

Dados do Projeto de Pesquisa	
Título do Projeto de Pesquisa:	Institucionalização Cognitiva e Social da Pesquisa Científica no Brasil: extração automática e análise de citação para identificação de redes de colaboração
Grande área/área segundo o CNPq:	Ciências Sociais Aplicadas / Ciência da Informação
Grupo de Pesquisa vinculado ao projeto:	GRUPO MAPA - Memória, Acervos e Patrimônio http://dgp.cnpq.br/dgp/espelhogrupo/2813669542970174
Linha de pesquisa do grupo de pesquisa vinculado ao projeto:	Representação de Acervos em Ambientes Informacionais
Categoria do projeto:	() projeto em andamento, já cadastrado na PRPI () projeto não iniciado, mas aprovado previamente (x) projeto novo, ainda não avaliado
Palavras-chave:	Análise Automática de Textos; Análises Métricas; Expressões Regulares; Institucionalização Científica.

1. INTRODUÇÃO

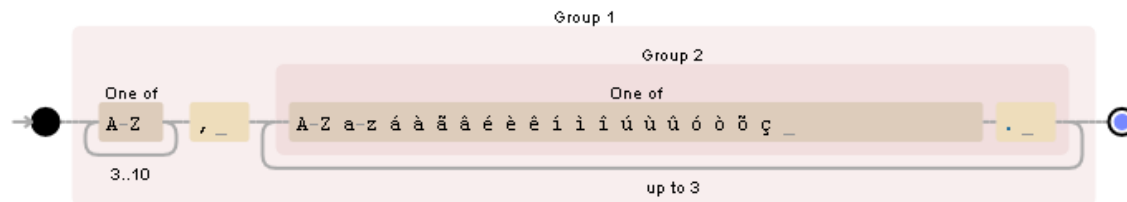
Os estudos voltados para as análises métricas (principalmente a altmetria, bibliometria, cientometria e webometria) são de extrema importância para a Ciência da Informação. Isto pode ser percebido, por exemplo, pela existência não apenas de eventos científicos voltados especificamente para a área (Encontro Brasileiro de Bibliometria e Cientometria, que em 2016 teve sua 5ª edição), mas também grupo de trabalho voltado para discussões que abordam prioritariamente a temática no mais importante evento científico da Ciência da Informação (GT7 - Produção e Comunicação da Informação em Ciência, Tecnologia & Inovação - do Enancib).

Uma outra área de estudos que tem se expandido junto com as análises métricas são as voltadas para as teorias sócio cognitivas na Ciência da Informação, e dentro destas uma crescente importância de natureza teórico-metodológica nos estudos de análise de domínio, principalmente a partir das contribuições de Hjørland (2002, 2004), entre outros, assim como os estudos da institucionalização científica, tendo como principal autor Whitley (1974, 2000).

Este projeto encontra-se na intercessão destes campos, tendo ainda a contribuição da análise sintática-semântica da Ciência da Computação, a partir das expressões regulares (HABIBI, 2003; THE OPEN GROUP, 1997), para buscar dentro da produção acadêmica as relações sociais que fundamentam os campos da ciência. Este projeto é parte integrante do projeto **MCTI/CNPQ N° 28/2018 - Universal: Web Semântica e Web Pragmática: estudo de representação e recuperação de informação em acervos informacionais digitais**, funcionando como instrumentalização tecnológica para ações realizadas no mesmo.

Durante pesquisas anteriores, foram identificados alguns contextos que são passíveis de representação em ambientes informacionais digitais, descritos a seguir.

Figura 2 – Representação Gráfica da Expressão Regular para Autor em Referência ABNT



Fonte: Elaborado pelo autor, 2017.

A segunda etapa consiste na codificação dos elementos anteriormente representados em código que permita sua execução de forma isolada. Esta codificação ocorrerá, preferencialmente, na linguagem Java, pois uma IDE para programação nesta linguagem vem instalada em alguns computadores da UFCA, mas pode ser em outra linguagem que os bolsistas se sintam mais confortáveis em programar ou que permita maior integração com outros SRI.

Após criado o código, este será testado tanto isoladamente como de forma integrada com outros SRI, como a Brapci¹ ou o Portal de Periódicos² da UFCA, por exemplo, para avaliar seu funcionamento e possibilidade de uso conjunto com outras plataformas para gerar metadados que permitirão novas perspectivas de pesquisa.

4. PRINCIPAIS CONTRIBUIÇÕES CIENTÍFICAS, TECNOLÓGICAS OU DE INOVAÇÃO DO PROJETO

Com isto, este projeto possibilita contribuições em quatro frentes: acadêmica, social, pessoal e tecnológica. Do ponto de vista acadêmico, este trabalho possibilita o aprofundamento dos estudos e usos das Expressões Regulares, item pouco usado devido a ser considerado de alta complexidade; com o uso destas expressões, é possível avançar nos estudos do reconhecimento de padrões textuais e, mais especificamente, de referências bibliográficas ou outros elementos de textos acadêmico-científicos que permitam análises automáticas ou semiautomáticas de textos, abrindo com isto novas possibilidades de pesquisa. Este projeto, especificamente, é apenas a etapa inicial de um projeto maior, que envolve a análise de repositórios de trabalhos acadêmicos e outras bases de dados que agregam informações referentes a esta produção.

Do ponto de vista social, este projeto retorna para a comunidade científica a possibilidade de avaliar relações sociais e de influência textual das diferentes áreas do conhecimento, contribuindo com isto para novas visões da comunidade e de influências de teorias, paradigmas e autores que muitas vezes passam despercebidas no dia a dia.

Sob a perspectiva pessoal, este projeto é fruto da conjunção de conhecimentos e habilidades adquiridas pelo pesquisador durante sua formação mas que, devido às

¹ <http://www.brapci.ufpr.br/brapci/>

² <http://periodicos.ufca.edu.br/ojs/>

características das disciplinas lecionadas, não tem grande permeabilidade nas atