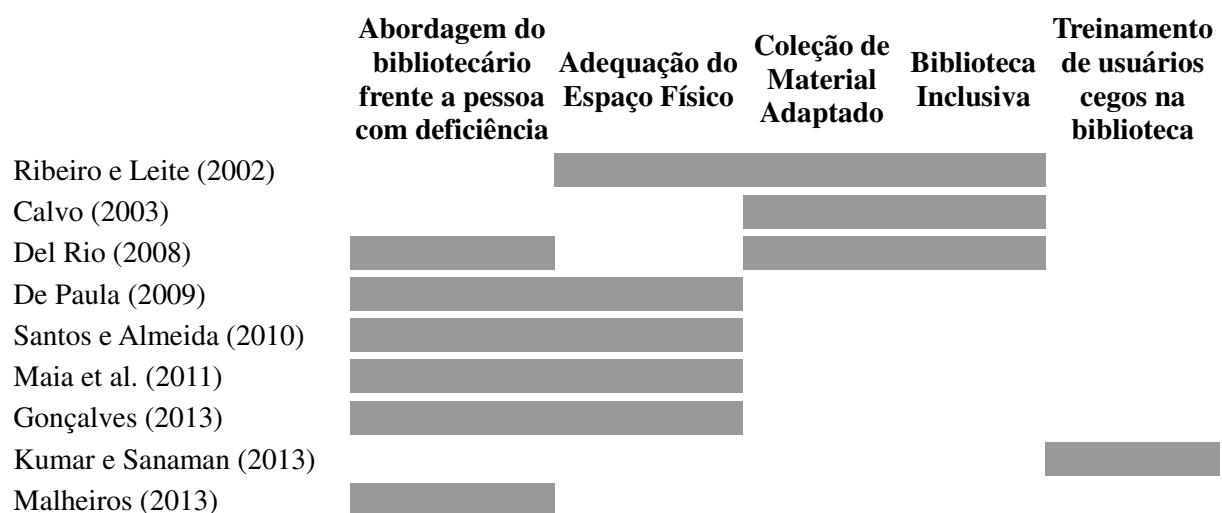
	Biblioteca no Geral	Biblioteca Virtual	Biblioteca Digital	Site da Biblioteca
Casey (1999)				
Guenther (2001)				
Fulton (2008)				
Wijayaratne e Singh (2010)				
Booth(2012)				
Lee et al. (2012)				
Riley-Huff (2012)				
Tatomir e Tatomir (2012)				
Cervone (2013)				
Higgins (2013)				

¹³⁵ <http://www.periodicos.capes.gov.br/>

¹³⁶ <http://scholar.google.com.br/>

Este trabalho limita-se à acessibilidade Web, contudo, como foram muitos trabalhos relacionados às bibliotecas e acessibilidade no geral (34,61%), resolveu-se investigar as abordagens relatadas nos artigos, conforme apresentado no Quadro 2.

QUADRO 2 – Trabalhos que abordam bibliotecas e acessibilidade no geral



No que diz respeito aos trabalhos específicos da Ciência da Informação, 11,53% abordaram acessibilidade no geral. Esses trabalhos dizem respeito aos serviços de informação acessíveis (CAREY, 2007; RIBERA et al., 2009; MARTÍN, 2013). Os demais, que referem-se à acessibilidade Web (7,69%), dizem respeito ao conhecimento do assunto na Ciência da Informação (INGLE et al., 2009; COMEAUX, 2001).

As pesquisas realizadas investigando especificamente acessibilidade Web dos SRI não obtiveram sucesso. Contudo, as pesquisas por acessibilidade Web dos OPAC (que são um tipo de SRI) foram bem sucedidas (11,53% dos trabalhos). Esses trabalhos, no geral, dizem respeito aos problemas de acessibilidade Web encontrados nas OPAC e apresentam estratégias para torná-los acessíveis (VAN BODENGRAVEN; POLLITT 2003; PETERS; BELL, 2006; ASTLEITNER, 2008).

Dos trabalhos selecionados no Quadro 2, três deles abrangem, dentro dos serviços oferecidos pela biblioteca, o uso de TIC (Tecnologias de Informação e Comunicação) por pessoas com necessidades especiais (MAIA et. al, 2011); treinamento de pessoas com deficiência que incluem os OPAC, recursos de acesso à Web e acesso e uso de ferramentas de busca e portais da biblioteca (

KUMAR e SANAMAN, 2013) e; no caso da biblioteca inclusiva, proposta por Ribeiro e Leite (2002), serviços como: consulta do OPAC da biblioteca da Faculdade, acesso à

Internet, consulta de textos digitalizados, consulta de índices ou partes de obras escaneadas por pessoas com deficiência.

5. CONCLUSÃO

O artigo buscou explorar a literatura a fim de verificar os esforços na Ciência da Informação que dizem respeito à acessibilidade Web dos SRI. O trabalho não considerou as publicações da área que referiam-se à acessibilidade Web no geral, ou no aspecto social.

Muitos trabalhos de acessibilidade encontrados referiram-se a bibliotecas. Esses trabalhos englobaram a adaptação do espaço físico das bibliotecas, a abordagem do bibliotecário frente a uma pessoa com deficiência no espaço da biblioteca e bibliotecas inclusivas e bibliotecas especializadas para pessoas com deficiência. Neste último caso, a adaptação de materiais para pessoas com deficiência foi abordada quase que exclusivamente nessas bibliotecas. Outrossim, com relação à adaptação de material, também foi possível perceber uma preocupação em atender pessoas com vários tipos de deficiência, incluindo deficiência intelectual e auditiva.

No que diz respeito à acessibilidade Web dos SRI, foram encontrados trabalhos relacionados à biblioteca digital, à OPAC e ao uso de TIC por pessoas com deficiência. Esses trabalhos não foram maioria, contudo, qualitativamente os trabalhos foram muito bem colocados.

Conforme foi relatado, não foram encontrados trabalhos que abordam especificamente acessibilidade Web como requisito de um SRI, entretanto, isso não quer dizer também que os SRI não são acessíveis, já que profissionais de TI são contratados para o desenvolvimento desses.

Por fim, entende-se que, garantir o direito de acesso à informação a todos os usuários independente de suas limitações também é um papel dos profissionais que desenvolvem estratégias para mediar a informação. A Internet é um mecanismo importante na inclusão de pessoas com deficiência e os SRI podem contribuir para a intermediação da informação para esses usuários, assim, é fundamental que a acessibilidade Web seja mais difundida entre a comunidade de especialistas em informação.

REFERÊNCIAS

ABOU-ZAHRA, S. Web Accessibility Evaluation, In: HARPER, S.; YESILADA, Y. (editors). Web Accessibility – A Foundation for Research. London: Springer-Verlag Limited, 2008. p. 79-106.

ASTLEITNER, W. Accessibility vs. Aleph OPAC. JKU Linz/Information Management, nov. 2008. Disponível em:
http://www.exlibrisgroup.org/download/attachments/13766794/access_webopac.pdf?version=1&modificationDate=1228400463000 Acesso em: jan.2014

BELL, L.; PETERS, T. AccessibleIT: Making OPACs blow. Computers in libraries, v.26, n.8, p. 20-22, 2006.

BOOTH, C. (Ed.). **Making Libraries Accessible: Adaptive Design and Assistive Technology.** American Library Association, 2012.

BRASIL. Decreto-lei 5296, de 2 de dezembro de 2004. Regulamenta as Leis nºs 10.048, de 8 de novembro de 2000, que dá prioridade de atendimento às pessoas que especifica, e 10.098, de 19 de dezembro de 2000, que estabelece normas gerais e critérios básicos para a promoção da acessibilidade. Publicado no D.O.U, nº 232, 03 de dezembro de 2004.

BRASIL. Recomendações de Acessibilidade para a Construção e Adaptação de Conteúdos do Governo Brasileiro na Internet. e-MAG, Acessibilidade de Governo Eletrônico. Modelo de Acessibilidade. Departamento de Governo Eletrônico, Secretaria de Logística e Tecnologia da Informação, Ministério do Planejamento, Orçamento e Gestão. Doc. de Ref., versão 2.0. 2005.

BRASIL. e-MAG Modelo de Acessibilidade em Governo Eletrônico. Ministério do Planejamento, Orçamento e Gestão, Secretaria de Logística e Tecnologia da Informação; Ministério da Educação, Secretaria de Educação Profissional e Tecnológica - Brasília : MP, SLTI, 2011.

CALVO, F. J. M. La biblioteca pública y los lectores con discapacidad visual: una reflexión sobre el caso español. **Integración: Revista sobre ceguera y deficiencia visual**, v.41, p. 15-21, 2003.

CAREY, K. The opportunities and challenges of the digital age: A blind user's perspective. **Library Trends**, v. 55, n. 4, p. 767-784, 2007.

CASEY, C. Accessibility in the virtual library: Creating equal opportunity Web sites. **Information technology and libraries**, v. 18, n. 1, p. 22-25, 1999.

CERVONE, F. H. Selected practices and tools for better accessibility in digital library projects. **OCLC Systems & Services**, v. 29, n. 3, p. 130-133, 2013.

CHU, H. Information representation and retrieval in the digital age. **Information Today, Inc., 2003.**

COMEAUX, D.; SCHMETZKE, A. Accessibility of academic library web sites in North America: Current status and trends (2002-2012). **Library Hi Tech**, v. 31, n. 1, p. 8-33, 2013.

DE PAULA, S. N. Acessibilidade à Informação em Bibliotecas Universitárias e a Formação do Bibliotecário. Dissertação. Centro de Ciências Sociais e Aplicadas, Pontifícia Universidade Católica de Campinas, Campinas, 2009.

DEL RÍO, M. E. P. **Información sin barreras, formación para ti.** La experiencia de la Biblioteca Insular Especializada en Discapacidad de Tenerife. p. 423 – 431, 2008.

FALGUERAS, E. A.; CODINA, L. **Bases de datos documentales:** características, funciones y método. Madrid: Editorial Síntesis, 2005.

FERREIRA, S. M. S. P. Novos paradigmas e novos usuários de informação. **Ciência da Informação**, v. 25, n. 2, p. 1-10, 1995.

FULTON, C. Web accessibility, libraries, and the law. **Information Technology and Libraries**, v. 30, n. 1, p. 34-43, 2011.

GONÇALVES, E. F. P. As Tecnologias Assistivas e a Atuação do Bibliotecário como Intermediário entre as Fontes de Informação e o Deficiente Visual. **Múltiplos Olhares em Ciência da Informação**, v. 2, n.1, p. 1-9, 2013.

GUENTHER, K. Making your digital library accessible. Computers in Libraries-Westport, v.21, n. 9, p. 61-63, 2001.

GUINCHAT, C.; MENOU, M. **Introdução geral às ciências e técnicas da informação e documentação**. 2. ed. Brasília: IBICIT, 1994.

HARPER, S.; YESILADA, Y. Web Accessibility and Guidelines. In: HARPER, S.; YESILADA, Y. (Ed.). **Web Accessibility – A Foundation for Research**. London: Springer-Verlag Limited, 2008. p. 61-78.

HIGGINS, S. **Access to Digital Libraries for Disadvantaged Users**. Library Philosophy and Practice (e-journal). Paper 916, 2013. Disponível em: <<http://digitalcommons.unl.edu/libphilprac/916>> Acesso em: jan. 2014.

INGLE, E. et al. How Accessible Are Public Libraries' Web Sites? A Study of Georgia Public Libraries. **Journal of Access Services**, v.6, n.1, p.101-115, 2009.

ISO/IEC 9126/1. **Information technology—Software product quality — Part 1: Quality model**, 2001.

KAVCIC, A. Software Accessibility: Recommendations and Guidelines. In: Computer as a Tool, The International Conference on EUROCON 2005. **Proceeding...** Belgrade, Serbia and Montenegro, IEEE, v.2, p.1024 -1027, 2005.

KUMAR, S.; SANAMAN, G. Orientation/Training Programmes in NCR Libraries for People with Disabilities: Study of User's Perspective. **Library Philosophy and Practice** (e-journal). Paper 927, 2013. Disponível em: <<http://digitalcommons.unl.edu/libphilprac/927>> Acesso em jan. 2014.

LE COADIC, Y. F. **A ciência da informação**. Briquet de Lemos Livros, 1996.

LEE, S. et al. Revising Cataloging Rules and Standards to Meet the Needs of People with Disabilities. **Library Resources & Technical Services**, v.57, n. 1, p.18-29, 2013.

LOPEZ-YEPES, J. El desarrollo de los sistemas de información y documentación. **Revista general de información y documentación**, v.1, n.2, p. 23-34., 1991.

MAIA, M. A. et al. O Bibliotecário como Mediador no Processo de Transferência da Informação para Pessoas com Deficiência Visual. In: XXIV Congresso Brasileiro de Biblioteconomia, Documentação e Ciência da Informação, Maceió, 2011. **Anais...**

MALHEIROS, T. M. C. **Necessidade de informação do usuário com deficiência visual:** um estudo de caso da biblioteca digital e sonora da Universidade de Brasília. Dissertação - Faculdade De Ciência Da Informação - Universidade De Brasília. Brasília, 2013.

MARCHIONINI, G. Interfaces for end-user information seeking. **Journal Of The American Society For Information Science**, v.43, n. 2, p. 156-163, 1992.

MARTÍN, S. G. Visibilidad y accesibilidad web de las tesinas de licenciatura en Bibliotecología y Documentación en la Argentina. **Información, cultura y sociedad**, v.28, 2013.

PACIELLO, M. G. **Web accessibility for people with disabilities.** CMP Books. Kansas-USA, 2000.

RIBEIRO, A.; LEITE, J. Contributos para um conceito de "Biblioteca Inclusiva". **Integrar**, n. 19, p. 50-56, 2002, Disponível em: <<http://repositorio-aberto.up.pt/handle/10216/18473>> Acesso em: jan. 2014.

RIBERA, M. *et al.* Web Content Accessibility Guidelines 2.0 - A further step towards accessible digital information. **Program: Eletronic Library and Information Systems**, v. 43, n. 4, p. 392-406, 2009.

RILEY-HUFF, D. A. Chapter 4: Web Accessibility and Universal Design. In: **Library Technology Reports**, v.48, n. 7, p. 29-35, 2012.

SANTOS, C. G. D.; ALMEIDA, E. M. **Estudo do usuário com deficiência visual:** um importante instrumento sócio-inclusivo de pesquisa e formação para o profissional de biblioteconomia perante as necessidades informacionais de pessoas com deficiência. XVI Seminário de Pesquisa do CCSA, 2010.

SMIT, J. W. Novas abordagens na organização, no acesso e na transferência da informação. In: SILVA, H. de C.; BARROS, M. H. T. C. de (Org.). Ciência da Informação: múltiplos diálogos. Marília: Oficina Universitária UNESP, p. 57-66, 2009.

SOUZA, R. R. Sistemas de recuperação de informações e mecanismos de busca na web: panorama atual e tendências. **Perspectivas em Ciência da Informação**, v.11, n.2, p. 161-173, 2006.

TATOMIR, J. N.; TATOMIR, J. C. Collection Accessibility: A Best Practices Guide for Libraries and Librarians. **Library Technology Reports**, v.48, n.7, p.36-42, out, 2012,

THATCHER, J. *et al.. Web Accessibility:* Web Standards and Regulatory Compliance. Springer-Verlag New York - USA, 2006.

VAN BODENGRAVEN, M.; POILITT, C. Making Websites and OPACS accessible. **IFLA journal**, v. 29, n. 4, p. 357-363, 2003.

VERA, T. P. **Organización y Representación del Conocimiento:** Incidencias de las Tecnología de la Información y Comunicación. Buenos Aires: Alfagrama, 2011.

W3C/WAI. **Introduction to Web Accessibility.** 2005. Disponível em: <<http://www.w3.org/WAI/intro/accessibility.php>>. Acesso em: set. 2013

W3C/WAI. **Web Content Accessibility Guidelines (WCAG) 2.0. 2008.** Disponível em: <<http://www.w3.org/TR/WCAG20>>. Acesso em: set. 2013

WIJAYARATNE, A. ; SINGH, D. Is There Space in Cyberspace for Distance Learners with Special Needs in Asia? A Review of the Level of Web Accessibility of Institutional and Library Homepages of AAOU Members. **International Information and Library Review**, v. 42, n. 1, p. 40-49, 2010.

ANÁLISE DA AÇÃO DE INFORMAÇÃO RELACIONAL NO LAVID DA UFPB

ANALYSIS OF ACTION OF RELATIONAL INFORMATION IN LAVID/UFPB

Emy Pôrto Bezerra
 Ítalo José Bastos Guimarães
 Zayr Cláudio Gomes da Silva

Resumo: Na sociedade contemporânea a colaboração científica vem sofrendo alterações nas suas estruturas dinâmicas através das redes de coautoria. A rede de coautoria dos docentes que compõem o Laboratório de Aplicações de Vídeo Digital (LAViD) da UFPB é analisada através da Análise Estrutural de Redes Sociais, a partir de métricas e representação gráfica geradas pelo software NodeXL. Tal rede é considerada aqui como produto da ação de informação relacional que compõe o Regime de Informação do LAViD. São descritas as propriedades (composição, estrutura e dinâmica) dessa rede. Foram analisadas métricas que compõem essas propriedades, como: grau de conexão, centralidade, laços sociais, agregação e desagregação. Percebemos uma forte centralização da coautoria a um determinado ator, com aparecimento de laços fortes e fracos entre este e a maioria dos atores, evidenciando o potencial grau de liderança deste na rede. Também pudemos observar um processo inicial de *clusterização*, demonstrado a partir de um movimento de cooperação, adaptação e organização por parte de alguns atores.

Palavras-chave: Ação de Informação Relacional. Coautoria. Análise de Redes Sociais.

Abstract: In contemporary society scientific collaboration has undergone changes in its dynamic structures through co-authorship networks. The co-authorship network of teachers comprising the Laboratory of Digital Video Applications (LDVA) UFPB is analyzed through the Structural Analysis of Social Networks from metrics and graphical representation generated by NodeXL software. This network is considered here as a product of the action of relational information that makes up the Scheme Information LDVA. (Composition, structure and dynamics) properties of such a network are described. Were analyzed metrics that make up these properties, such as degree of connection, centrality, social ties, aggregation and disaggregation. We noticed a strong centralization of co-authoring a particular actor, with the emergence of strong and weak ties between this and most actors, highlighting the potential degree of leadership in this network. We also observe an initial process of clustering, shown from a cooperative movement, adaptation and organization on the part of some actors.

Keywords: Action Relational Information. Coauthoring. Social Network Analysis.

1 INTRODUÇÃO

Com o presente artigo, propomos analisar a ação de informação relacional no âmbito do Regime de Informação (Ri) do Laboratório de Aplicações de Vídeo Digital (LAViD) do Departamento de Informática (DI) da Universidade Federal da Paraíba (UFPB).

Ao longo de sua história, a análise dos diversos fenômenos relacionados ao objeto da informação – permanece sendo o principal foco da pesquisa em Ciência da Informação (CI). No que se refere a Políticas e Gestão de Informação, estamos testemunhando um crescente e frutífero debate na elaboração do conceito de Regime de Informação (Ri). Este, por natureza, é constituído a partir de “ações de informação” realizadas por grupos de atores sociais no

contexto onde estão definidos os elementos que compõem o fluxo da produção, organização, comunicação e transferência de informações, em um dado espaço social.

A partir da descrição e interpretação da ação de informação relacional que constituem o *Ri* do LAVID, será possível situar a produção do laboratório no âmbito do *Ri* da sociedade em rede, bem como, especular sobre o desenvolvimento de uma inteligência coletiva no ambiente que é circunscrito por um específico *Ri*. Deste modo, trazemos aqui um resultado parcial da pesquisa ora em andamento. Esperamos aprofundar a discussão no tema, sem a pretensão de fazer um mapeamento exaustivo, mas preocupando-se em analisar a construção de espaços públicos de pesquisa nos quais se desenvolvam *Ri*.

2 DESENVOLVIMENTO

Na sociedade contemporânea, o campo teórico da gestão da informação abre cada vez mais a discussão sobre o valor da informação e a democratização do acesso às suas fontes na web. Neste contexto, autores como Frohmann (1984), González de Gómez (2012), Freire (2013), Braman (2004) e Ekbia (2009) vêm levantando o debate sobre as ações de informação que compõem o *Ri* de tipos de formações (instituições públicas ou privadas) sociais, na sociedade em rede. O presente artigo sustenta-se na afirmação de González de Gómez (2003, p. 34), que define Regime de Informação como

um modo de produção informacional dominante numa formação social, conforme o qual serão definidos sujeitos, instituições, regras e autoridades informacionais, os meios e os recursos preferenciais de informação, os padrões de excelência e os arranjos organizacionais de seu processamento seletivo, seus dispositivos de preservação e distribuição.

Aqui podemos inserir os laboratórios de fomento à pesquisa científica em instituições públicas federais. Neste sentido, laboratórios como o LAVID podem ser entendidos como exemplos de espaços de produção e compartilhamento de informação e conhecimento, cabendo sua análise à luz do conceito de *Ri*. Criado em 2003, o LAVID é atualmente uma referência nacional e internacional em desenvolvimento de tecnologia da TV Digital no país, contando com a colaboração de mais 60 jovens pesquisadores que estão interconectados com pesquisadores de todo o Brasil e do mundo, trazendo as atuais tendências tecnológicas mundiais nas áreas de vídeo e TV Digital.

2.1 Sobre o Regime de Informação no LAVID

O conceito de *Ri* vem sendo desenvolvido por diversos pesquisadores da CI nos últimos anos, sempre com características gerais similares, mas sem analisar as particularidades e necessidades do contexto (ambiente) onde se originam ou se instalaram.

Quando analisado no contexto acadêmico, prevalece no *Ri* as características do trabalho científico: a produção e o intercambio crítico de conhecimentos e experiências.

Tendo em vista que as ações de informação são essencialmente um fenômeno social, pois são partes integrantes da vida diária das comunidades e organizações, torna-se cada vez mais evidente que a integração entre essas ações e as tecnologias “colaborativas” transporte consigo uma profunda revisão no *Ri* de uma determinada instituição.

Assim, a elaboração e desenvolvimento deste artigo parte inicialmente da premissa de que a análise das ações de informação em um determinado “espaço de informação”, não só configura o desenvolvimento do *Ri* deste espaço, como também estabelece o processo de construção de uma inteligência coletiva entre os atores sociais participantes da rede que o comporta e o “alimenta”.

2.2. Sobre a metodologia

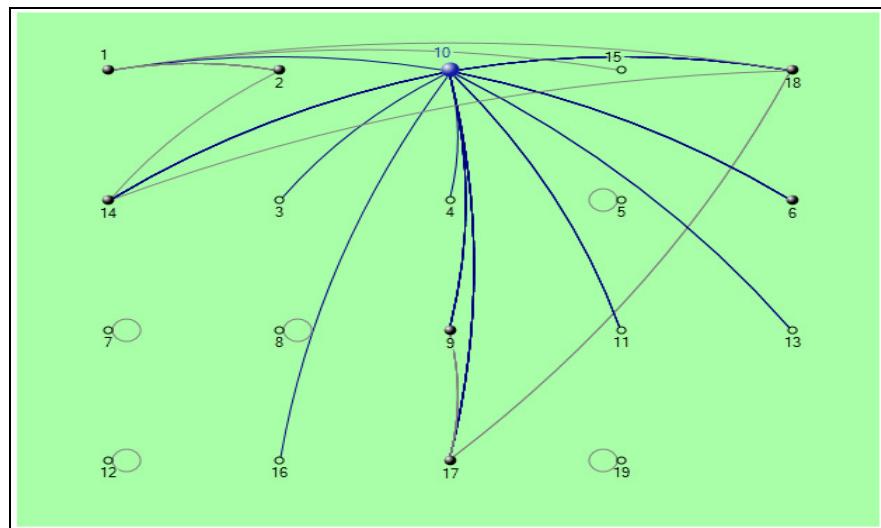
Nossa pesquisa teve caráter exploratório e descritivo. Sendo assim, exploramos e descrevemos as características da produção colaborativa dos docentes do LAViD e suas conexões através da Análise Estrutural de Redes Sociais (ARS), a partir de Fragoso *et al.*, (2013). Foram analisados os atores (docentes) da rede de coautoria do LAViD, visando identificar as suas propriedades de composição, estrutura e dinâmica. A metodologia utilizada para criação da base de dados da rede LAViD ocorreu por meio de quatro etapas: pesquisa documental da produção em coautoria dos docentes a partir dos seus currículos na Plataforma Lattes; construção da matriz referente a essa produção; transferência dos dados para o NodeXL e formatação da rede.

Tomamos como universo da pesquisa a produção científica colaborativa dos docentes do LAViD em anais de congressos. Também optamos por analisar essa produção apenas no triênio 2011-2012-2013. Justificamos nossos recortes por serem os anais uma das principais modalidades de divulgação da produção científica utilizada pelos pesquisadores do laboratório. Já o triênio pelo fato de ter sido este o último período de avaliação feita pelo MEC, o que se pressupõe uma atualização dos currículos por parte dos docentes. Como ferramenta para geração e análise da rede de coautoria do LAViD, utilizamos o software NodeXL, um software gratuito de código aberto, que permite ser programado e melhorado por outras pessoas. Ele utiliza o Excel como plataforma principal, tornando mais fácil a transferência e manipulação dos dados, permitindo a criação de gráficos de rede com boas visualizações por pessoas que não sejam programadoras.

3 DISCUSSÃO E INTERPRETAÇÃO DOS RESULTADOS PARCIAIS

Foram analisadas as produções em coautoria dos anais de congressos no triênio (2011, 2012, 2013). As propriedades dos dados foram analisadas a partir dos critérios de estrutura, composição e dinâmica da rede, para ter uma ideia desses critérios por meio de algumas métricas de ARS. Nesse sentido delimitamos a pesquisa à análise das métricas de laços sociais (enquanto composição), grau de conexão e centralidade (enquanto estrutura) e agregação e desagregação (enquanto dinâmica) da rede. A FIGURA 1 apresenta a estrutura da rede LAViD construída a partir dos dados processados pelo o software NodeXL.

FIGURA 1 – Rede de Coautoria do LAViD



Fonte: Dados da Pesquisa – Gerado pelo NodeXL

3.1 Quanto à Estrutura

No que se refere à estrutura percebemos que o ator 10 possui um alto grau de conexão na rede, visível na sua produção em coautoria com 11 dos 19 pesquisadores (aproximadamente 58% do total). Segundo Fragoso *et al.*, (2013, p. 122)

a estrutura compreende a forma da rede e as características que podem ser extraídas desta (laços sociais e capital social). A composição refere-se à qualidade dessa forma, ou seja, aos atores determinados e à qualidade de suas conexões (grau de conexão, densidade, centralidade e centralização). Os processos dinâmicos das redes são consequência direta dos processos de interação entre os atores (ordem, caos, agregação, desagregação e ruptura).

O grafo se apresenta com o formato de uma “estrela”, onde um só nó centraliza a maior parte das conexões. Segundo Fragoso et al. (2013), o grau de conexão é uma descrição de quantos nós compõem a vizinhança de um determinado nó. Este critério de avaliação comprova sua popularidade, evidenciando a sua centralidade para a rede como um todo.

3.2 Quanto à Composição

A composição da rede pode ser avaliada através dos laços sociais ou conexões estabelecidas entre os atores sociais. Granovetter (1973; 1983), por exemplo, discute que os laços sociais poderiam ser fortes e fracos, na medida em que conecte indivíduos que dividem valores sociais diferentes. No caso do LAViD e no que se refere a produção científica em coautoria, os laços mais fortes foram estabelecidos entre o ator 10 e os atores 6, 9, 14, 17 e 18 (esferas na rede) com os quais foram produzidos de seis à dez artigos em coautoria. Laços mais fracos também ocorrem com outros membros da rede (atores 3, 4, 11, 13 e 16). Os atores 1, 14, 17 e 18 possuem trabalhos em coautoria com outros membros além do ator 10. O que evidencia o início de uma possível e provável descentralização da rede (*clusterização*).

3.3 Quanto à Dinâmica

De acordo com Fragoso, Recuero e Amaral (2013), ao se estudar uma rede social é necessário pensar que a rede, apesar de representar uma estrutura estática, está sempre em movimento. Por se tratar de um laboratório de pesquisa científico-tecnológica, esse movimento é uma constante no LAViD. Desde o seu surgimento em 2003, muitos foram os pesquisadores que já passaram pelo laboratório em seus variados projetos. Os processos dinâmicos das redes são consequência direta dos processos de interação entre os atores. Redes são sistemas dinâmicos e, como tais, sujeitos a processos de ordem, caos, agregação, desagregação e ruptura (NICOLIS; PRIGOGINE, 1989). Por isso, no que se refere à dinâmica, optamos por analisar critérios de agregação e desagregação na rede com o intuito de observar como comporta esse fenômeno no LAViD.

Percebemos, portanto, que seis dos dezenove pesquisadores foram incorporados recentemente ao grupo docente que compõe o LAViD, não possuindo ainda produção em coautoria relevante para o período analisado, o que pode justificar o aparecimento de alguns nós isolados na rede (5,7, 8, 12 e 19). Outro motivo que pode justificar esse aparente “isolamento” é o fato de a análise ter sido feita apenas sobre a produção em coautoria em anais de congressos, não sendo analisadas outras produções em coautoria como artigos em periódicos, dissertações defendidas, capítulos de livros etc. A seguir apresentamos a análise dos resultados supracitados à luz da ação de informação relacional.

3.4 Quanto à Ação de Informação Relacional

A partir da análise dos dados da rede LAViD pudemos compreender a correlação existente entre ação de informação relacional e a produção científica em coautoria docente no

laboratório. As três propriedades de ARS e suas métricas serviram de subsídio para nossa interpretação. Elas nos mostram que a partir da estrutura, composição e dinâmica da rede, os atores interagiram direta e indiretamente entre si no processo de produção em coautoria. Esta por sua vez, não se mostra uniforme, e sim, centralizada a um único ator da rede, porém esboçando um processo natural de descentralização ou *clusterização*. Isso demonstra maior interação entre os outros atores da rede, caracterizando-se como uma ação de informação relacional.

Para González de Gómez (2003) as ações relacionais são “realizadas por sujeitos articuladores ou ‘relacionantes’, que executariam em grande parte uma forma de trabalho que teria a maior expansão no mundo contemporâneo: o trabalho relacional ou interativo para transformar a informação e a comunicação que orientam o agir coletivo”. Interpretamos aqui esse agir coletivo como o processo de *clusterização* da rede, ou seja, a produção em coautoria com outros membros, e o surgimento de novos “líderes”. Tal processo foi evidenciado na rede LAViD, uma rede onde seus atores são sujeitos sociais articuladores e reflexivos que exercem suas atividades sociais com o intuito de transformar a informação e a comunicação que orientam o agir coletivo para produção científica em coautoria.

5 CONSIDERAÇÕES FINAIS

Atualmente, estamos no início do caminho sobre a análise de dados de produção em coautoria na internet, principalmente com base em Análise de Redes Sociais. As discussões conceituais sobre a produção científica em coautoria, sua história, enfoques e temáticas, trazem possibilidades e limitações que se refletem nas escolhas do pesquisador, abrindo caminho para novos estudos e debates sobre o tema. Chamamos a atenção para o nosso recorte (a produção docente LAViD no triênio 2011-2012-2013). Este fator certamente interferiu de forma profunda na estrutura e dinâmica da rede LAViD. Salientamos que não foram analisadas aqui todas as produções científicas dos docentes, como: dissertações, artigos em periódicos científicos, livros, capítulos de livros etc. Apenas as produções em anais de congressos no referido período, o que interferiu tanto na sua forma estrutural, quanto no direcionamento das escolhas das métricas para sua análise.

Deste modo, sugerimos pesquisas que busquem alargar a rede a partir da inserção de novas métricas e atores, períodos e tipos de publicações, englobando assim, tanto quem fez como quem faz parte da Rede LAViD. Para trabalhos futuros, achamos pertinente a análise das ações de informação formativa e de mediação, além da agregação de outros métodos de pesquisa aos já existentes, como: etnografia virtual, bibliometria, observação participante,

webometria etc. Isto permitirá um retrato aprofundado e fiel da produção em coautoria nas instituições de pesquisa científica.

REFERÊNCIAS

- BRAMAN, S. The emergent global information policy regime, In: _____ (Ed.) **The emergent global information policy regime**. Hampshire, Palgrave, 2004.
- EKBIA, H.; EVANS, T. Regimes of information: land use, management, and policy. **The Information Society**, v. 2, v. 5, p. 328–343, 2009.
- FRAGOSO, S; RECUERO, R; AMARAL, A. **Métodos de pesquisa para a internet**. Porto Alegre: Sulinas, 2013.
- FREIRE, I. M. Sobre o Regime de Informação no Laboratório de Tecnologias Intelectuais – LTi. InCID. **R. Ci. Inf. e Doc.**, Ribeirão Preto, v. 4, n. 1, p. 70-86, jan./jun. 2013.
- FROHMAN, B. Talking information policy beyond information science: Applying the actor network theory. In H. A. Olson, & D. B. Ward (Eds.) **Proceedings of the 23rd Annual conference of the Canadian Association for Information Science**, 7–10 June 1995, Edmonton, Alberta. Disponível em: <http://www.caicsi.ca/proceedings.1995/frohmann_1995.pdf>. Acesso em: 12 jul. 2014.
- GONZÁLEZ DE GÓMEZ, M. N. Escopo e abrangência da Ciência da Informação e a Pós-Graduação na área: anotações para uma reflexão. **Transinformação**, Campinas, v. 15, n. 1, p.31-43, 2003.
- GONZÁLEZ DE GÓMEZ, M. N. Regime de informação: construção de um conceito. **Informação e Sociedade: Estudos**, João Pessoa, v. 22, n. 3, p. 43-60, set./dez. 2012.
- GRANOVETTER, M. **The Strength of Weak Ties**: network theory revisited. **Sociological Theory**, v. 1, p.203-233, 1983.
- NICOLIS, G.; PRIGOGINE, I. Exploring Complexity: an introduction. New York: W. H. Freeman end Company. 1989.

BIBLIOTECAS E WEB SEMÂNTICA: ANÁLISE SOBRE O ESTADO DESTE RELACIONAMENTO

LIBRARY AND SEMANTC WEB: ONE ANALYSIS ON THE STATE OF THIS RELATIONSHIP

Edgar Bisset Alvarez

Resumo: Apresenta-se uma análise sobre quais são as iniciativas adotadas pelas bibliotecas para se juntar ao movimento de desenvolvimento da Web Semântica. Apresenta-se brevemente como as Tecnologias de Informação e Comunicação têm impactado nos processos bibliotecários, e quais são alguns dos principais desafios a enfrentar no processo de adaptação ao novo contexto da Web Semântica, fazendo uma análise sobre como as bibliotecas poderiam beneficiar o desenvolvimento da Web semântica e apresentaram-se algumas ações baseadas em exemplos desenvolvidos pelas bibliotecas até a atualidade, para alcançar este propósito, analisou-se a evolução da Web nas últimas décadas, considerando as principais características, vantagens e desvantagens que oferece a proposta da Web Semântica, concluindo-se que a Web Semântica, ainda se desenvolvendo de forma lenta, é o futuro ao qual devemos chegar e, nesse sentido, o futuro das bibliotecas na Web Semântica apresenta-se bastante incerto, porém, terão desafios enormes e complexos por superar. Foi utilizada a análise documental como método de pesquisa.

Palavras-chave: Web Semântica. Bibliotecas. Tecnologias da Informação e da Comunicação.

Abstract: This paper presents an analysis on what are the initiatives taken by libraries to join the proposal of semantic web. Besides it presents briefly how Information and Communication Technologies have impacted the libraries, which are some of the main challenges in the process of adaptation to the new context of the semantic web, doing an analysis on how libraries could benefit the development of the Semantic Web and to achieve this purpose, we analyzed the evolution of the web in the last decades, taking into account the main characteristics, advantages and disadvantages that offers the proposal of the semantic web.. Also we showed up some actions based on examples developed by libraries until now, concluding that, although the semantic web will be developing slowly, the future of libraries in the Semantic Web has become quite uncertain, however, will have enormous and complex challenges to overcome. Documental analysis was used as a research method.

Keywords: Semantic Web. Libraries. Information and Communications Technologies.

1 INTRODUÇÃO

Nas últimas décadas do século XX, o desenvolvimento alcançado pela ciência no campo da informação proveniente dos avanços nas áreas da informática, eletrônica, telecomunicações, tecnologias de satélites dentre outros representou um enorme avanço para a sociedade atual, a qual passou de uma sociedade puramente industrial a uma sociedade baseada no frequente uso da informação e do conhecimento, e é por isso que na atualidade fala-se de uma “Sociedade da Informação e do Conhecimento”. Essa transformação converteu-se a uma realidade cada vez mais presente nos nossos dias com a aparição e a aplicação de termos como a Web e suas derivações, como Web 2.0 (Web social), Web

semântica entre outras e, hoje, tem levado a utilização das Tecnologias de Informação e Comunicação (TIC), sendo a internet o exponente mais fiel, a planos socioculturais.

Durante muito tempo, a Web serviu como um sistema de publicação rápida, em que, sem muitas complicações, podia-se disseminar qualquer tipo de informação, mesmo gráfica ou textual. Isto foi facilitado pela popularização de editores orientados à criação e à edição de textos em formato HTML, mas foram poucos os que se importaram com a necessidade de dar um significado semântico aos dados inseridos na Web.

Desde o começo do século XXI, vem se desenvolvendo por iniciativa do World Wide Web Consortium (W3C) um movimento que promoverá uma grande transformação no contexto da WWW, chamada de “Web Semântica”. Consistência-se numa extensão da Web atual, mas com uma construção que irá permitir uma melhor relação homem-máquina.

[...]com a popularização dos serviços e ferramentas da web 2.0. Se estabelecem padrões intuitivos de enlaçamento semântico seletivo por parte dos usuários finais, que aproveitam as funcionalidades disponíveis a seu alcance. Uma maior compreensão da arquitetura e funcionamento dos sistemas de gestão de conteúdos e da importância de armazenar a informação de forma estruturada em bases de dados traz uma sensibilidade notável para a importância de desenvolver esforços que façam possível etiquetar, enlaçar e reutilizar a informação automaticamente, em virtude de seu conteúdo semântico. (TRAMULLAS, 2011, p). (tradução própria)

No contexto bibliotecário, a entrada das TIC tem provocado o surgimento de novas formas de trabalho, novos serviços, como bibliotecas digitais, referência digital, consultas online com especialistas, digitalização dos catálogos, criação de repositórios de informação entre outros, tendo como alvo um novo tipo de usuário, segundo Orera-Orera (2007, p.330) “... um novo modelo de biblioteca para um novo modelo de sociedade”.

No entanto, algumas questões ainda persistem diante do acelerado desenvolvimento tecnológico da sociedade como, por exemplo, em quais práticas as bibliotecas vêm acertando para aproveitar as facilidades da Web semântica? Ou, com quais recursos elas contam? E como as bibliotecas não querem ficar excluídas, terão que respondê-las.

A seguir, faz-se uma análise sobre como as bibliotecas poderiam beneficiar o desenvolvimento da Web semântica, tendo como método de pesquisa a análise documental.

2 DA WEB A WEB SEMÂNTICA

Desde o surgimento da Word Wide Web (WWW) no final do século XX, sua evolução se subdividiu em três etapas: na primeira, como Web 1.0, a participação do usuário em sua construção era inexistente, pois todo o conteúdo era mantido por webmasters, tornando-a uma ferramenta apenas de leitura, com consultas de informação unidireccionais. Já, na segunda,

como Web 2.0, o usuário começa a ter maior participação na construção desta. As pessoas se conectam, colaboram, interagem e compartilham informações por meio do uso das redes sociais, blogs, wikis, cloud computing etc. E, por último, na terceira fase, que ainda está acontecendo, temos a Web semântica, em que se apresentam aplicações Web se conectando com outras aplicações Web, com o objetivo de enriquecer e ajudar na busca e na criação de informação, melhorando a experiência do usuário final.

Segundo Tim Berners-Lee, James Hendler y Ora Lassila (2001), a Web Semântica é uma extensão da Web atual, na qual a informação se posiciona com significado bem definido e estruturado, que permite sua compreensão pelos computadores, facilitando uma relação de cooperação entre humano e computador, em que a informação legível por máquina representa o pensamento das pessoas e o sistema informático faz a análise dos padrões de interação.

A Web Semântica pretende transformar os sistemas informáticos em agentes inteligentes, entidades que recolhem, filtram, processam, recuperam e fazem conclusões sobre a informação contida na Web para tomar decisões sem precisar da participação do usuário no processo. Tem como objetivo transformar a Web de hoje, de uma coleção de documentos, em uma base de conhecimento.

Permite solucionar problemas na busca de informação através do uso de uma infraestrutura que especifica o significado do conteúdo dos recursos, permitindo compartilhar, processar e trocar informação de forma automática. Para cumprir com estas funções, a Web semântica se sustenta em uma arquitetura organizada em camadas de tecnologias.

Este tipo de Web caracteriza um futuro não muito distante e que já está oferecendo alguns benefícios em áreas que tem aplicado algumas de suas ferramentas. A materialização deste fenômeno transformará significativamente a vida das pessoas. Todavia, a Web Semântica apresenta algumas vantagens, mas também apresenta desvantagens no seu desenvolvimento.

Algumas das vantagens que apresenta a Web Semântica são: fornece conteúdo semântico aos documentos colocados na internet, permitindo uma melhor organização da informação, o que garante uma busca por significado e não por conteúdo textual; os motores de busca oferecem resultados mais relevantes para as pesquisas dos usuários; maior simplicidade na hora de compartilhar, processar e transferir informação.

Entre as desvantagens da Web Semântica, podemos citar: custoso e complicado processo de adaptação e reestruturação dos documentos já existentes na internet para processá-los semanticamente; tecnicamente, existe um alto nível de complexidade na

codificação semântica, sendo preciso unificar padrões semânticos, além de ter que lidar com os problemas das diferenças de idiomas.

Apesar do certo equilíbrio que se apresentam entre vantagens e desvantagens da Web Semântica, suas tecnologias decerto tornarão mais acessíveis e compreensíveis às informações disponibilizadas nos ambientes informacionais digitais da Web.

3 AS BIBLIOTECAS E A WEB SEMÂNTICA

A Web Semântica propõe organizar e estruturar toda a informação disponível na Web, de tal forma que seja inteligível tanto pelos seres humanos quanto pelas máquinas, agregando metadados semânticos e fazendo uso de tecnologias que permitam o cruzamento dessas informações independentemente do seu suporte.

Em paralelo, uma das fontes de informação estruturada mais importante do mundo encontra-se nas bibliotecas e são as bases de dados que suportam os OPAC's(catálogo automatizado de acesso público on-line dos documentos de uma biblioteca, pelas siglas em inglês). Nesse sentido, a Biblioteconomia tem se encarregado há muito tempo de criar vocabulários, linguagens e formatos que permitam representar e organizar a informação acumulada nos seus fundos, facilitando os processos de busca e recuperação da informação nas unidades de informação:

- **Formato MARC** (Machine Readable Cataloging: Catalogação legível por computador)
- **Protocolo Z39.50**; protocolo cliente servidor de padrão internacional para o intercambio, transferência e recuperação de informação em computadores distribuídos em redes.
- **Esquema Dublin Core**, esquema de metadados que permite a descrição de objetos digitais na web.
- **OAI-PMH** (Open Archives Initiative Protocol for Metadata Harvesting: protocolo da iniciativa de arquivos abertos para a recollecção de metadados) protocolo desenvolvido pelo Open Archives Initiative que define um mecanismo para coleta de registros de metadados em repositórios entre outros.

Embora estas ferramentas tenham sido importantes para o desenvolvimento e melhora dos processos bibliotecários, elas foram projetadas para o uso exclusivo das comunidades bibliotecárias, porém, não para a reutilização dos dados resultantes da sua utilização, o que dificulta o processo de vinculação dos dados das bibliotecas com outros dados disponíveis na internet. Na atualidade os dados bibliotecários encontram-se fechados em grandes depósitos, com uma qualidade variável que não permite sua integração com outros conjuntos de dados, nem a vinculação à Web Semântica.

Um estudo do W3C Library Linked Data Incubator Group no ano 2011 publicou um informe expondo quais são os principais obstáculos que enfrentam as bibliotecas para aumentar a interoperabilidade global de seus dados na Web e se incluir no movimento da Web Semântica: os dados bibliotecários não estão integrados com recursos Web; a comunidade bibliotecária e a comunidade semântica usam diferentes terminologias para conceitos similares de metadados; os padrões bibliotecários se desenham só para a comunidade bibliotecária; as mudanças na tecnologia bibliotecária dependem do desenvolvimento dos vendedores de sistemas.

Neste documento o grupo também expõe quais deveriam ser as ações das bibliotecas para alcançarem a integração de seus conjuntos de dados na Web Semântica por meio do Linked Data¹³⁷ (dados ligados entre si): fazer uso dos dados bibliotecários disponíveis como Linked Data; usar Linked Data externos para enriquecer os serviços bibliotecários; identificar conjuntos de dados de valor, plausíveis de ser publicados dentro de projetos assumíveis e adequadamente dimensionados (nem tudo tem valor); mudar o enfoque da criação e desenvolvimento de padrões, levando em conta a participação de especialistas das comunidades científicas da web semântica; analisar desde uma perspectiva técnica, a criação, a gestão e a preservação de Identificadores de Recursos Uniforme (URIs, pela sigla em inglês), bem como o desenvolvimento e gestão de vocabulários em RDF¹³⁸ (Resource Description Framework).

Quando analisamos profundamente este documento, não encontramos nada novo e que já não saímos. O que preocupa é que as bibliotecas ainda não tenham assumido uma posição mais inovadora. Não é possível que as bibliotecas continuem sem desenvolver sistemas capazes de suprir as suas necessidades, ignorando as propostas e avanços tecnológicos alcançados em outros contextos, adotando uma posição inerte e sem iniciativas e esperando sempre adaptar modelos de externos a seu contexto.

Sendo assim, a adoção de um modelo que permita a interoperabilidade entre os dados fornecidos pelas bibliotecas e os disponíveis na Web em forma de linked data transita por transformar todo o legado da Biblioteconomia para uma nova linguagem de dados, mais acessíveis e reutilizáveis, com um uso mais promissor, para além do contexto bibliotecário: criar endereços Web usando URIs, o que ajudaria a eliminar redundâncias nos metadados,

¹³⁷ Conjunto de práticas iniciadas por Berners-Lee com função de publicar dados estruturados na Web.

¹³⁸ Linguagem utilizado para representar informação na Internet.

garantindo a padronização de termos utilizados por domínios de conhecimentos diferentes e que se referem a mesma coisa; desenvolver padrões de dados para as bibliotecas que sejam compatíveis e plausíveis de ser vinculados com dados de outros contextos onde já se esteja utilizando as tecnologias da Web Semântica, levando em conta que estas representam uma forma diferente de interpretar os dados e formatos de dados do século passado; o conhecimento dos bibliotecários no gerenciamento de metadados os tornaria importantíssimos parceiros nos grupos de desenvolvimento de padrões da Web Semântica; a grande experiência no gerenciamento de arquivos de autoridade e listas controladas iria permitir que as bibliotecas expandissem rapidamente sua presença na nuvem de dados ligados; a experiência bibliotecária na preservação digital se converte em fator favorável para a necessária gestão de vocabulários.

Na medida em que as bibliotecas compreendam a importância e a necessidade da vinculação de seus dados com outros domínios no contexto da Web semântica e o entendimento comum dessa necessidade se expanda, os benefícios da publicação de dados ligados irão ficando mais claro, assim demonstrado pelos resultados obtidos por aquelas bibliotecas que aderiram e colocaram em prática as tecnologias da Web Semântica:

- a) O projeto Europeana, que disponibilizou 3,5 milhões de registros de objetos disponíveis como dados abertos ligados.
- b) A Biblioteca do Congresso (LC) que ofereceu seu File Name Authority como dados abertos ligados, reconhecendo a necessidade da adoção de um modelo mais robusto de intercâmbio de dados baseado no uso do RDF.
- c) Yahoo, Google e Bing lançam o schema.org que usa dados estruturados na otimização de motores de busca;
- d) A Biblioteca da Universidade de Cambridge publicou um dataset de 1,3 milhões de registro, e fornece um kit de ferramentas para converter registros de MARC21 para linked data.
- e) O Virtual International Authority File (VIAF), que acolhe os registros de autoridade de mais de uma dúzia de agências nacionais e regionais e tem ligado e disponibilizado muitos de seus registros para DBpedia.

Todas estas instituições obtiveram grandes vantagens na disponibilização de seus dados, mas, se a aplicação destas tecnologias guiou essas bibliotecas a um melhor aproveitamento dos dados por elas armazenados, existe ainda um problema a ser enfrentado, e diz respeito à disponibilização dos dados abertos interligados na internet, conhecido como “Linked Open Data”, pois, presume-se, que se os dados estão disponíveis na internet e podem

ser descarregados, eles são dados abertos e livres. Mas esta afirmação não é verdadeira, pois, para que um dado seja aberto, além de ser preciso que ele esteja disponível na internet e possa ser descarregado, ainda é necessário:

[...] possuir algum tipo de licença legal para poder utilizar, reutilizar y redistribuir, mesclando-os incluso com outros dados, sujeitos como mínimo á “atribuição” (reconhecimento de autoria, quem o tem feito), ao “compartilhar igual” (que a exploração que se faça destes dados –incluindo as obras derivadas- mantenham a mesma licença ao serem divulgadas). (HERNÁNDEZ-PÉREZ; GARCÍA-MORENO, 2013, p.2). (tradução própria)

Quando nos referimos a dados interligados, estamos nos referindo à possibilidade de acessar e navegar por diferentes fontes de dados seguindo os links RDF. Sejam os dados científicos, geográficos, históricos, governamentais ou de administrações públicas, o mais importante é que possam ser considerados como abertos e estejam interconectados. Neste sentido, Tim Berners-lee (2009) ofereceu um esquema, em que ele atribui uma qualificação para os diferentes níveis de disponibilização de dados abertos: Uma estrela, por disponibilizar dados na Web em qualquer formato, com uma licença aberta; Duas estrelas, quando esses dados são disponibilizados em um formato estruturado; Três estrelas, quando é utilizado um formato não proprietário na disponibilização dos dados; Quatro estrelas, quando os dados utilizam URIs; Cinco estrelas, quando eles estão entrelaçados a outros dados, oferecendo um contexto.

Um dos principais benefícios para o mundo bibliotecário, apontados pelo W3C Library Linked Data Incubator Group (W3C, 2011), referente ao uso de Linked Open Data, é a criação de “uma grande piscina global de dados abertos, compartilhados que podem ser usados e reusados para descrever os recursos, com uma quantidade limitada de esforço redundante em comparação com os processos de catalogação atuais”.

É certo que existem notáveis projetos e serviços bibliotecários transitando pelas tecnologias da Web Semântica, mas os passos dados até agora se mostraram insuficientes. Sendo assim, as bibliotecas precisam de uma mudança radical no seu funcionamento para alcançar o êxito, o que representaria quase abandonar os sistemas de compartilhamento e criação de dados fortemente arraigados em seus sistemas e recomeçar com novas regras e procedimentos, fato que representaria uma enorme perturbação para uma profissão que é muito lenta para evoluir.

Em um contexto digital que avança cada vez mais rápido no desenvolvimento e aproveitamento das tecnologias da Web Semântica, as bibliotecas devem considerar alguns elementos expostos por López e Zorita (2008) e que são considerados essenciais para uma

mudança em nossas bibliotecas: considerar a biblioteca como uma espécie de “contêiner” capaz de estruturar seus dados e oferecer relações semânticas do tipo “este objeto é parte de” ou “este objeto é membro de”; aumentar a interoperabilidade dos sistemas bibliotecários, permitindo o intercâmbio de informação com outros sistemas e a possibilidade de nossos objetos serem utilizados e reutilizados em outras aplicações de forma transparente; fazer dos sistemas bibliotecários sistemas mais flexíveis, podendo incorporar outros esquemas de metadados e taxonomias, mais cômodas para as necessidades dos pesquisadores e seus materiais altamente especializados; construção de sistemas com arquiteturas mais abertas, criando aplicações modulares, possibilitando maior flexibilidade no momento de criar serviços mais complexos; a criação de repositórios de dados, além dos repositórios institucionais já existentes nas bibliotecas, capaz de integrar-se entre eles e com os outros sistemas das instituições onde estão instalados.

5 CONSIDERAÇÕES FINAIS

Para as bibliotecas alcançarem os propósitos da Web Semântica, faz-se necessário uma padronização de suas tecnologias, linguagens e metadados descritivos, de forma que respondam a determinados princípios comuns, como armazenar e descrever dados de maneira que sejam utilizados de forma natural por pessoas e máquinas.

O futuro das bibliotecas na realidade da Web Semântica apresenta-se bastante incerto, com desafios enormes e complexos por superar, mas também com oportunidades para aproveitar.

Continuaríamos nos enganando se perdurasse a inércia imperante em nossas práticas e história, desperdiçando as oportunidades que o novo contexto tecnológico coloca em nossas mãos, correndo o risco ainda de sermos considerados ultrapassados e com pouca relevância no contexto tecnológico informacional.

REFERÊNCIAS

- BERNERS-LEE, T.; HENDLER, J.; LASSILA, O. The Semantic Web: A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities. Scientific American Special, número especial, 2001. Disponível em: <<http://www.cs.umd.edu/~golbeck/LBSC690/SemanticWeb.html>>. Acesso em: 25 jun. 2014
- BERNERS-LEE, Tim. Linked data - Design issues. 2009, Disponível em: <<http://www.w3.org/DesignIssues/LinkedData.html>>, Acesso em: Acesso em: 15 jul. 2014.
- HERNÁNDEZ-PÉREZ, T.; GARCÍA-MORENO, M. A. Datos abiertos y repositorios de datos: nuevo reto para los bibliotecarios. El profesional de la información, mayo-junio, v. 22, n. 3, p. 259-263, 2013.

LÓPEZ, A.; ZORITA, L. Les biblioteques universitàries i la gestió de la informació en l'entorn digital: diverses consideracions per a dipòsits digitals. BiD: textos universitaris de biblioteconomia i documentació, v.20, n. junho, 2008. Disponible em: <<http://bid.ub.edu/20lopez1.htm>> Acesso em: 4 jul. 2014.

ORERA-ORERA, L. La biblioteca universitaria ante el nuevo modelo social y educativo. El profesional de la información, v.16, n.4, p. 329-337, 2007.

TRAMULLAS, Jesús. Web semántico en bibliotecas: del material del que están hechos los sueños. 2011. Disponible em: <<http://www.ub.edu/blokdebid/es/content/web-sem%C3%A1ntico-en-bibliotecas-del-material-del-que-est%C3%A1n-hechos-los-sue%C3%B3os>>. Acesso em: 15 jun. 2014

W3C Library Linked Data Incubator Group. Draft report with transclusion. 2011. Disponible em: <<http://www.w3.org/2005/Incubator/lld/wiki/DraftReportWithTransclusion>>. Acesso em: 15 jun. 2014.

ANÁLISE DA ARQUITETURA DA INFORMAÇÃO DO WEBSITE DO SIBI UNIRIO

ANALYSIS OF THE ARCHITECTURE OF INFORMATION THE WEBSITE OF SIBI UNIRIO

Alanda do Valle Vitorino
Claudio Jose Silva Ribeiro

Resumo: Com o aumento da utilização da *web*, os *websites*, se tornaram grandes ferramentas de relacionamento das bibliotecas com seus usuários. A Arquitetura da Informação surge neste cenário com uma alternativa metodológica para organizar a informação em ambiente *web*. Dentro deste contexto, conforme os preceitos dos autores Rosenfeld e Morville (2006) que apresentam uma metodologia para projetos de Arquitetura da Informação além de quatro sistemas para Arquitetura da Informação, sistema de organização, sistema de rotulagem, sistema de navegação e sistema de busca, propõe-se nessa pesquisa analisar o *website* do Sistema de Bibliotecas da Universidade Federal do Estado do Rio de Janeiro-UNIRIO sob duas perspectivas, do processo de concepção e do produto final (o *website*). Aplica a técnica de entrevistas com roteiro para identificação do processo de elaboração do *website* e adota os quatro sistemas da Arquitetura da Informação para observar o produto final. Pretende identificar aspectos positivos e negativos do *website* e apresentar uma proposta utilizando a Arquitetura da Informação para aprimoramento do *website*.

Palavras-chave: *Websites* de bibliotecas. Arquitetura da Informação. Sistema de Bibliotecas da Universidade Federal do Estado do Rio de Janeiro.

Abstract: With the increased use of the web, websites, became great tools of libraries relationship with their users. The Information Architecture arises in this scenario with an alternative methodology to organize the information in a web environment. Within this context, according to the precepts of the authors Rosenfeld and Morville (2006) who present a methodology for design of information architecture and four systems for information architecture, system organization, labeling system, navigation and search system, is proposed in this research to analyze the website of the Library System of the Federal University of the State of Rio de Janeiro-UNIRIO two perspectives, the design process and the final product (the website). Apply the technique of interviewing roadmap for identifying the process of developing the website and takes the four systems of the Information Architecture to watch the final product. Aims to identify positive and negative aspects of the website and submit a proposal using the information architecture to improve the website.

Keywords: *Websites* libraries. Information Architecture. Library System of the Federal University of the State of Rio de Janeiro.

1 INTRODUÇÃO

A popularização da *Internet* na década de 1990 e a evolução das Tecnologias da Informação e Comunicação (TICs) desencadearam mudanças na maneira como a sociedade vem produzindo e buscando informação. A enorme quantidade de conteúdos e formatos disponibilizados na *web* transformou o ambiente informacional.

A tarefa de buscar e localizar informações tornou-se cada vez mais complicada. Rosenfeld e Morville (2006) citam que a incapacidade de encontrar uma informação é um dos fatores que mais frustram os usuários.

Nesse contexto, Wurman (2001 *apud* SILVA; DIAS, 2008, p.4) afirma que Arquitetura da Informação “organiza a informação para torná-la clara, [...] permitindo a criação no ambiente *web*, de um espaço informacional por onde o usuário é capaz de navegar/mover-se de forma mais eficiente às suas necessidades de informação”.

Analizando a literatura dos últimos seis anos sobre avaliação de Arquitetura da Informação de *websites* nas áreas de Biblioteconomia e Ciência da Informação fica evidente na literatura a unanimidade na preferência pelos estudos de Rosenfeld e Morville (2006) como metodologia aplicada aos estudos, podemos citar (AQUINO e OLIVEIRA, 2012; SILVA *et al.*, 2012; CICON, 2013; DANTAS e SILVA, 2013; MIRANDA, *et al.*, 2013; SANTOS e CAMPOS, 2013), razão pela qual esse estudo também optou em utilizá-los em sua análise.

Os *websites* são grandes ferramentas para as bibliotecas universitárias disponibilizarem e disseminarem informações. Brinkley (1999, *apud* AMARAL, 2005, p. 19) assegura que o “*website* é o melhor meio para promover e publicar os serviços oferecidos pelas bibliotecas”.

Sendo assim, conforme os preceitos dos autores Rosenfeld e Morville (2006) que apresentam uma metodologia para projetos de Arquitetura da Informação além de quatro sistemas para (AI), sistema de organização, sistema de rotulagem, sistema de navegação e sistema de busca, propõe-se nessa pesquisa analisar o *website* do Sistema de Bibliotecas da Universidade Federal do Estado do Rio de Janeiro-UNIRIO sob duas perspectivas, do processo de concepção e do produto final (o *website*).

Considerando que falhas na organização de um *website* provocam em seus usuários confusão e decepção dificultando o uso do mesmo e repercutindo negativamente na imagem da instituição, este estudo faz-se necessário pois visa contribuir para a melhoria da qualidade do *website* do Sistema de Bibliotecas da UNIRIO, a fim de que os usuários possam atingir seus objetivos e satisfazer suas necessidades informacionais.

2 ARQUITETURA DA INFORMAÇÃO

Em 1976, Richard Wurman, desenhista e Arquiteto por formação, cunhou o termo Arquitetura da Informação (AI), segundo o qual definiu a AI como “sendo a ciência e a arte de criar instruções para espaços organizados. Aplicando esse conceito na organização de informações em suportes físicos, como guias e mapas. (SILVA *et al.*, 2011).

Na década de 1994, Luís Rosenfeld e Joseph Janes, ambos com formação em Ciência da Informação e Biblioteconomia, fundaram a Argus Associates, empresa dedicada a trabalhar com Arquitetura da Informação na *web*. Nesse mesmo ano Rosenfeld publicou uma coluna

intitulada “Arquiteto da web” na revista *Web Review*, unindo-se subsequente com Peter Morville, também bibliotecário, sendo contratado como primeiro empregado da Argus Associates (CAMARGO; VIDOTTI, 2011).

Juntos Rosenfeld e Morville foram chamados em 1998 pela *O'Reilly Publishing* para escrever um livro, intitulado “*Information Architecture for Word Wide Web*” com última edição em 2006.

Rosenfeld e Morville (2006) apresentam quatro definições para arquitetura da informação:

- O *design* estrutural de ambientes de informação compartilhados.
- A combinação de sistemas de organização, rotulagem, navegação e busca em *websites* e *intranets*.
- A arte e a ciência de dar forma a produtos de informação e experiências para apoiar a usabilidade e a *findability*.
- Uma disciplina emergente e comunidade de prática focada em trazer princípios do *design* e da arquitetura para a paisagem digital.

Os autores acima citados defendem que a Arquitetura da Informação consiste em compreender três pilares: contexto, conteúdo e usuário. Esta tríade é única para cada *website* e cabe ao arquiteto da informação balanceá-las.

- ✓ **Contexto** – Objetivos do *website*, cultura e política da empresa, ambiente de uso, restrições tecnológicas, recursos humanos, etc.
- ✓ **Conteúdo** – Volume, formato, estrutura, propriedade, dinamismo, metadados etc.
- ✓ **Usuário** – Necessidades, comportamento de busca da informação, experiências, tarefas, etc.

2.1 Processo e Metodologia de Arquitetura da Informação

O projeto de Arquitetura de Informação de um *website* envolve dificuldades no *design* dos quatro sistemas da Arquitetura da Informação. Vencer essas dificuldades torna o projeto de AI um problema complexo, sendo necessário uma metodologia para organizar o trabalho do arquiteto e garantir a qualidade do produto final (REIS, 2007).

Para Rosenfeld e Morville (2006) a criação de *websites* requer uma abordagem por fases. Então os autores criaram uma metodologia que compreende cinco fases: pesquisa, estratégia, *design*, implementação e administração. No entanto em seu livro *Information Architecture for the word wide web* eles focam nas três primeiras fases, pesquisa, estratégia e

design, por considerarem as fases de implementação e administração a execução das etapas anteriores.

A fase da **pesquisa** começa com a revisão dos materiais de apoio existentes, reuniões com a equipe de estratégia visando a obtenção de entendimento sobre as metas, contexto, Arquitetura da Informação existente, o conteúdo e o usuário. A pesquisa fornece uma compreensão contextual que constitui a base para o desenvolvimento de uma estratégia de arquitetura da informação.

De uma perspectiva *top-down*, a **estratégia** define dois ou três níveis de estruturas de organização e navegação do *website*. Da perspectiva *bottom-up* sugere tipos de documentos e um esquema de metadados. A estratégia oferece um *framework* de primeiro nível para a AI estabelecendo a direção e o alcance que vai orientar o projeto de Arquitetura da Informação até a implementação.

Na fase do **design** é aonde dá-se forma a estratégia de uma Arquitetura da Informação, é a criação de *blueprints* detalhados, *wireframes* e esquemas de metadados.

A **implementação** ocorre quando o projeto é colocado a prova, o *website* é construído, testado e lançado.

Na **administração** ocorre a avaliação e melhoria continua da Arquitetura da Informação. Tarefas diárias como inclusão e exclusão de novos documentos, monitoramento do uso e *feedback* ao usuário.

2.2 Sistemas da Arquitetura da Informação

Rosenfeld e Morville (2006) ainda dividem a Arquitetura da Informação em quatro sistemas.

- **Sistema de Organização (Organization System):** Define a classificação de todo o conteúdo.
- **Sistema de Rotulação (Labeling System):** Estabelece as formas de representação, de apresentação, da informação definindo rótulos para cada elemento informativo.
- **Sistema de Navegação (Navegation System):** Especifica as maneiras de navegar, de se mover pelo espaço informacional e hipertextual.
- **Sistema de Busca (Search System):** Determina as perguntas que o usuário pode fazer e o conjunto de respostas que irá obter.

Albuquerque e Lima-Marques (2011, p. 64) analisam que “a proposição de quatro sistemas interdependentes para a arquitetura da informação de *website* (Organização,

navegação, rotulação e busca), feita por estes autores, torna-se, até um certo ponto, um desdobramento dos primeiros princípios (contexto, conteúdo e usuários)”. Reis salienta que

A divisão da arquitetura de informação nesses quatro sistemas é apenas conceitual e destina-se a organizar o trabalho do arquiteto de informação. Todos esses sistemas apresentam uma grande interdependência de modo que os problemas de um deles normalmente afetam os demais. Apesar disso, analisá-los separadamente facilita a busca das dificuldades de projetá-los porque cada um deles é conceituado com bases teóricas advindas de disciplinas diferentes (REIS, 2007, p.72).

3 METODOLOGIA

Tendo em vista os objetivos propostos, a investigação se caracteriza como uma pesquisa descritiva de abordagem qualitativa.

Quanto ao procedimento, é possível classificar essa pesquisa em bibliográfica, documental e estudo de campo. A pesquisa bibliográfica para a fundamentação teórica, a documental para a coleta e análise dos dados e pesquisa de campo para conhecer a metodologia de desenvolvimento do projeto de Arquitetura da Informação.

Neste estudo, a pesquisa bibliográfica constituiu-se na análise de textos impressos e eletrônicos obtidos em bibliotecas, Portal da CAPES, na base BRAPCI, repositórios digitais das universidades e anais de eventos da área de Biblioteconomia e Ciência da Informação. As palavras chaves utilizadas foram: Arquitetura da Informação; website de biblioteca; biblioteca universitária.

Para o estudo de campo será utilizada a técnica de entrevista da pesquisa qualitativa de entrevistas em profundidade seguindo um roteiro, elaborado a luz da revisão de literatura, com o objetivo de identificar a metodologia adotada no processo de desenvolvimento do *website*.

Para a pesquisa documental, será utilizada a *homepage* do *website* do Sistema de Bibliotecas da UNIRIO (<http://www.biblioteca.unirio.br>), onde será verificado a Arquitetura da Informação com base nos sistemas de Rosenfeld e Morville (2006).

4 CONCLUSÕES PARCIAIS

As bibliotecas são fontes de informação confiáveis para a sociedade, devem ser capazes de utilizar a *web* como forma de disseminar informações e oferecer serviços de modo a não ficarem excluídas desse novo cenário, o *website* deve representar a biblioteca na *internet* e se tornar referencial de pesquisa para seus usuários.

Diante da possibilidade que a *web* oferece de armazenar grandes quantidades de informação de diferentes fontes em diversos formatos e suportes suscita-se a seguinte questão

“como as bibliotecas podem vencer o desafio de apresentar esse vasto conteúdo em seus *websites* de forma que seus usuários não se sintam perdidos e confusos?”. Nesse sentido, os preceitos da Arquitetura da Informação podem auxiliar as bibliotecas na estruturação e organização da informação de seus *websites*. Seguir uma metodologia de AI no processo de desenvolvimento do *website* e aplicar esses preceitos no produto final é fundamental para as bibliotecas vencerem esse desafio. A organização de ambientes de informação é mais eficiente quando se segue um método.

REFERÊNCIAS

- ALBUQUERQUE, A. R. R.; LIMA-MARQUES, M. Sobre os fundamentos da Arquitetura da Informação. **Perspectivas em Gestão & Conhecimento**, João Pessoa, v. 1, Número Especial, p. 60-72, out. 2011.
- AMARAL, S. A. Websites: uso de tecnologias no cumprimento das funções da biblioteca. **Informação & Sociedade: Estudos**, João Pessoa, v. 1, n. 1, p. 15-40, jul./dez. 2005.
- AQUINO, M. A.; OLIVEIRA, H. P. C. Contribuições da arquitetura da informação para o website "A Cor da Cultura". **Informação & Sociedade: Estudos**, João Pessoa, v. 22, n. 1, p. 129-143, jan./abr. 2012. Disponível em: <<http://www.ies.ufpb.br/ojs2/index.php/ies/article/view/12158>>. Acesso em: 26 nov. 2013.
- BRINKLEY, M. The library web site in 1999: a virtual trip to the library. In: INTERNET & LIBTECH INTERNATIONAL, 99., 1999. **Anais...** London, 1999: Information Today, 1999. Disponível em:<<http://www.esatclear.ie/~mbrinkley/ili99.html>>. Acesso em:16 set. 2014.
- CAMARGO, L. S. A.; VIDOTTI, S. A. B. G. **Arquitetura da Informação**: uma abordagem prática para o tratamento de conteúdo de Interface em ambientes informacionais digitais. Rio de Janeiro: LTC, 2011.
- CICON, C. R. **A organização da informação em ambiente Web**: um estudo do portal do Procon-PR. 2012. 150f. Dissertação (Mestrado Profissional em Gestão da Informação) – Centro de Educação, Comunicação e Artes, Universidade Estadual de Londrina. Disponível em:<<http://www.bibliotecadigital.uel.br/document/?code=vtls000 181054>>. Acesso em: 26 nov. 2013.
- DANTAS, C. M.; SILVA, H. O. Arquitetura da Informação, acessibilidade e usabilidade: princípios básicos para a análise do website da Fundação Centro Integrado de Apoio à Pessoa Portadora de Deficiência(FUNAD). In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 14., 2013, Florianópolis. **Anais...** Florianópolis: ENANCIB, 2013. Disponível em: <<http://enancib.sites.ufsc.br/index.php/enancib2013/XIVenancib/paper/view/334>>. Acesso em: 26 nov. 2013.
- MIRANDA, Z. D. *et al.* Análise do ambiente do SIS Médicos e a Cultura a partir da Arquitetura da Informação: enfocando o sistema de organização. **Tendências da Pesquisa Brasileira em Ciência da Informação**, João Pessoa, v. 5, n. 1, 2013. Disponível em: <<http://www.brapci.ufpr.br/documento.php?dd0=0000013075&dd1=395f7>>. Acesso em: 26 nov. 2013.

REIS, G. **Centrando a arquitetura da informação no usuário.** Dissertação (Mestrado em Ciência da Informação) –Escola de Comunicação e Artes, Universidade de São Paulo, São Paulo, 2007. Disponível em: <<http://www.teses.usp.br/teses/disseveis/27/27151/tde-23042007-141926/pt-br.php>>. Acesso em: 30 mar. 2014.

ROSENFELD, L.; MORVILLE, P. **Information Architecture for the Word Wide Web.** 3. ed. Sebastopol, CA: O'Reilly, 2006.

SANTOS, S. de O.; CAMPOS, M. L. A. Análise do Portal IBGE a luz da arquitetura da informação de da classificação facetada. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 14., 2013, Florianópolis. **Anais...** Florianópolis: ENANCIB, 2013. Disponível em: <<http://enancib.sites.ufsc.br/index.php/enancib2013/XIVenancib/paper/view/244>>. Acesso em: 26 nov. 2013.

SILVA, M. A.T. *et al.* O que é arquitetura da informação? **Biblionline**, João Pessoa, v. 7, n. 1, p. 47-57, 2011.

SILVA, M. A. T. *et al.* Arquitetura da informação aplicada a leitores de E-book: avaliando o sistema de organização da interface do Kindle III wifi. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 13., 2012, Rio de Janeiro. **Anais....** Rio de Janeiro: ENANCIB, 2012. Disponível em: <<http://www.eventosecongressos.com.br/metodo/enancib2012/arearestrita/pdfs/19214.pdf>>. Acesso em: 26 nov. 2013.

SILVA, P. M.; DIAS, G. A. A arquitetura da informação centrada no usuário: estudo do website da biblioteca virtual em saúde (BVS). **Revista Eletrônica de Biblioteconomia e Ciência da Informação**, Florianópolis, n. 26, 2ºsem.2008. Disponível em: <<http://www.periodicos.ufsc.br/index.php/eb/article/viewFile/7200/6647>>. Acesso em: 23 dez. 2013.

WURMAN, R. S. **Ansiedade de Informação:** como transformar informação em compreensão. São Paulo: Cultura, 2001.

AVALIANDO A EDITORAÇÃO DE E-BOOKS EM AMBIENTES DE EDITORAS UNIVERSITÁRIAS: UMA APLICAÇÃO DO OPEN MONOGRAPH PRESS

*EVALUATING THE E-BOOKS PUBLISHING IN UNIVERSITY PRESSES ENVIRONMENTS:
AN APPLICATION OF THE OPEN MONOGRAPH PRESS*

Adriana Carla Silva de Oliveira

Guilherme Ataíde Dias

Resumo: O livro acompanhou o desenvolvimento da tecnologia se adaptando à mesma, assim como às necessidades de alguns leitores. Isso contribuiu para a emergência do e-book, expandindo as funcionalidades do livro impresso, sendo, uma tecnologia em franca evolução no cenário corporativo, e acadêmico, onde editoras de universidades federais brasileiras já os editoram e os disponibilizam. Esta é uma pesquisa bibliográfica e exploratória, que objetiva identificar, por meio de questionário, os formatos de *e-books* utilizados, bem como, a estrutura de editoração de *e-books* na ambiência de editoras universitárias brasileiras. Os resultados da pesquisa servirão de subsídio para a implantação da prática de editoração de *e-books* na Editora da Universidade Federal da Paraíba, por meio da utilização de um *software* denominado *Open Monograph Press*. Identificou-se, que os formatos de *e-books* mais utilizados pelas editoras pesquisadas foram o *Portable Document Format (PDF)* e o *Electronic Publication (EPUB)*, e que a editoração de *e-books* no cenário acadêmico, está em franca evolução. Foi identificado que 44,44% das editoras universitárias brasileiras, editoram *e-books*.

Palavras-chave: *E-books*. Editoras universitárias. *Open Monograph Press*.

Abstract: The book followed the development of technology adapting to it, as well to the needs of some readers. This contributed to the emergence of e-book, expanding the printed books features. The e-books technology is developing in the corporate and academic environments. Brazilian federal universities presses already edit them and make them available. This is a literature and exploratory research that aims to identify through a questionnaire, the formats of e-books used, as well as the structure of publishing e-books on the ambience of Brazilian university presses. The survey results will help the implementation of the e-books publishing practices at Universidade Federal da Paraíba press, by using a software product known as Open Monograph Press. It was identified that the formats of e-books used by most publishers surveyed were the Portable Document Format (PDF) and the Electronic Publication (EPUB). The publishing of e-books in academic environments is in steady evolution. It was identified that 44.44% of the Brazilian university presses publish e-books.

Keywords: *E-books*. Scholarly presses. *Open Monograph Press*.

1 INTRODUÇÃO

O livro, desde a sua gênese, acompanha o desenvolvimento das tecnologias, e vem passando por inúmeras modificações, no que diz respeito aos seus suportes informacionais, produto de influências culturais, econômicas e sociais.

Como seu primeiro suporte informacional, o livro teve a pedra e/ou a argila, materializando-se em objeto de grande volume e, portanto, de difícil manuseio, deslocamento, e armazenamento. Posteriormente, um novo material começa a ser usado, o papiro. Contrário

à pedra, possuía fácil manuseio, deslocamento e armazenamento, originando os cofres/depósitos de livros, denominados *bibliothéke*. Do papiro, o livro evoluiu para o pergaminho, que além daquelas características, conservavam por mais tempo as informações contidas neles. Em seguida, vem o modelo *códex*, que eram pergaminhos costurados, onde o livro já se figurava no formato de livro impresso, como é conhecido atualmente. Por volta do século XII, o pergaminho passa a ser substituído pelo papel branco, que passa a compor o livro. Mais tarde, em 1455, Johannes Gutenberg inventa os tipos móveis, propiciando a sua popularização e acesso para a sociedade.

O livro acompanhou o desenvolvimento da tecnologia se adaptando à mesma, como também se adapta às necessidades de alguns leitores. “Assim como a sociedade evolui, não é diferente com os suportes informacionais. Eles evoluíram junto com a sociedade de modo a se adequar às necessidades do espaço e do tempo”. (ARAÚJO et al, 2013, p. 14).

Surgiram os *e-books*, que se apresentam como inovações do tradicional livro impresso, possibilitando acesso rápido e fácil a uma quantidade infinita de obras, virtualmente. Possibilita expandir as funcionalidades do livro impresso, do livro tradicional. Assim, podemos compreender um *e-book* como um dispositivo de Tecnologia da Informação que disponibiliza conteúdos informacionais, de forma a mimetizar o livro tradicional (DIAS, 2010). É uma tecnologia emergente com abrangência nacional e internacional, que está tendo bastante destaque, tanto no cenário comercial, como no cenário acadêmico.

Como evidência, em âmbito comercial, temos empresas multinacionais, como a *Amazon.com* e a *Apple* que, além de comercializarem diversas obras nesta modalidade de livro, também desenvolveram e comercializam dispositivos eletrônicos de leitura para os mesmos, os *e-readers*, como é o caso do *Kindle* e do *iPad*, respectivamente. No Brasil, as vendas de *e-books* estão em crescimento, algumas empresas como a Livraria Cultura já comercializam obras e, também, o dispositivo eletrônico para leitura de *e-books*, chamado neste caso particular de *Kobo*.

Em âmbito acadêmico, os *e-books* se tornaram realidade. As universidades federais, a partir das suas respectivas bibliotecas disponibilizam para a comunidade, conteúdos editorados por editoras comerciais. De forma equivalente, algumas universidades, através de suas editoras já os editoram e os disponibilizam em variados formatos para a comunidade.

A pesquisa objetiva avaliar, por meio de pesquisa bibliográfica e exploratória, a editoração de *e-books* em ambientes de editoras universitárias brasileiras, bem como os padrões de *e-books* utilizados por elas, objetivando a implantação de um *software* de código aberto, livre, o *Open Monograph Press* (OMP), que automatiza o fluxo editorial para a

editoração de *e-books* e, posteriormente, seu acesso e/ou venda, na ambiência da Editora da Universidade Federal da Paraíba (EDUFPB).

A partir destas explanações iniciais, indagamos: **como se configura a editoração de e-books em Editoras Universitárias Brasileiras no que tange aos formatos para a sua disseminação e aos sistemas utilizados para a automação do processo editorial?**

2 OPEN MONOGRAPH PRESS

O *OMP* é um *software* de código aberto da plataforma *Public Knowledge Project* (PKP), projetado para auxiliar as editoras universitárias e congêneres interessadas em publicar *e-books* em diferentes formatos eletrônicos, seja em acesso aberto ou através de compra (PKP, 2010). Sua introdução na ambiência de editoras busca simplificar o processo de editoração e publicação de *e-books*.

O *OMP* visa gerenciar e agilizar o fluxo editorial da publicação eletrônica de *e-books*, automatizando todas as etapas editoriais de publicação, além de disponibilizar ferramentas para criação de catálogos *online* de acesso aberto, o gerenciamento de depósito, a avaliação e disponibilização das publicações. Os critérios e políticas de publicação e disponibilização do material são de responsabilidade da editora, muito embora o *OMP* possua um fluxo editorial próprio, conforme descrito no item a seguir.

2.1 Processo de editoração do *OMP*

O processo de editoração no *OMP* divide-se em diversas etapas. Inicia-se pelo processo de submissão realizado pelo autor, onde o mesmo irá preencher, no ato da submissão, as informações do material. O próprio autor realiza a catalogação do seu material.

Em seguida, o material submetido é enviado eletronicamente ao editor, que avalia se as políticas e critérios estabelecidos pela editora foram atendidos. Em caso positivo, o material é encaminhado aos pareceristas (revisores interno e externo), que avaliam o mesmo e atribuem o parecer final: aprovado ou recusado. Em caso negativo, o editor pode entrar em contato com o autor, solicitando que o mesmo faça as alterações adequadas, da mesma forma que os pareceristas podem solicitar alterações no material enquanto o avaliam.

Em sendo aprovado, o material é encaminhado para a fase de edição, produção e elaboração final da obra, sob as funções dos Produtores e *Designers*. O fluxo editorial finaliza quando o material, o *e-book*, é publicado e inserido no catálogo da editora, obedecendo aos critérios de disponibilização, sob a função, novamente, do Editor. A Figura 1 ilustra o processo de editoração do *OMP*.

FIGURA 1 – Processo Editorial do OMP



Fonte: Dados da pesquisa (2014).

É interessante ressaltar que, durante todo este processo de submissão até a publicação final, o autor acompanha *online* todo o fluxo editorial do seu material.

A utilização do *OMP* contribui para aumentar a eficácia e eficiência em todo o gerenciamento do fluxo editorial de *e-books* em editoras universitárias. A sua aplicação em editoras universitárias está em crescimento no âmbito nacional.

No cenário da EDUFPB, a adoção desse *software* almeja um gerenciamento eficiente do processo editorial dos *e-books* com disponibilidade de acesso aberto para toda a comunidade acadêmica.

3 METODOLOGIA

Trata-se de uma pesquisa bibliográfica e exploratória, de natureza quanti-qualitativa.

Pesquisa bibliográfica, a partir de um levantamento sobre os padrões de *e-books* utilizados em editoras de universidades federais brasileiras, bem como, um levantamento sobre a adoção do *OMP*, e a sua utilização, que ainda é pequena nacionalmente.

A pesquisa de caráter exploratório visa coletar informações sobre a temática estudada, com vistas à ampliação do entendimento acerca do fenômeno, já que as informações acerca desse assunto ainda são poucas. Indicada em situações nas quais as informações sobre o problema em estudo são insignificantes ou insuficientes, a pesquisa exploratória tem por finalidade ampliar o conhecimento sobre o tema pesquisado, esclarecendo conceitos e fornecendo subsídios para as etapas subsequentes da investigação (VERGARA, 2006).

A pesquisa assume um caráter quantitativo e qualitativo no que diz respeito às informações coletadas através da aplicação do questionário que foi realizado pelo autor. Assume um caráter quantitativo, acerca da representação em forma numérica dos dados obtidos através da pesquisa, e qualitativo para interpretação e análise dos dados coletados.

3.1 Procedimento metodológicos

O universo da pesquisa teve como escopo as editoras das universidades federais brasileiras. Inicialmente foi realizado um levantamento, visando elaborar uma lista de todas as universidades federais existentes no Brasil, com a finalidade de saber quais dessas possuem

editoras universitárias. Esse universo totalizou em 63 (sessenta e três) universidades federais brasileiras.

Para identificar quais universidades tinham editoras universitárias, visitamos os *sites* oficiais das universidades, à procura de suas editoras, e seus *e-mails* para contato, para posteriormente realizarmos contato e enviarmos o questionário. Algumas universidades informam a existência de editora universitária no próprio *site*; outras não, sendo necessário realizar o contato com as mesmas, via *e-mail*, para colher esta informação. Em alguns casos, não foi possível obter retorno.

A pesquisa teve como objetivo geral avaliar os principais formatos de *e-books* utilizados em editoras universitárias, como parte do processo de implantação do *OMP* na EDUFPB.

A pesquisa dividiu-se em cinco partes: 1) Levantamento bibliográfico das Editoras de Universidades Federais Brasileiras; 2) Levantamento bibliográfico sobre o Open Monograph Press; 3) Aplicação do questionário; 4) Análise e interpretação dos dados coletados; 5) Apresentação dos resultados.

Para coleta das informações necessárias para o andamento da pesquisa foi utilizado, como instrumento para coleta de dados, um questionário intitulado “**Pesquisa sobre formatos utilizados para a disponibilização de *e-books* nas Editoras Universitárias de Universidades Federais**”, enviado, via *e-mail*, para todas as editoras universitárias.

O questionário foi composto por 09 (nove) questões, estruturadas da seguinte forma: 04 (quatro) abertas, 02 (duas) mistas, e 03 (três) fechadas. Com a aplicação do questionário foi construído um panorama da realidade das Editoras Universitárias, em relação aos padrões de *e-books* utilizados para disponibilização e editoração dos mesmos nestes ambientes.

4 ANÁLISE E INTERPRETAÇÃO DOS DADOS COLETADOS

Dentre as 63 (sessenta e três) universidades pesquisadas, 15 (quinze) universidades não dispõem de editoras universitárias, representando um percentual de 23,80% do total. Apenas 03 (três) universidades não deram retorno, o que corresponde a aproximadamente 4,77% do total.

E identificou-se que, aproximadamente, 71,43% do total, ou seja, 45 universidades federais brasileiras dispõem de editoras.

Para aplicação do instrumento, se fez necessário coletar os e-mails para contato das editoras, entrando em contato com assessoria de comunicação das universidades, ou apenas visitando os sites das editoras. Posteriormente o questionário foi enviado eletronicamente para

as 45 editoras de universidades federais, e ficou ativo para preenchimento durante dois meses, durante o período de 27/03/2014 a 27/05/14. Deste universo, apenas 44,44% deram retorno, totalizando 20 editoras.

Como mencionado, o questionário foi composto por 09 questões. As três primeiras questões diziam respeito às informações básicas sobre as editoras, tais como: nome da editora, instituição a qual pertencem e *e-mail* para posterior contato.

As três questões seguintes, dizem respeito à editoração e publicação de *e-books*. Tem como finalidade identificar se as editoras trabalham com editoração e publicação de *e-books*, os formatos de *e-books* utilizados, em caso afirmativo, bem como, em caso negativo, a justificativa do porquê ainda não trabalham com editoração e publicação de *e-books*.

Identificou-se que 50% das editoras que preencheram o questionário, não trabalham com editoração e publicação de *e-books*, mas há a pretensão, por parte das mesmas, em se trabalhar com esta modalidade de livro. E que as outras 10 editoras, representando 50%, trabalham com editoração e publicação de *e-books*. Dentre os formatos de *e-books* utilizados por elas, identificou-se que 68,33% utilizam o formato PDF; 18,33% usam o formato ePUB; 3,34% utilizam o formato Mobi e 10% utilizam outros formatos para editoração e publicação de *e-books* que não foram listados no questionário

Quanto à utilização de sistemas para editoração de *e-books*, das 20 editoras questionadas, 15 editoras, representando 75% do total, não utilizam sistemas para editoração e publicação de *e-books* e, apenas 25% utilizam este recurso.

Verificou-se que 05 editoras, das 20 questionadas, utilizam algum sistema eletrônico no processo de edição de *e-books*. Dentre esses, o *OMP* se destaca, por ser utilizado por 02 das 05 editoras que utilizam sistemas eletrônicos. As outras editoras utilizam outros sistemas que não foram listados no questionário, para o processo de edição de *e-books*.

No que tange ao quesito disponibilização, 13 editoras, 65% do total, disponibilizam gratuitamente seus *e-books* aos usuários, enquanto que 07 delas não os disponibilizam de forma gratuita, representando 35% do total.

5 CONCLUSÕES

Com base nos dados coletados, através do questionário aplicado às editoras universitárias brasileiras, a editoração e publicação de *e-books*, neste contexto, ainda estão em consolidação. Como visto, 44,44% das editoras questionadas trabalham com *e-books*. Quanto ao formato de publicação, verificou-se que 68,33% utilizam o formato PDF, sendo este o formato, para publicação de *e-books*, mais utilizado nesta ambência.

Consequentemente, 55,56% das editoras questionadas ainda não trabalham com editoração e publicação de *e-books*.

Quanto à utilização de sistemas automáticos para editoração e publicação de *e-books*, foi visto que dentre as 20 editoras participantes, 05 delas utilizam algum sistema para tal. O sistema com maior destaque foi o *OMP* utilizado por 02 das 05 editoras que utilizam sistemas para editoração e publicação de *e-books*, uma delas sendo a Editora da Universidade Federal da Paraíba.

A implantação do *OMP*, na EDUFPB, se deu a partir do uso individual por parte dos funcionários da mesma, visando à obtenção de práticas e posterior conhecimento, como também através de treinamentos e reuniões.

Quanto à disponibilização de forma gratuita dos *e-books* editorados e publicados pelas editoras para com os seus usuários, 65% delas os disponibilizam gratuitamente.

Por fim, conclui-se que na ambiência da Editora da Universidade Federal da Paraíba, considerando que o sistema utilizado para editoração e publicação de *e-books* será o *OMP*, e que o mesmo não exige um formato de *e-book* específico, os formatos de *e-books* mais adequados para a disseminação da informação nesta ambiência seriam o PDF, por ser o mais recorrente nas editoras universitárias brasileiras, bem como o ePUB, por ser o formato mais recomendado para os *e-books*.

REFERÊNCIAS

ARAUJO, W. J. *et al.* Elemento tecnológicos de edição, manipulação, e uso dos livros digitais. **Informação & Sociedade: Estudos**, João Pessoa, v. 23, n. 1, p. 13-25, jan. /abr. 2013.

DIAS, G. A. **E-books: alguns insights...** 2010. Disponível em:
<http://wcro.ccsa.ufpb.br/wrc0/?p=63>. Acesso em: 12 fev. 2014.

PUBLIC Knowledge Project. 2012. **Open Monograph Press**. Disponível em:
<http://pkp.sfu.ca/omp/>. Acesso em: 10 mar. 2014.

VERGARA, S. C. Projetos e relatórios de pesquisa em administração. São Paulo: Atlas, 2006.

BIG DATA APLICADO A SISTEMAS CIBER-FÍSICOS DA LOGÍSTICA: PROPOSTA DE MODELO CONCEITUAL

BIG DATA APPLIED TO CYBER-PHYSICAL LOGISTIC SYSTEMS: A PROPOSED CONCEPTUAL MODEL

Moisés Lima Dutra¹³⁹

William Barbosa Vianna¹⁴⁰

Enzo Morosini Frazzon¹⁴¹

Resumo: Os Sistemas Ciber-físicos combinam aspectos cibernetícios da computação e comunicação com a dinâmica dos sistemas físicos operados no mundo real. Juntamente com a grande relevância do gerenciamento e análise das informações com vistas ao planejamento e controle dos sistemas logísticos, considera-se que utilização de técnicas e tecnologias de Big Data e Business Analytics possa impulsionar a tomada de decisão dentro destes sistemas. Neste sentido, o processo de tomada de decisão é estratégico para se lidar com enormes quantidades de dados, analisar a informação e determinar como se trabalhar com tal volume e variedade de informações na velocidade adequada. Este trabalho é uma pesquisa aplicada, que propõe um modelo conceitual para combinar práticas de gestão com a aplicação de técnicas inovadoras para aquisição e análise de dados nos Sistemas Ciber-físicos. Ao final, conclui-se que as organizações não irão colher os benefícios de uma transição para o uso de Big Data, a não ser que sejam capazes de gerenciar efetivamente este processo de mudança.

Palavras-chave: Big Data. Business Analytics. Sistemas Ciber-físicos. Logística.

Abstract: Cyber-Physical Systems combine the cybernetic aspects of computing and communication with the dynamics and physics of physical systems operating in the real world. Along with the high relevance of information management and analysis for the planning and control of logistic systems, we consider that the concepts and tools of Big Data and Business Analytics may play an important role regarding the decision-making process within these systems. In this sense, the decision-making process is strategic to deal with large volumes of data, analyze the information and determine how working with this volume and variety of information at proper speed. This work is an applied research, which aims to propose a conceptual model and to investigate management perspectives regarding the application of innovative techniques for data acquisition and analytics in Cyber-Physical Logistic Systems. In the end, we can conclude that organizations will not reap the benefits of a transition with the use of Big Data, unless they are able to manage the change effectively.

Keywords: Big Data. Business Analytics. Cyber-Physical Systems. Logistics.

1 INTRODUÇÃO

Considera-se a Logística uma parte da gestão da cadeia de suprimentos que se ocupa da movimentação de produtos desde a emissão do pedido até a distribuição, por meio de estágios, tais como aquisição de matéria prima, produção, movimentações internas e externas

¹³⁹ UFSC.

¹⁴⁰ UFSC.

¹⁴¹ UFSC.

de suprimentos e de distribuição sendo que, neste contexto, destaca-se como ponto crítico o gerenciamento das informações relativas a este processo (DUTRA et al., 2013).

A competitividade de uma cadeia de suprimentos depende cada vez mais da sua flexibilidade. As melhores não são apenas rápidas e eficientes, mas igualmente ágeis e autoadaptáveis (LEE, 2004). Por exemplo, períodos de desenvolvimento mais curtos, produção de bens variáveis e individualmente configuráveis são estratégias apropriadas para satisfazer o consumidor e responder tão rápido quanto possível às mudanças de demanda existentes (RICHTER, 2007). Logo, a logística precisa compensar os requisitos crescentes da flexibilidade por meio da utilização de uma maior densidade da informação e do emprego de técnicas inovadoras para a aquisição e análise de dados. Consequentemente, o fluxo da informação que comanda todo este processo representa uma questão crítica a ser abordada, e tem propulsionado uma grande quantidade de investimentos em tecnologias de apoio (BANDEIRA; MAÇADA, 2008; LASETER; OLIVER, 2005), tais como as tecnologias e técnicas de Big Data.

Este estudo faz parte de um projeto de pesquisa em andamento que envolve a Ciência da Informação e a Engenharia de Produção. Nele, buscou-se apresentar uma proposta de modelo conceitual que utilize as técnicas de Big Data aplicadas dentro de um contexto de Sistemas Ciber-físicos da Logística.

2 SISTEMAS CIBER-FÍSICOS

O conceito de um sistema que une funções físicas e perspectivas relacionadas à informação vem sido bastante trabalhado ultimamente (BROY, 2010). As capacidades dos assim chamados Sistemas Ciber-físicos (SCF) permitem a divisão do sistema em diversas entidades capazes de se comunicar, de reconhecer seu ambiente e de tomar decisões. Um SCF acumula os benefícios advindos dos sistemas embarcados com a possibilidade de comunicar-se numa variedade grande de tecnologias de comunicação. Além disso, eles permitem a conexão entre o fluxo informacional e o fluxo de material nos processos logísticos (HRIBERNIK et al., 2010).

A combinação de objetos físicos com inteligência cibernética, trazida pela aplicação dos SCF, cria um novo potencial para eficiência melhorada, *accountability* (capacidade para a prestação de contas), sustentabilidade e escalabilidade nos sistemas logísticos. De fato, esta combinação é considerada por muitos pesquisadores e usuários como uma das grandes tendências e um dos maiores desafios da indústria no futuro. Com respeito a esta integração, é primordial que uma boa gestão de informação seja a viga mestre para o desenvolvimento de

sistemas de apoio à decisão apropriados (DUTRA et al., 2013). O conhecimento deste contexto é um dos fundamentos dos sistemas ciber-físicos. É possível agregar informação de sistemas distribuídos de dados relacionados à logística por meio da combinação de padrões e sistemas já existentes. Através da agregação dos SCF com as redes de comunicação, é fomentada a interação entre o mundo físico e o mundo cibernetico (HANS et al., 2008).

3 BIG DATA E BUSINESS ANALYTICS

De acordo com Krishnan (2013), Big Data consiste de uma vasta quantidade de dados disponíveis em diferentes níveis de complexidade – criados por humanos ou por máquinas em diferentes ritmos –, e que apresentam grandes níveis de ambiguidade, de forma que não podem ser processados computacionalmente por meio da utilização de tecnologias, dispositivos de comunicação, métodos de processamento e algoritmos tradicionais, além de qualquer outra solução similar. Segundo Katal et al. (2013), Big Data se refere a grandes quantidades de dados que necessitam de novas tecnologias e arquiteturas para efetuar a recuperação da informação, que deve ser feita através de processos específicos de coleta e análise.

Business Analytics (BA) é um conceito definido como a aplicação de várias técnicas avançadas de análise de dados, de maneira a responder determinadas indagações ou resolver problemas numa determinada área do conhecimento (DAVENPORT, 2009). Esta definição se refere ao uso de modelos quantitativos e preditivos e de apoio baseado em fatos para o suporte à decisão. O significado de BA abrange mineração de dados, análise preditiva, estatística e análise aplicada. Desta forma, pode-se ver que o conceito de BA possui muitas similaridades com o conceito de Big Data, pois ambos procuram recuperar informações a partir de dados.

Ward e Barker (2013) relatam que, depois de analisarem diversas definições de Big Data, chegaram à conclusão de que todas fazem alusão a três importantes aspectos: (i) tamanho: o volume de dados é um fator crítico; (ii) complexidade: a estrutura, seu comportamento e as permutas feitas entre os *datasets* também são um fator crítico; e (iii) tecnologia: as ferramentas e técnicas usadas para processar os *datasets* são essenciais.

No entanto, em consonância com a análise acima, três aspectos diferenciam Big Data de BA: volume, velocidade e variedade. O seja, os dados do Big Data, juntamente com a tecnologia necessária para coletá-los e processá-los, diferem do BA pelos enormes valores nesses três aspectos. Além disso, em alguns casos, a velocidade da criação de dados é até mais importante do que o volume dos mesmos, por exemplo, informação em tempo-real, ou em tempo-quase-real, pode ser decisiva para o sucesso ou fracasso de sistemas que operem nesta

velocidade (WARD; BARKER, 2013). Com relação às características do conteúdo existente no Big Data, pode-se verificar que são diversas e variadas entre si. A Figura 1 mostra algumas destas características.

Dados dentro do Big Data podem ser coletados sob diversas formas, Por exemplo, sob a forma de atualizações em redes sociais (Facebook, Instagram), mapeamento remoto (Google Maps, GPS), fotos (Flickr), vídeos (Youtube), ferramentas de busca (Google, DuckDuckGo), entre outras. Todas essas formas de dados formam um subconjunto Big Data conhecido como dados não-estruturadas. Cada pessoa, hoje em dia, é um produtor de dados em potencial. As tecnologias mais tradicionais, que processam dados estruturados, tais como bases de dados estruturadas, não servem completamente para o armazenamento e processamento no ambiente Big Data. As novas tecnologias (armazenagem, memória, processamento e largura de banda) que surgiram nos últimos anos foram as grandes impulsionadoras da análise de dados no contexto Big Data (ARELLANO, 2013).

FIGURA 1 – Conteúdo Big Data



Fonte: Os autores

Além disso, o desenvolvimento dos processos de virtualização e a distribuição de produtos/serviços dentro dos ambientes colaborativos requerem a integração de processos e a interoperabilidade semântica de dados e informação, bem como o uso de ontologias para facilitar tal integração. Há uma série de tarefas potencialmente beneficiárias do uso de ontologias no contexto da integração logística numa cadeia de suprimentos. Representar formalmente um processo, produto ou serviço significa dar um passo fundamental na direção de se remover a ambiguidade, que é um problema crucial no ambiente Big Data. Além disso, uma boa representação ontológica permite inferir e recuperar a informação que não está

explícita à primeira vista, o que possibilita identificar qualquer situação conflitante ou inconsistente dentro do contexto das relações entre as entidades do sistema. Com uma longa tradição nos campos da Filosofia, Ciência da Computação e Ciência da Informação, as ontologias provêm suporte para a engenharia do conhecimento e para a inteligência artificial, por meio da modelagem das diversas áreas do conhecimento em termos de conceitos, atributos e relações, geralmente classificadas em relações hierárquicas do tipo especialização/generalização (SIMPERL, 2009; MCKINSEY GLOBAL INSTITUTE, 2011).

4 MODELO BIGLOGDATIX

Este trabalho é uma pesquisa aplicada, que propõe um modelo conceitual para combinar práticas de gestão com a aplicação de técnicas de Big Data para aquisição e análise de dados dentro de um contexto de Sistemas Ciber-físicos da Logística.

As modernas organizações que estão envolvidas em projetos de desenvolvimento de novos produtos/serviços, para continuarem competitivas, deveriam adotar métodos flexíveis de trabalho de maneira a satisfazer as numerosas e variadas demandas do mercado global. Esta flexibilidade deveria não somente se constituir da capacidade de oferecer boas respostas aos seus clientes, mas também da capacidade de se detectar potenciais mudanças e futuras tendências em todo o sistema. No cenário dos sistemas ciber-físicos, esta estratégia significa combinar e coordenar todos os atores envolvidos – humanos e não-humanos –, desde a entrada da matéria-prima até a entrega do produto/serviço final.

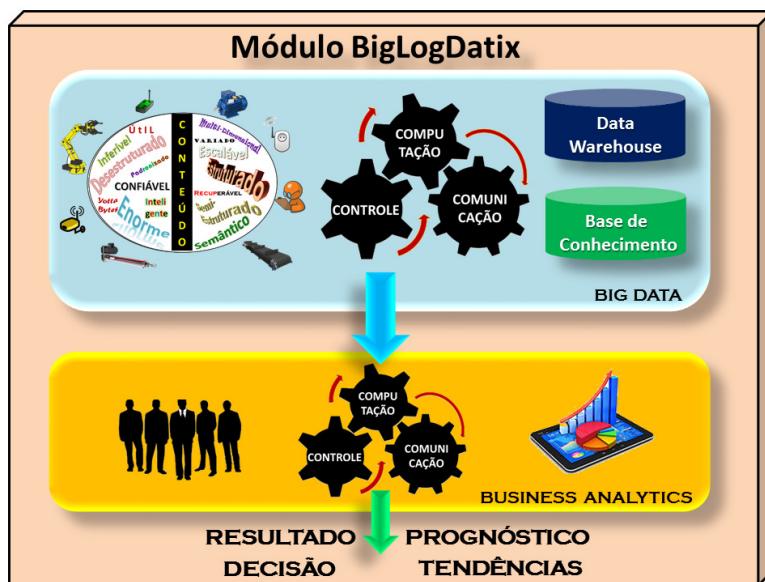
Esta combinação depende do complexo processo de integração de três problemas essenciais: computação, controle e comunicação (CYBER-PHYSICAL SYSTEMS, 2012). Integrar estas três dimensões significa considerar os diferentes participantes e seus pontos de vista, diferentes áreas de conhecimento, diferentes topologias de rede e equipamentos, diferentes requisitos e diferentes estágios de atividades locais e globais. Quando este processo é completado com sucesso, ele é capaz de produzir um sistema autoadaptável e mais ágil.

O modelo proposto (Figura 2) procura satisfazer critérios reativos e pró-ativos, pois o mesmo processa não somente informação de *feedback* (por exemplo, opinião de consumidores sobre determinado produto), como também procura revelar prognósticos e descobrir informações que se configurem tendências. De maneira a manter este sistema sustentável e escalável, considera-se que o compartilhamento de informação é um atributo fundamental, pois fornece entradas de dados pré-processadas a atores humanos e não-humanos, de maneira que estes possam utilizá-las para tomar decisões melhores. As informações de *feedback* irão permitir ajustes, correções e atualizações em todo o SCF. Informações do tipo prognóstico e

tendências futuras – inferidas e produzidas através de técnicas de análise – serão a base para ações da organização visando inovação, antecipação e estratégias criadoras de valor, que visam o ganho competitivo em relação aos competidores. Para que tal cenário seja alcançado, o modelo BigLogDatix irá se beneficiar de técnicas de Big Data e Business Analytics.

Destacamos os esforços feitos para que o modelo BigLogDatix combine aspectos de computação, controle e comunicação do Big Data, já que a integração destes três domínios é um dos principais desafios da pesquisa nos SCF (CYBER-PHYSICAL SYSTEMS, 2012). O BigLogDatix é estruturado em módulos distribuídos e ligados em rede. Cada um desses módulos é configurável e instanciável para um cenário específico da cadeia de suprimentos. Isto é, estes módulos são customizáveis para diferentes atores da logística, tais como o controle do chão de fábrica, o gestor da relação com o consumidor, o grupo de BA, entre outros. Apesar das circunstâncias serem diferentes, os métodos e técnicas utilizados são similares.

FIGURA 2 – Módulo BigLogDatix



Fonte: Os autores.

Um Data Warehouse é proposto para integrar os dados coletados nas diferentes origens. Este repositório central será alimentado através de processos ETL (*extract, transform, load*), que transformarão os dados de seus formatos previamente coletados em formato BigLogDatix. Além disso, propomos igualmente uma representação ontológica da informação, de maneira que se possa efetuar um processo de inferência da informação e do conhecimento. O resultado deste motor de inferência irá alimentar uma Base de

Conhecimento, que será utilizada para a descoberta e reconhecimento de informações e conhecimentos tácitos. Bohn e Short (2010) dizem que:

[...] há muito critérios em potencial para medir o valor da informação, incluindo julgamento subjetivo, preço de venda, disposição dos consumidores em pagar, custo de desenvolvimento tamanho do público. Mas não há maneira de comparar valor, especialmente quando se compara informação de diferentes tipos.

Considerando esta afirmação, propomos uma máquina de inferência para inferir e comparar, a partir de representações ontológicas, o valor da informação.

Desta forma, o modelo conceitual BigLogDatix pretende apoiar o processo de tomada de decisão no contexto Big Data. Consideramos que os problemas não têm apenas uma existência física, mas que eles dependem da intervenção de um ou mais atores humanos e não-humanos que interagem por meio dos processos de comunicação, controle e computação, dentro da tomada de decisão nos SCF aplicados à logística.

5 CONCLUSÕES

Consideramos estar no limiar de uma grande mudança da maneira pela qual decisões são tomadas dentro das organizações. Podemos estar testemunhando um período no qual a micro-segmentação de tempo-real de cidadãos e consumidores irá atingir o seu pico, através da evolução e do uso das técnicas de Big Data e Business Analytics. Estratégias sofisticadas de análise podem melhorar substancialmente o processo de tomada de decisão, minimizar riscos e descobrir informações importantes, que de outra forma permaneceriam ocultas.

AGRADECIMENTOS

Este trabalho foi parcialmente financiado pelo CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico), no escopo da chamada MCTI/CNPq/MEC/CAPES – 43/2013, que apoia financeiramente projetos que visem contribuir significativamente para o desenvolvimento científico e tecnológico e para a inovação do País nas áreas de Ciências Humanas, Sociais e Sociais Aplicadas.

REFERÊNCIAS

ARELLANO, N. **3 key differences between big data and analytics.** It World Canada, 22 Set. 2013. Site dedicado à veiculação de notícias de tecnologia de informação. Disponível em: <<http://www.itworldcanada.com/post/3-key-differences-between-big-data-and-analytics>>. Acesso em: 30 Jul. 2014.

BANDEIRA, R. A. M.; MAÇADA, A. C. G. Tecnologia da informação na gestão da cadeia de suprimentos: o caso da indústria gases. **Produção**, São Paulo, v. 18, n. 2, 2008.

BOHN, R. E.; SHORT, J. E. **How Much Information? 2009 Report on American Consumers.** Global Information Industry Center, University of California, San Diego, Janeiro 2010. Available in: <http://hmi.ucsd.edu/pdf/HMI_2009_ConsumerReport_Dec9_2009.pdf>. Acesso: 30 Jul. 2014.

BROY, M. Cyber-Physical Systems — Wissenschaftliche Herausforderungen Bei Der Entwicklung, In: **acatech DISKUTIERT**, 0:17-31. 2010.

CYBER-PHYSICAL SYSTEMS: Milestones and research challenges. Computer Communications. Editorial. Elsevier, n. 36, pg 1-7, 2012.

DAVENPORT, T. **Realizing the Potential of Business Analytics:** Plenty of Food for those with the Appetite. Wellesley, MA, USA: Babson, 2009. Disponível em: <<http://analytics.typepad.com/files/retailanalytics.pdf>>. Acesso em: 30 Jul. 2014.

DUTRA, M.L.; VIANNA, W. B.; MATIAS, M.; PINTO, A. L. Uma Abordagem Baseada em Ontologias para a Área da Logística. In: ALMEIDA, F.A.S.; SILVA, A.M.; FRANCO, M.J.B.; BRITO, P.Q.; FREITAS, C.C. (Org.). Coletânea Luso-Brasileira IV: Gestão da informação, Inovação e Logística. 1ed. Goiânia: Senai, 2013, v. 1, p. 533-560.

HANS, C.; HRIBERNIK, K. A.; THOBEN, K. D. An approach for the integration of data within complex logistics systems. In: **Dynamics in Logistics**. Springer Berlin Heidelberg, 381-390. 2008.

HRIBERNIK, K.; KRAMER, C.; HANS, C.; THOBEN, K. D. A Semantic Mediator for Data Integration in Autonomous Logistics Processes. In: **Enterprise Interoperability IV**. Springer, London, 157-167. 2010.

KATAL, A.; WAZID, M.; GOUDAR, R.H. Big data: Issues, challenges, tools and Good practices. In: **International Conference on Contemporary Computing (IC3)**, VI, 8-10 Ago. 2013. Disponível em: <<http://ieeexplore.ieee.org/xpls/icp.jsp?arnumber=6612229&tag=1>>. Acesso em: 30 Jul. 2014.

KRISHNAN, K. Data Warehousing in the Age of Big Data. Newnes. 2013. ISBN: 978-0-12-405891-0.

LASETER, T.; OLIVER, K. The Missing Link – When will Supply chain management grow up?. Strategy + business, Booz Allen & Hamilton, 2005.

LEE, H. L. Cadeias de suprimentos do século XXI: Cadeia de suprimentos triplas. Harvard Business Review. p.74-84. 2004.

MCKINSEY GLOBAL INSTITUTE. Big data: The next frontier for innovation, competition, and productivity. Technical Report, McKinsey Global Institute. 2011.

RICHTER, K. Aufbau globaler Netzwerke als Erfolgsfaktor in der Automobilindustrie. In: **Die Automobilindustrie auf dem Weg zur globalen**, Springer, 251-264. 2007.

SIMPERL, E. Reusing ontologies on the Semantic Web: A feasibility study. Data and Knowledge Engineering. 2009.

WARD, J. S; BARKER, A. **Undefined By Data: A Survey of Big Data Definitions**. Saint Andrews: University of Saint Andrews, 20 Set. 2013. Disponível em: <<http://arxiv.org/pdf/1309.5821v1.pdf>>. Acesso em: 30 Jul. 2014.

VISUALIZAÇÃO DE CORRESPONDÊNCIAS SEMÂNTICAS NO UNIVERSO BIG DATA

VISUALIZING SEMANTIC CORRESPONDENCES IN BIG DATA ENVIRONMENTS

Moises Lima Dutra

Marcio Matias

Resumo: O Big Data consiste de uma vasta quantidade de dados disponíveis em diferentes níveis de complexidade – criados por humanos ou por máquinas, em diferentes ritmos – e que apresentam grandes níveis de ambiguidade, de forma que não podem ser processados computacionalmente por meio da utilização de tecnologias, dispositivos de comunicação, métodos de processamento e algoritmos tradicionais, além de qualquer outra solução similar. Desta forma, pode-se dizer que a recuperação da informação nessas grandes aglomerações de dados necessita de novas tecnologias e técnicas de apoio, que deverão modificar os processos tradicionais de coleta e análise da informação. Este trabalho utiliza abordagem qual-quantitativa. Trata-se de uma pesquisa exploratória e aplicada, que envolve pesquisa bibliográfica e faz parte de dois projetos de pesquisa em andamento. O objetivo deste trabalho é apresentar uma proposta preliminar para o desenvolvimento de visualizador de correspondências semânticas no contexto Big Data, de maneira a facilitar a interação humano-computador. Por fim, concluiu-se que a possibilidade de visualizar correspondências semânticas dentro do universo Big Data, com interfaces adequadas quanto a sua usabilidade, pode melhorar sensivelmente o processo humano de tomada de decisão, pois possibilita descobrir e trazer à tona informações importantes, que de outra forma permaneceriam ocultas.

Palavras-chave: Big Data. Visualização. Correspondências semânticas. Ontologias. Usabilidade.

Abstract: Big Data consists of a vast amount of data available in different complexity levels, created by humans or machines at diverse paces and presenting several levels of ambiguity, so that they cannot be computed by using traditional technologies, communication devices, processing methods, algorithms or any off-the-shelf solution. In this sense, we can say that retrieving information in these huge datasets requires new supporting technologies and techniques, which should in turn change the traditional processes of information gathering and analysis. This work is a bibliographic and applied research, and is part of two ongoing research projects, supported by CNPq. It presents a preliminary proposed visualizer for semantic correspondences in Big Data, in order to facilitate the human-computer interaction. In the end, we can conclude that visualizing semantic correspondences with usability in the Big Data context substantially improves the human decision-making process, since it enables to unfold important information, which would otherwise remain unknown.

Keywords: Big Data. Visualization. Semantic correspondences. Ontologies. Usability.

1 INTRODUÇÃO

Big Data é um termo utilizado para descrever as volumosas quantidades de informações estruturadas, semi-estruturadas e não-estruturadas que as organizações produzem. Essas informações levariam muito tempo e necessitariam de muitos recursos para serem todas carregadas em uma base de dados convencional, de maneira que pudessem ser analisadas pelos modelos tradicionais. Embora não haja uma definição formal quanto à

quantidade desses dados, convencionou-se usar o termo *Big Data* para situações em que se trata de *petabytes* ou *exabytes* de dados (MCKINSEY GLOBAL INSTITUTE, 2011).

Segundo IBM (2012), diariamente são criados e disponibilizados 2,5 quintilhões de *bytes* de dados. Este volume é de tal monta que 90% dos dados existentes no mundo hoje foram criados nos dois últimos anos. Além disso, 80% dos dados existentes não estão estruturados. “Esses dados vêm de toda parte: sensores usados para coletar informações climáticas, textos postados em redes sociais, fotos e vídeos digitais, registros de transações de compra e venda, informações de GPS e de telefones celulares, apenas para nomear alguns. Isto é *Big Data*”.

O conceito de *Big Data* se estende por três dimensões: volume, velocidade e variedade. Para IBM (2012, tradução livre), estas dimensões estão assim representadas:

- Volume: as organizações estão “atoladas” em um número cada vez mais crescente de dados de todos os tipos, facilmente acumulando *terabytes* – às vezes mesmo *petabytes* – de informação. Exemplos de desafios: Como analisar os 12 *terabytes* de mensagens postadas no Twitter diariamente? Ou converter as 350 bilhões de leituras anuais dos medidores para melhor prever o consumo de energia?
- Velocidade: Algumas vezes, 2 minutos podem ser tempo demais. Para processos baseados no tempo, como a detecção de fraudes eletrônicas, a ideia de *Big Data* pode ser utilizada pelas organizações, com o intuito de maximizar suas potencialidades. Exemplos: i) Examinar minuciosamente 5 milhões de transações comerciais feitas diariamente, para identificar potenciais fraudes. ii) Analisar 500 milhões de chamadas telefônicas diárias em tempo real, para identificar clientes em potencial.
- Variedade: *Big Data* é qualquer tipo de dado – estruturado ou não –, como texto, dados de sensores, áudio, vídeo, sucessão de cliques, registro de operações em determinado sistema, e muito mais. Novas ideias surgem quando todos esses tipos de dados são analisados juntos. Exemplos: i) Monitorar 100% das imagens de vídeos obtidas a partir de câmeras de segurança de maneira e identificar pontos de interesse. ii) Explorar 80% do aumento de imagens, vídeos e documentos em geral, de maneira a incrementar a satisfação dos usuários.

De acordo com o HMI Report of Global Information Industry Center (BOHN; SHORT, 2010), a quantidade de informação armazenada em sistemas locais ou distribuídos e disponibilizada publicamente na Web – ou de maneira privada, via bases de dados

proprietárias – está em constante crescimento. Estima-se que o consumo de informação *per capita* diário gire em torno de 34GB por pessoa. Para Robredo (2011),

[...] essas quantidades de dados irão, cada vez em maior escala, para alguma nuvem, e uma grande parte desses dados – mas não sempre os mesmos – estarão permanentemente em linha. [...] dos 35 ZB em 2020, 12 ZB passarão em algum momento por uma nuvem, e 5 ZB se fixarão numa nuvem. [...] dados médicos, de segurança nacional, estratégicos, dados bancários, cadastros de clientes, endereços de e-mails, sistemas cooperativos, grupos sociais, e mil coisas mais já estão armazenadas em alguma nuvem.

Um *zettabyte* (ZB) equivale a um sextilhão de *bytes*. Segundo Sanchati e Kulkarni (2011), em breve as unidades de informação estarão construindo e gerenciando seus próprios *data centers*. Os autores também acham que a utilização da nuvem computacional é uma escolha bastante atrativa, e que será desafiada pelo crescimento do número de documentos indexados e pelas novas ferramentas tecnológicas que surgirem mais à frente.

Para Mayer-Schönberger e Cukier (2013, tradução livre),

Na sua essência, o Big Data trata de previsões. Embora seja descrito como um ramo da Ciência da Computação chamado Inteligência Artificial, e mais especificamente como uma subárea desta, chamada Aprendizado de Máquina (*Machine Learning*), esta classificação não está totalmente correta. Big Data não diz respeito a “ensinar” um computador a “pensar” como humanos. Ao invés disso, diz respeito à aplicação de técnicas matemáticas em grandes quantidades de dados de maneira a inferir probabilidades, por exemplo: a possibilidade de que um e-mail seja um *spam*; que a palavra digitada “cinêcia” seja na verdade “ciência”; que a velocidade e a trajetória de uma pessoa distraidamente atravessando uma rua implique que um carro auto-dirigido deva ajustar sua velocidade de maneira gradual, para dar tempo à mesma de chegar ao outro lado. A chave para o bom desempenho destes sistemas é a alimentação dos mesmos com enormes quantidades de dados, sobre os quais previsões serão feitas. Além disso, os sistemas são construídos para se autoaperfeiçoar com o passar do tempo, através do registro das melhores informações e dos melhores padrões obtidos, a partir dos quais novas informações serão procuradas futuramente. [...] No futuro – e mais cedo do que imaginamos – muitos aspectos de nosso mundo serão melhorados ou substituídos por sistemas computacionais que facilitarão a extensão do julgamento humano.

Esta capacidade de “aprender a pensar” está relacionada com um conceito herdado da filosofia: a ontologia. De acordo com o Novo Dicionário Aurélio da Língua Portuguesa (2004), *Ontologia* é uma “parte da filosofia que trata do ser enquanto ser, *i. e.*, do ser concebido como tendo uma natureza comum que é inerente a todos e a cada um dos seres”. Noy et al. (2001) afirmam que ontologias combinadas juntamente com um conjunto de instâncias de classes constituem uma base de conhecimento. “Há uma linha tênue que define onde a ontologia acaba e a base de conhecimento começa” (NOY et al., 2001, tradução livre).

Em uma visão voltada à Ciência da Informação, as ontologias tratam das construções que estruturam conteúdos explícitos e que são capazes de codificar as regras implícitas de

uma parte da realidade, ainda que trabalhem com declaração explícita independente do fim do domínio da aplicação (GARCÍA JIMÉNEZ, 2002), no qual o foco nas suas representações em prol de uma eficiente recuperação da informação, baseado em taxonomias e tesouros e linguagem natural da área. Para Bloehdorn et al. (2009), ontologias são essenciais para facilitar a interoperabilidade semântica e a integração de dados e processos. Eles também acreditam que nós estamos atualmente entrando em uma fase do desenvolvimento de sistemas na qual ontologias são produzidas em grande número e grande complexidade. Desta forma, a criação e gestão de infraestruturas para ontologias seriam necessárias para servir de suporte ao crescente número de aplicações, especialmente da Web Semântica, que trabalham com interoperabilidade semântica dentro das organizações.

Podemos considerar, portanto, a utilização de ontologias como sendo de grande importância no contexto Big Data, pois as mesmas permitem um tipo de representação que facilita a descoberta de correspondências semânticas entre conceitos *a priori* distintos. Estas correspondências poderão ser utilizadas como base para a recuperação de informações e conhecimentos implícitos.

O objetivo deste trabalho é apresentar uma proposta preliminar para o desenvolvimento de visualizador de correspondências semânticas no contexto Big Data, com usabilidade.

2 VISUALIZAÇÃO E USABILIDADE

A proposta de desenvolvimento de um visualizador de correspondências semânticas no contexto Big Data utilizará conceitos e princípios de usabilidade, ou de facilidade de uso. Vale ressaltar, que a usabilidade é uma característica essencial para facilitar o uso de sistemas, e a compreensão das informações apresentadas por um sistema de informação, ou por quaisquer elementos gráficos que possam representar informações para os usuários em um ambiente computacional.

Segundo a norma ABNT NBR ISO 9241-11, a usabilidade é definida como: “medida na qual um produto pode ser usado por usuários específicos para alcançar objetivos específicos com eficácia, eficiência e satisfação em um contexto específico de uso” (ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS, 2011).

Na interação entre a engenharia de usabilidade e a Ciência da Informação, surgem oportunidades para que especialistas em usabilidade e profissionais da informação possam trabalhar conjuntamente, interagir e compartilhar conhecimento a fim de projetar e desenvolver sistemas que possuam maior qualidade em suas interfaces.

De acordo com Wilkes (2012), especialistas em usabilidade possuem o conhecimento para melhorar usabilidade de dados, garantindo não apenas que as pessoas que precisam de certos dados, possam acessá-los da forma que esperam (a partir de notebooks, smartphones e/ou outros dispositivos móveis), mas que também estejam de forma adequada: significativa, agradável, atrativa e útil. Além disto, os especialistas em usabilidade e os profissionais da informação possuem conhecimento para sugerir soluções nas quais os metadados sejam melhor integrados com mostradores e painéis de dados, com a indicação dos dados que possuem maior qualidade e confiabilidade.

A complexidade envolvida na visualização de um grande volume de dados tem sido um tema desafiador para a Ciência da Informação, a Ciência da Computação, o Design, a Ergonomia e a Engenharia de Usabilidade; e algumas pesquisas, têm buscado desenvolver e testar possibilidades que possam gerar alternativas que permitam representar os mesmos fenômenos a partir de um conjunto menor de dados.

Paul e Bruns (2013), por exemplo, realizaram um estudo nesta linha com o Twitter no qual analisaram mais de 164.000 tweets coletados durante o terremoto de 2011 para descobrir que tipo de informação específica sobre localização, os usuários mencionaram nos seus tweets, e em que momento eles trocaram informação sobre isto. Baseados em sua análise, descobriram que mesmo um conjunto de dados reduzido, não caracterizado como Big Data, pode ser útil para encontrar rapidamente áreas específicas de desastres. Com esta estratégia, nesta situação, conseguiram descobrir que área requer mais ajuda, e também os nomes dos lugares que foram atingidos mais gravemente durante o desastre.

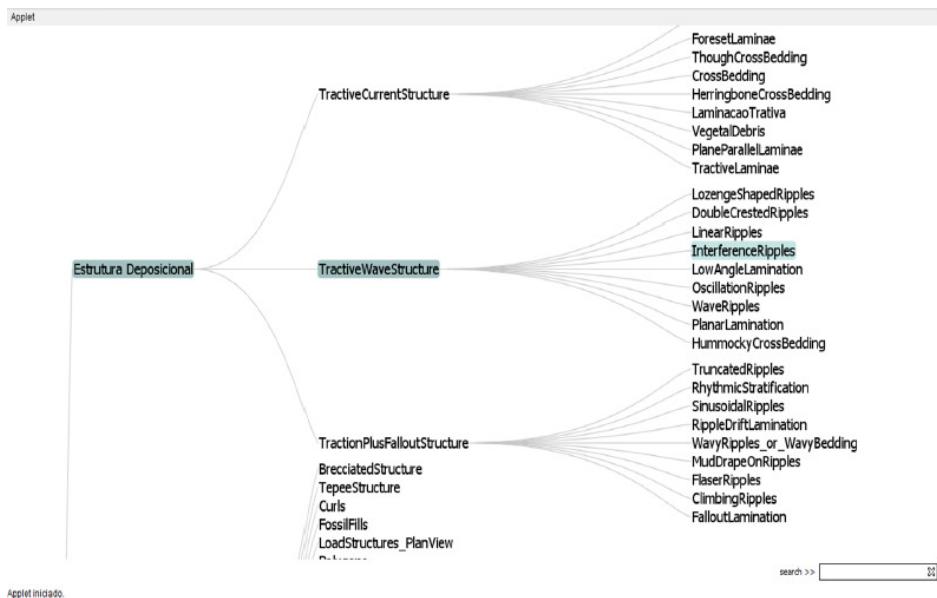
Entretanto, o grande desafio tem sido utilizar efetivamente grandes volumes de dados caracterizados como Big Data. Diversos visualizadores têm sido desenvolvidos e utilizados como instrumentos para entregar a informação de forma facilitada para os usuários.

Takeda et al. (2012) desenvolveram uma ferramenta de visualização chamada *Irregular Trend Finder* (ITF) que foi projetada para análise de grandes volumes de dados, e que utiliza princípios de usabilidade, pois permite a utilização de estruturas hierárquicas, a fim de que o usuário possa obter primeiro uma visão geral, e depois, possa obter informação mais detalhada sob demanda. Os resultados dos estudos com esta ferramenta mostraram que abordagem utilizada mostrou-se eficiente para os usuários buscarem informações particulares em grandes volumes de dados.

Schievelbein et al. (2014) descrevem um exemplo de ferramenta de visualização de ontologias e de captura e modelagem de conhecimento visual para um grupo de usuários

especialistas da área de Geologia em um ambiente Web 2.0. A Figura 1 apresenta um exemplo de tela desta ferramenta chamada OBAITA.

FIGURA 1 – Exemplo de ferramenta de visualização de ontologias – OBAITA



Fonte: Schievelbein et al. (2014)

3 ASPECTOS METODOLÓGICOS

Esta pesquisa é exploratória, e utiliza a abordagem quali-quantitativa. A abordagem qualitativa está associada à avaliação da usabilidade da ferramenta a ser desenvolvida; e a quantitativa está presente principalmente no tratamento dos dados referentes à visualização das correspondências semânticas.

Este estudo utiliza pesquisa bibliográfica, o paradigma da orientação a objeto para especificação e desenvolvimento da ferramenta de visualização, e as técnicas de avaliação de interfaces humano-computador da engenharia de usabilidade: avaliação heurística e inspeção de usabilidade (CYBIS, 2010).

Este trabalho envolverá as seguintes etapas a serem realizadas: levantamento bibliográfico sobre Big Data, visualização, correspondências semânticas, ontologias e usabilidade; levantamento sobre ferramentas existentes que integrem estes temas e conceitos; especificação da ferramenta de visualização proposta; desenvolvimento da ferramenta proposta; avaliação da usabilidade da ferramenta desenvolvida.

4 CONSIDERAÇÕES FINAIS

Vale ressaltar que este estudo faz parte de um projeto de pesquisa em andamento, e que, portanto, seus resultados parciais envolvem a construção da proposta que se caracteriza por explorar a interação entre as áreas de Ciência da Informação, Ciência da Computação e

Engenharia de Usabilidade e, mais especificamente, a aplicação integrada dos conceitos de big data, visualização, correspondências semânticas, ontologias e usabilidade.

A partir da literatura já levantada e consultada, pode-se afirmar que tal abordagem de integração mostrou-se viável e potencialmente profícua.

AGRADECIMENTOS

Este trabalho foi parcialmente financiado pelo CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico), no escopo das chamadas MCTI/CNPq/MEC/CAPES – 43/2013 (que apoia financeiramente projetos que visem contribuir significativamente para o desenvolvimento científico e tecnológico e para a inovação do País nas áreas de Ciências Humanas, Sociais e Sociais Aplicadas) e MCTI/CNPq – 14/2013 (Chamada Universal).

REFERÊNCIAS

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. NBR ISO 9241-11: requisitos ergonômicos para o trabalho com dispositivos de interação visual – Parte 11 – orientações sobre usabilidade. Rio de Janeiro, 2011.

BLOEDORN S., HAASE P., HUANG Z., SURE Y., VÖLKER J., VAN HARMELEN F., STUDER R. (2009). Ontology Management. In J. Davies et al. (eds.), **Semantic Knowledge Management**, Springer Berlin Heidelberg, 2009.

BOHN, R. E.; SHORT, J. E. **How Much Information?** 2009 Report on American Consumers. HMI Report of Global Information Industry Center, University of California, San Diego, Janeiro, 2010. Disponível em:
http://hmi.ucsd.edu/pdf/HMI_2009_ConsumerReport_Dec9_2009.pdf. Acesso em: 30 de julho de 2014.

CYBIS, W. **Ergonomia e usabilidade:** conhecimentos, métodos e aplicações. 2. ed. São Paulo: Novatec, 2010.

GARCÍA JIMÉNEZ, A. **Organización y gestión del conocimiento en la comunicación.** Madrid: TREA, 2002.

IBM. **What is big data?** In: Bringing big data to the enterprise. Disponível em:
<http://www.ibm.com/big-data/us/en/>. Acessado em: 30 de julho de 2014.

MAYER-SCHÖNBERGER, V.; CUKIER, K. **Big Data:** A Revolution that will transform how we live, work, and think. New York: Houghton Mifflin Harcourt, 2013.

MCKINSEY GLOBAL INSTITUTE. Big data: The next frontier for innovation, competition, and productivity. Technical Report, **McKinsey Global Institute**, Junho 2011.

NOVO DICIONÁRIO AURÉLIO DA LÍNGUA PORTUGUESA: revista e atualizada do Aurélio Século XXI, 3. edição, 1. Impressão, Editora Positivo, 2004.

NOY N.F., McGUINNESS D.L. Ontology Development 101: A Guide to Creating Your First Ontology. **Stanford Knowledge Systems Laboratory Technical Report KSL-01-05 and Stanford Medical Informatics Technical Report SMI-2001-0880**, March 2001.

PAUL, A.; BRUNS, A. Usability of small crisis data sets in the absence of big data. In Ariwa, Ezendu, Zhao, Wenbing, & Gandhi, Meenakshi (Ed.) **Proceedings of the 2013 International Conference on Information, Business and Education Technology**, Atlantis Press, Beijing, China, p. 718-721.

ROBREDO, J. Do documento impresso à informação nas nuvens: reflexões. **Liinc em Revista**, Rio de Janeiro, v.7, n.1, março, 2011.

SANCHATI, R.; KULKARNI, G. Cloud Computing in Digital and University Libraries. **Global Journal of Computer Science and Technology**, v. 11, n. 12, Jul. 2011.

SCHIEVELBEIN, G. *et al.* **Usabilidade em ambientes colaborativos para construção de ontologias de domínio com conhecimento visual**. Disponível em:
<http://www.lume.ufrgs.br/handle/10183/47562>. Acesso em: 29 de julho de 2014.

TAKEDA, S. et al. Irregular Trend Finder: visualization tool for analyzing time-series big data. **IEEE Symposium on Visual Analytics Science and Technology**, 2012.

WILKES, S. Some Impacts of "Big Data" on Usability Practice. **Communication Design Quarterly Review**, v. 13, n. 2, p. 25, Jun. 2012.

SÍNDROME DE GABRIELA” EM UMA COMUNIDADE RURAL DE JOÃO PESSOA: RESISTÊNCIA E ACEITAÇÃO DAS TECNOLOGIAS DE INFORMAÇÃO E COMUNICAÇÃO NO PROCESSO DE ALFABETIZAÇÃO E INCLUSÃO DIGITAL

SÍNDROME DE GABRIELA” IN A RURAL COMMUNITY OF JOAO PESSOA: RESISTANCE AND UPTAKE OF INFORMATION AND COMMUNICATION TECHNOLOGIES IN THE PROCESS OF LITERACY AND DIGITAL INCLUSION

Raíssa Carneiro de Brito

Resumo: Diante da experiência no projeto de alfabetização de inclusão digital Luz do Saber em uma comunidade rural de João Pessoa, foi percebida a existência de um fenômeno: em uma localidade com aproximadamente 400 famílias, atualmente há 28 alunos matriculados, sendo apenas 13 alunos o que frequentam assiduamente. A dificuldade do sujeito em vivenciar o uso das tecnologias denomina-se, aqui, “síndrome de Gabriela”, termo que passou a ser utilizado popularmente após a novela de Gabriela, Cravo e Canela, de Jorge Amado, tendo como protagonista a Gabriela, mulher simples que não se adaptava aos costumes da vida urbana e cuja trilha sonora, música de Gal Costa, diz “... Eu nasci assim, eu sou mesmo assim, vou ser sempre assim, do meu jeito assim, Gabriela...”. O projeto Luz do Saber na comunidade rural Engenho Velho, em João Pessoa, no estado da Paraíba, é uma oportunidade de letramento direcionada às pessoas que lá residem através do computador com atividades interativas em jogos didáticos, em que o aluno é alfabetizado. A análise do fenômeno encontrado nesta comunidade é amparada pela teoria das Limitações Digitais, desenvolvida por Bellini et al. (2010), que apresenta um modelo tridimensional que aborda limitações de acesso, limitações cognitivo-informacionais e limitações comportamentais. A pesquisa será conduzida através da construção do referencial teórico sobre o tema abordado e contará com o apoio da técnica de coleta de dados do Grupo Focal, cuja discussão abordará questões orientadas no modelo TAM (modelo de aceitação de tecnologia), através da análise de conteúdo das opiniões expressas. Referenciando tal procedimento, serão construídas recomendações para que outros projetos sociais com características semelhantes possam ser conduzidos a fim de alcançar seus objetivos e assim alcançar maior aceitação da população.

Palavras-chaves: Projeto social; Comunidade Rural; Limitações Digitais; Grupo Focal; Síndrome de Gabriela.

Abstract: From the experience in the project of digital inclusion littering called by “Luz do Saber” (Light of Knowledge) in a rural community at Joao Pessoa, it emerged the searching about a phenomenon present in this community, where is observed that from approximately 400 families, we have now 28 students enrolled, but only 13 attending assiduously. This challenge of changing is called, in this research, by “Gabriela’s Syndrome”, term that starts to be used commonly after the soap opera “Gabriela cravo e canela”, written by Jorge Amado, in which Gabriela was the protagonist. She was a simple woman who did not get used to the urban life. The soundtrack says “... I was born on this way. I am just like this way. I will be always like this way that is my own way, Gabriela...”. The project “Luz do Saber”, at the rural community in Engenho Velho, Joao Pessoa, is an opportunity for people who live there being able to literate through interactive activities like didactics games on the computer that help the student to became literate. However, despite all these resources that the program offers, the project has a strong resistance to the population changing of the rural community. Helped by the theory of Digital Limitations, developed by Bellini et al 2010, which presents a tridimensional model of Access Limitation, Informational Cognitive Limitation and Behavioral Limitation, The research will be conducted through the construction of theoretical

references about the project's subject and will count on Focal Group's support of data collect technic, where it will be discussed questions based on TAM's model (model of acceptance technology). From analysis of public opinions content, it will be produced advices that other similar social projects can take in order to target their goals and a higher popular acceptance.

Key-words: Social Project; Rural Community; Digital Limitations; Gabriela Syndrome; Focus Group.

1 INTRODUÇÃO

Presencia-se na sociedade da informação grandes desigualdades sociais, principalmente em países em desenvolvimento como é o caso do Brasil e de outros países da América latina. Na contraposição de tantos avanços nos âmbitos das Tecnologia de Informação e Comunicação (TICs), vivencia-se problemas de desemprego, educação precária com grandes índices de analfabetismo que geram a falta de condições básicas para o exercício da cidadania.

Na contemporaneidade as relações sociais estão cada vez mais sendo intermediadas por equipamentos eletrônicos. A sociedade avança tecnologicamente criando uma lacuna com os aspectos sociais, sem procurar entender como essas inovações impactam na vida dos indivíduos, nos aspectos cognitivos e comportamentais, sendo um desafio atual alcançar a democratização das TICs. Para alcançar a referida democratização, se faz necessário compreender como são as diferentes formas de reações da sociedade, sendo necessários estudos que investiguem as reações individuais e de grupos pertencentes ao conjunto dos incluídos e excluídos digitalmente, procurando compreender como é o processo de aprendizagem dos indivíduos mediada pelas TICs e propondo mecanismos que corroborem e estimulem a efetiva aceitação e uso a proposta de desenvolvimento das competências informacionais necessárias para o pleno exercício da cidadania.

Acompanhando a história da humanidade, houve períodos onde a informação era restrita, poucos podiam ter acesso a ela. Atualmente a informação está disponível para todos e a todo o momento, através das TICs. Porém, vemos que diante de toda essa facilidade em acessar a informação, grande parcela da sociedade, principalmente em países com grandes níveis de desigualdades sociais como é o caso do Brasil, uma grande parcela da sociedade tem dificuldade em acessar essas informações. Tal dificuldade se dá por diversos fatores, entre eles, o baixo nível de escolaridades da população, pois o Brasil ainda registra números elevados de pessoas analfabetas e muitas que são consideradas alfabetizadas não conseguem compreender o que leem ou que escrevem (analfabetismo funcional). Outro fator que dificulta o acesso e uso das TICs é a falta de motivação ou interesse das pessoas, que algumas vezes dispõem de oportunidades para aprenderem e a se capacitarem em algo que tenham

dificuldades como, por exemplo, melhorar as habilidades com o computador, mas mesmo assim muitas não se sentem motivados e permanecem em suas “zona de conforto” sem superarem a resistência a mudança.

Para tal, faz-se necessário sair da “zona de conforto” e passar a ter atitudes em prol de desenvolver essas habilidades, sendo necessário romper a dificuldade de mudança de comportamento, que aqui no Brasil denomina-se popularmente como “síndrome de Gabriela”, devido a uma telenovela exibida em 1975, onde retratava a vida de uma jovem simples do sertão, cuja trilha sonora da personagem dizia, “... eu nasci assim, eu sou mesmo assim, vou ser sempre assim, Gabriela...”. O termo é usado para expressar quando o indivíduo assume uma postura de que não vai modificar ou adotar determinada atitude, simplesmente porque não tem nenhum interesse para assumir tal mudança ou postura.

2 SOCIEDADE DA INFORMAÇÃO

O termo sociedade da informação é bastante utilizado para retratar a maneira como a sociedade atual lida com a informação, sendo possível, então, fazer uma reflexão do que esse termo representa na contemporaneidade. A expressão “sociedade da informação” é utilizada atualmente para se referir a “sociedade pós-industrial” e retratar o novo paradigma técnico-econômico desse último século. A informação se tornou um insumo básico e necessário desta nova conjuntura social, sendo ampliada sua importância e acesso pelos avanços tecnológicos na microeletrônica e telecomunicação. Essas mudanças ocorrem em uma velocidade cada vez mais rápida.

Outra definição para o termo sociedade da informação refere-se à enorme proliferação da informação, estimulada pelo aproveitamento da microelectrônica e pelas primeiras manifestações do seu potencial impacto social e econômico. Em contrapartida, o conceito de sociedade de aprendizagem transporta em si a concepção embrionária do modo de vida moderno, fortemente recomendado devido à crescente integração das tecnologias de informação e comunicação e ao receio de que a globalização possa prejudicar a competitividade. A sociedade do conhecimento distingue-se pela maneira como encara a mudança estrutural da economia a longo prazo. Segundo esta visão, a produção, divulgação e utilização do conhecimento irão desempenhar um papel ainda mais importante na criação e aproveitamento da riqueza (Lindley, 2000, p36).

Conforme Shanon (1949), no livro A teoria matemática da comunicação, trata a informação como sendo tudo aquilo que reduz a incerteza. Nesta perspectiva, o termo sociedade da Informação não poderia ser aplicado adequadamente, uma vez que temos uma

quantidade imensurável de dados denominados como informação, mas que, isoladamente, não têm significado e não produzem compreensão.

Castells (1999) em seu livro intitulado *A Sociedade em rede*, afirma que a revolução tecnológica iniciou a partir da Arpanet, surgindo a partir daí uma nova ordem social: a informacional, em que a informação passou a ter valor de capital, considerando os discursos sobre a revolução da tecnologia de informação e comunicação como exagerados, proféticos e ideológicos. No entanto, este mesmo autor considera este evento histórico da mesma importância da Revolução Industrial do Século XVIII, induzindo um padrão de descontinuidade nas bases materiais da economia, sociedade e cultura. A tecnologia da informação é para esta revolução o que as novas fontes de energia foram para as Revoluções Industriais sucessivas. (CASTELLS, 1999).

Para Toffler (2001, p. 37), a inovação tecnológica não só propõe mudanças em máquinas, mas alteram também o meio ambiente intelectual do homem (como ele pensa e encara o mundo), propondo soluções inéditas para problemas filosóficos e até sociais. A utilização da informação para responder a uma questão, solucionar um problema, tomar uma decisão, negociar uma posição ou dar sentido a uma situação. Sendo, a informação é fabricada por indivíduos a partir de sua experiência passada e de acordo com as exigências de determinada situação na qual a informação deve ser usada. (CHOO, 2003)

Dentro desta mesma linha de pensamento, pode ser entendido que os avanços tecnológicos e as inovações para disponibilizar a informação não alcançam toda a sociedade, ficando boa parte da humanidade sem ter acesso a esses recursos tecnológicos sendo alguns desses motivos: falta de recursos econômicos para adquiri-los, falta de instruções educacionais para se entender ao que se tem acesso, ou limitações cognitivas.

3 COGNIÇÃO

Por Cognição entende-se um tipo específico de representação dos objetos e fatos ou qualquer tipo de representação da informação proveniente do meio. O dicionário de psicologia Mesquita e Duarte (1996 p. 45) define a cognição como um conceito utilizado para designar comportamentos, pensamentos, atitudes e crenças, conscientes dos indivíduos.

Para Piaget (1986) existem estruturas inatas do sujeito, que se organizam a partir das experiências do meio ambiente, resultando os processos perceptivos. Desta forma, a cognição tem estreita relação e é construída através do meio e todo comportamento é influência do meio. De acordo com Sternberg (2000) um dos teóricos considerados neopiagetianos, ressalta

a cognição como o processar e coordenar elementos que possibilitam a diferenciação de informação na determinação de objetivos secundários para atingir uma meta.

Neste cenário, Jones (1988) afirma que a metacognição exerce influência sobre a motivação, isso se dá, devido ao fato da metacognição possibilitar, o controle e gestão dos próprios processos cognitivos, dando-lhe a noção da responsabilidade pelo seu desempenho, gerando assim, confiança nas suas próprias capacidades (Morais & Valente, 1991). Para Weinert (1987) as metacognições podem ser consideradas cognições de segunda ordem: pensamentos sobre pensamentos, conhecimentos sobre conhecimentos, reflexões sobre ações. Livingston (1997 p.4) diferencia a metacognição da cognição como a metacognição sendo o conhecimento e a cognição sobre um fenômeno cognitivo.

Como podemos observar nas definições de cognição e metacognição, alguns estudiosos na área afirmam que é difícil, em certo ponto, definir esses campos do conhecimento e distinguir suas diferenças, podendo ser compreendido como complemento um do outro. A partir da década de 1970 muitos pesquisadores intensificaram suas pesquisas sobre aprendizagem e as capacidades cognitivas, bem como, o que tange a cognição no sentido da memória, leitura e compreensão de texto. (BROWN, 1987)

3.1 Autodeterminação

Taxonomia da motivação humana dividiu-se em três grupos: 1. Sem motivação ou desmotivado, 2. Motivação extrínseca (de fora para dentro) este se subdivide em quatro tipos diferentes: regulação externa(onde a motivação é em troca de alguma recompensa), regulação introjetada (onde o indivíduo consegue fazer uma avaliação das consequências de agir ou não de determinada maneira, sendo a decisão influenciada por fatores externos), regulação identificada(nesse caso a uma interiorização da motivação sendo estimulada por algum objetivo externo), regulação integrada(há um equilíbrio entre os aspectos comportamentais, objetivos e valores que tem como objetivo maior a realização pessoal,) e por fim 3. Motivação intrínseca, a motivação parte do indivíduo sem necessitar de um estímulo externo, existe nesse caso uma satisfação e interesse em realizar determinada atividade ou adquirir determinado comportamento.

Alguns pesquisadores voltam suas pesquisas para compreender a influência da motivação na aprendizagem, essas pesquisas mostram que indivíduos intrinsecamente motivados, tendem a ter melhores desempenhos nas atividades que realizam e tendem a serem mais propícios a mudarem de comportamento em prol da realização de um objetivo. (SIQUEIRA & WECHSLER, 2006).

4 PROCEDIMENTOS METODOLÓGICOS

O presente trabalho pretende desenvolver a metodologia desta pesquisa a partir da construção de uma base conceitual sobre resistência à mudança de atitude, aprendizagem social e Limitações digitais, procurando compreender o que a literatura científica pode colaborar com essas reflexões. Realizar grupo focal, a fim de identificar o que motiva os alunos do programa Luz do Saber na superação da “síndrome de Gabriela”, afim de:

- Identificar quais fatores que influenciam a aceitação dos alunos pelo programa Luz do Saber, através do modelo TAM, que é indicado para estudar comportamento de aprendizagem e aceitação de tecnologia.
- Analisar os dados obtidos
- Construir recomendações para a efetivação de mudanças de comportamentos que desenvolvam capacidades informacionais e promova a efetivação da aprendizagem em projetos sociais, tais como o Luz do Saber.

REFERÊNCIAS

- BELLINI, C.G.P.; GIEBELEN, E.; CASALI, R.R.B. **Limitações digitais.** Informação & Sociedade, v. 20, n. 2, 2010, p. 25-35.
- BROWN, A. L. Metacognition, executive control, self-regulation, and other more mysterious mechanisms. In WEINERT, F. E. & KLUWE, R. (Org.). **Metacognition, motivation, and understanding.** 1987.
- CASTTELLS, M., **Society in Network.** São Paulo, 1999.
- CHOO, C. W. **A organização do conhecimento:** como as organizações usam a informação para criar significado, construir conhecimento e tomar decisões. São Paulo: Senac, 2003.
- JONES, B. F. Text learning strategy instruction: Guidelines from theory and practice. In WEINSTEIN, C. E.; GOETZ, E. T.; ALEXANDER, P. A. (Org.), **Learning and study strategies:** Issues in assessment, instruction, and evaluation. 233-260. N. Y. 1988.
- LINDLEY, R. Economias baseadas no conhecimento – o debate europeu sobre o emprego num novo contexto. In: RODRIGUES, M. J. (Org.). **Para uma Europa da Inovação e do conhecimento** – Emprego, reformas económicas e coesão social. Oeiras: Celta. p. 33-78. 2000
- LIVINGSTON, J. A. **Metacognition: An Overview,** 1997
- MESQUITA, R.; DUARTE, F. **Dicionário De Psicologia.** Ed. Plátano: São Paulo. 213 P, 1996.

MORAIS, M. M.; Valente, M. O. Pensar sobre o pensar: Ensino de estratégias metacognitivas para recuperação de alunos com dificuldades na compreensão da leitura na disciplina de língua portuguesa. **Revista de Educação**, v. 2, n. 1, p. 35-56, 1991.

PIAGET, J. **O nascimento da inteligência da criança**. Editora Crítica: São Paulo, 1986.

SHANNON, C. E.; WEAVER, W. **The Mathematical Theory of Communication**. Urbana: University of Illinois Press. 1949

SIQUEIRA, L. G. G.; WECHSLER, S. M. Motivação para a aprendizagem escolar: possibilidades de medida. **Avaliação Psicológica**, n. 5, v. 1, p. 21-3, 2006.

STERNBERG, R. J. **Psicologia Cognitiva**. M. R. B. Osório, Trad. Porto Alegre: ArtMed, 2000.

TOFFLER, A. **A terceira onda**, 26. ed. Rio de Janeiro: Record, 2001.

WEINERT, F. E. Metacognition and motivation as determinants of effective learning and understanding. In WEINERT, F. E.; KLUWE, R. H. (Ed.), **Metacognition, Motivation and Understanding**. Hillsdale: LEA, 1987.

A WEB 2.0 COMO CANAL DE COMUNICAÇÃO ENTRE A BIBLIOTECA E OS USUÁRIOS

THE WEB 2.0 AS COMMUNICATION CHANNEL BETWEEN THE LIBRARY AND USERS

Nivea Camara Rocha de Souza
Ana Paula de Oliveira Villalobos

Resumo: Este artigo objetiva analisar a importância da apropriação das ferramentas da Web 2.0 pelas bibliotecas da cidade de Salvador para divulgação de seus serviços, acervo e atividades, com vistas a potencializar a comunicação com os usuários, promover o acesso a informação e atrair novos visitantes. A Web 2.0, também conhecida como Web Social, viabiliza várias interfaces digitais que podem ser utilizadas pelas bibliotecas como forma de mediação e interação com os usuários, destacando que todos os tipos de bibliotecas podem empregar essas ferramentas. Em relação ao objetivo a pesquisa se caracteriza como descritiva, relativamente ao método trata-se de um estudo de caso, o instrumento de coleta de dados adotado será o questionário e a abordagem será de natureza quali-quantitativa. Após a análise dos dados os resultados serão confrontados com o referencial teórico, na perspectiva de investigar a importância das ferramentas da Web 2.0 como um eficiente canal de informação e comunicação entre as bibliotecas e os usuários.

Palavras-chave: Web 2.0. Acesso a informação. Biblioteca 2.0. Web Social.

Abstract: This article intended to analyze the importance of ownership of Web 2.0 tools by libraries in the city of Salvador to advertise their services, assets and activities, with a view to strengthening the communication with users, promote access to information and attract new visitors. Web 2.0, also known as Social Web, enables various digital interfaces that can be used by libraries as a form of mediation and interaction with users, stressing that all types of libraries can employ these tools. Regarding the objective of research is characterized as descriptive, concerning the method it is a case study, the data collection instrument will be adopted in the questionnaire and the approach will be qualitative and quantitative. After data analysis, the results will be compared with the theoretical framework, the aim of investigating the importance of Web 2.0 tools as an efficient channel of information and communication between libraries and users.

Keywords: Web 2.0. Information access. Social media. Library 2.0. Social web.

1 INTRODUÇÃO

Atualmente, é visível o poder da comunicação e divulgação da informação na internet e como ela fortalece as imagens dos órgãos públicos e empresas. Se esse mecanismo for utilizado pela biblioteca poderá haver uma resposta positiva por parte dos usuários, os quais esperam saber o que está acontecendo na sua unidade de pesquisa sem precisar visitar o espaço físico para obter essas informações. Essas novas tecnologias promovem um avanço nos canais de mediação, aumentando o alcance da difusão da informação e modificando os processos de comunicação, rompendo as barreiras de espaço e tempo.

Por meio dessas inovações surgiu a web 2.0, uma ferramenta que propicia um ambiente colaborativo entre as organizações e seus frequentadores, independente da localização.

A partir do conceito de biblioteca e web 2.0, a pesquisa objetiva realizar um estudo acerca da potencialidade da web 2.0 como um recurso utilizado pelas bibliotecas da cidade de Salvador para divulgação de serviços, acervo e atividades e também uma possibilidade de melhorar a comunicação, o relacionamento da biblioteca com os usuários e atrair novos visitantes. Além, especialmente, de promover o acesso a informação.

Já que a missão da biblioteca é disponibilizar o acesso à informação e levar o conhecimento, as criações de páginas nas redes sociais, blogs e sites multimídias seriam uma eficiente maneira de levar informações e promover uma maior interação com os usuários e futuros frequentadores. Apesar desta importância estar sendo discutida, ainda é escasso o número de bibliotecas em Salvador que utilizam esse recurso.

A pesquisa pretende evidenciar a importância das bibliotecas utilizarem a web 2.0 para desenvolverem um ambiente informacional mais interativo e colaborativo. Conforme Arroyo Vázquez e Merlo Vega,

A arquitetura de participação é um conceito fundamental na web 2.0 e consiste em envolver os usuários e obter sua própria colaboração de modo que eles, os próprios internautas que vão dar valor ao serviço, criando seus conteúdos e organizando os mesmos. (2007, p.2, tradução nossa)

As várias ferramentas da web 2.0 se destacam como uma nova via de comunicação e mediação pela biblioteca com seus frequentadores e futuros visitantes, sua inserção em um contexto digital e divulgação do trabalho realizado pela unidade. É importante destacar que todos os tipos de biblioteca podem empregar as ferramentas de web 2.0, não se restringindo às bibliotecas públicas.

Esta investigação é relevante porque comprova a importância do uso dessas ferramentas de interação pelas bibliotecas de Salvador, destacando as novas ferramentas da web 2.0 como potencializadoras no acesso a informação, na divulgação do acervo, atividades e serviços. Pode ser evidenciado também o marketing indireto, desempenhado quando essas informações ou notícias são compartilhadas pelos internautas nos seus perfis nas redes sociais.

2 SURGIMENTO DA WEB 2.0

O conceito de Web 2.0 surge em 2004 numa conferência promovida por Tim O'Reilly e origina a idéia de uma plataforma web construída a partir de uma coletividade, um ambiente

de co-participação e interação entre leitor e o desenvolvedor. Um ambiente mais dinâmico e flexível. Conforme, relatado por Tim O'Reilly,

O conceito de Web 2.0 começou com uma sessão de brainstorming conferência entre O'Reilly e Internacional MediaLive. Dale Dougherty, pioneiro web VP e O'Reilly, notou que longe de ter falhas, a web era mais importante do que nunca, com novas aplicações e sites surgindo com surpreendente regularidade. Além do mais, as empresas que haviam sobrevivido à colapso parecia ter algumas coisas em comum. (2007, p.2, tradução nossa)

A partir do surgimento da “web social” denominada também de web 2.0, os aplicativos e sites criados tem por objetivo permitir uma maior colaboração e participação da sociedade na busca e na disseminação da informação. Para Margaix Arnal (2007, p.99, tradução nossa), “A verdadeira revolução 2.0 é o novo conceito que temos dos usuários, para que eles interagem uns com os outros e com o pessoal da biblioteca”. Pode ser evidenciado quanto às ferramentas da Web 2.0 são dinâmicas e possibilitam novas formas de comunicar a informação.

A utilização dos aplicativos criados com a web 2.0 aumenta os canais de comunicação que podem ser explorados pelas bibliotecas para informar, divulgando as informações e novidades da biblioteca; interagir, trocar informações e idéias, e promover, compartilhar informação para a comunidade a que atende. No entanto, este mecanismo ainda é pouco explorado no âmbito das bibliotecas. Segundo Margaix Arnal (2007, p.101, tradução nossa) “Deve-se notar que a maioria dos pioneiros da biblioteca 2.0 executa o seu trabalho em bibliotecas públicas e são estes que estão aplicando as suas abordagens mais rapidamente e desenvolvendo o conceito”. Dessa maneira, a utilização das ferramentas da web 2.0 pela biblioteca é resultante do desenvolvimento de novas tecnologias e, consequentemente, da evolução Web 1.0 para a Web 2.0, ou Web Social.

3 BIBLIOTECA 2.0

Com os avanços tecnológicos e aumento do número de pessoas que possuem computador e *smartphones* se faz necessário a presença das bibliotecas na esfera virtual, para promover os seus serviços e possibilitar um canal de comunicação com os seus utilizadores. Conforme Alvin:

A Biblioteca 2.0 é uma atitude que orienta informação para determinados utilizadores, incorpora novas ferramentas e serviços, e constrói, com os utilizadores, conteúdos significativos. Com esta filosofia, não se abandonam os utilizadores habituais, acrescenta-se a possibilidade de mais participação aos que existem, e faz-se uma aproximação a uma população virtual mais alargada. Convém reconhecer que as bibliotecas evoluem lentamente e que será preciso trabalhar com os profissionais (2007, p. 3).

A partir do surgimento dessa nova proposta no âmbito da Ciência da Informação o conceito de biblioteca se amplia, incorporando novas ferramentas digitais, buscando disponibilizar informações e uma aproximação com seus usuários no ambiente virtual.

Com o surgimento da Biblioteca 2.0 e a utilização da Web Social como um novo meio de interação pelos centros de informação, os usuários podem opinar, deixar sugestões e perguntar no âmbito virtual através do site e páginas nas redes sociais. Essa facilidade no contato objetiva uma comunicação mais dinâmica e rápida. Cohen (2006, tradução nossa) afirma que criará sites abertos que permitam aos usuários se juntarem aos bibliotecários para contribuirem com conteúdo, a fim de melhorar a sua experiência de aprendizagem e prestar assistência a eles.

De acordo com Vieira *et. al* (2003) as bibliotecas ao utilizarem as várias ferramentas da Web 2.0 irão ampliar os meios de mediação com seus usuários e novos visitantes. As funcionalidades contidas nas ferramentas da Web Social oferecem múltiplos recursos para ajudar as bibliotecas a interagir e se comunicar com os seus usuários. Lopes (2006) acrescenta que

As TICs oferecem possibilidades para formas inovadoras de mediação. O bibliotecário como mediador poderá atuar na orientação de seus respectivos usos, nas atividades de acesso a mecanismos de busca, na seleção, análise, síntese de conteúdos de informação, no desenvolvimento de sistemas especialistas para responder questões de referência, na capacitação através de instruções bibliográficas, entre outras. (LOPES; SILVA, 2006)

A importância do desenvolvimento dessas novas ferramentas inovadoras de mediação é proporcionar a concepção de que a biblioteca não é um ambiente estático, mas em constante crescimento. As unidades de informação estão atentas na busca por aprimorar seus serviços e canais de comunicação com seus usuários, disponibilizando online informações significativas e mantendo uma via colaborativa e interativa com seus internautas.

A biblioteca 2.0 é uma comunidade virtual voltada para seus usuários, surge da ligação dos conceitos existentes entre os termos biblioteca e Web 2.0 (MANESS, 2007) e na visão de Arroyo Vázquez e Merlo Vega (2007) estas são algumas das ferramentas que poderiam ser utilizadas pelas bibliotecas 2.0, a exemplo: dos *podcasts*, os blogs, os espaços *wiki*, os sítios web para o compartilhamento de imagens (*Instagram / Flickr*), de vídeos (*YouTube / Dailymotion*), como também a utilização das redes sociais (*MySpace / Orkut / Google+ / Facebook / Twitter*) e as ferramentas de bookmark social (*Del.icio.us / Pinterest / Tumblr / Foursquare*).

Portanto, conforme elucidado por diversos autores, as bibliotecas ao utilizarem as ferramentas da web 2.0 ampliam seus meios de mediação com seus frequentadores e novos

visitantes. E para comprovar está eficácia nas bibliotecas de Salvador a seguir será explanada a metodologia que será utilizada.

4. INSTRUMENTOS METODOLÓGICOS

Trata-se de uma pesquisa que adota o método de estudo de caso, utilizando o questionário como um instrumento de pesquisa para a coleta de dados. A abordagem adotada será de natureza quali-quantitativa, por ser considerada a mais adequada para a interpretação dos dados.

A realidade empírica vai ser confrontada por uma pesquisa sobre quais bibliotecas utilizam ferramentas da Web 2.0 em Salvador, quais são os resultados obtidos após a criação deste canal de comunicação e se está havendo êxito no uso desse canal como meio de mediação e divulgação de informações e quais canais de comunicação estão tendo um maior número de usuários interagindo e colaborando.

Por fim, os dados serão analisados e confrontados com o referencial teórico e inferidas as conclusões e dadas às sugestões, afim que as bibliotecas tenham conhecimento dos benefícios dessas ferramentas inovadoras como meio de interação e divulgação.

5 CONSIDERAÇÕES FINAIS

Este trabalho busca refletir sobre o papel da Web 2.0 como um meio de comunicação utilizado pelas bibliotecas, tomando como referencia estudiosos da área da Ciência da Informação, no intuito de comprovar que as ferramentas que compõem a Web Social podem ser utilizadas em qualquer tipo de biblioteca, não se limitando a biblioteca pública, como tem sido a práxis

Após a análise do resultado das pesquisas empíricas, o qual revelará quais são as bibliotecas em Salvador que utilizam as ferramentas da web 2.0 e através de quais interfaces se concretiza essa interação e quais são os resultados obtidos após a criação deste canal de comunicação.

A partir deste estudo de cunho inovador, que poderá motivar a criação de páginas nas redes sociais, blogs e outras ferramentas da web social para aprimorar as possibilidades de interação e comunicação da biblioteca e seus frequentadores e atuar como um eficiente canal de mediação.

REFERÊNCIAS

ALVIM, Luísa. Blogues e bibliotecas: construir redes na Web 2.0. **Cadernos BAD**. Lisboa, n. 1. p. 38-74, 2007.

- ARROYO VÁZQUEZ, N; MERLO VEGA, J. A. La biblioteca como usuaria de la web 2.0. In: JORNADAS ESPAÑOLAS DE DOCUMENTACIÓN, 10., 2007, Santiago de Compostela. *Anais...* Santiago de Compostela: FESABID, 2007. p. 1-11. Disponível em: <http://eprints.rclis.org/9523/1/Arroyo%26Merlo_FESABID07.pdf>. Acesso em: 20 set. 2014
- COHEN, Laura. **A Librarian's 2.0 Manifesto.** 2006. Disponível em: <<http://katrinamalone23.blogspot.com.br/2013/01/a-librarians-20-manifesto-from-laura.html>>. Acesso em: 20 jun. 2014.
- LOPES, M. I.; SILVA, E. L. As bibliotecas universitárias e a mediação da informação na comunicação científica. In: SEMINÁRIO NACIONAL DE BIBLIOTECAS
- MANESS, J. M. Teoria da Biblioteca 2.0: Web 2.0 e suas implicações para as bibliotecas. Traduzido por Geysa Flávia Câmara de Lima Nascimento e Gustavo Henrique do Nascimento Neto. **Informação & Sociedade:** Estudos, v. 17, n 1, 2007. Disponível em: <<http://periodicos.ufpb.br/ojs2/index.php/ies/article/view/83>>. Acesso em: 20 jun. 2014.
- MARGAIX ARNAL, Dídac. Conceptos de web 2.0 y biblioteca 2.0: origen, definiciones y retos para las bibliotecas actuales. In: **El profesional de la información.** EPI SCP, Barcelona, Spain , 2007. p. 95-106. Disponível em: <<http://eprints.rclis.org/9521/>>. Acesso em: 20 jun. 2014.
- O'REILLY, Tim. **What is Web 2.0?** Design patterns and business models for the next generation of software. 2007. Disponível em: <<http://mpra.ub.uni-muenchen.de/4578/1/>>. Acesso em: 20 jun. 2014.
- UNIVERSITÁRIAS, 14, 2006, Salvador. *Anais...* Salvador: UFBA, 2006. CD-ROM.
- VIEIRA, David Vernon; BAPTISTA, Sofia Galvão; CUEVAS CERVERO, Aurora. Adoção da Web 2.0 em bibliotecas de universidades públicas espanholas: perspectivas de interação do bibliotecário com as redes sociais – relato de pesquisa. **Perspectivas em Ciência da Informação**, v. 18, n. 2, p. 167-181, 2013. Disponível em: <<http://dx.doi.org/10.1590/S1413-99362013000200011>>. Acesso em: 20 jun. 2014.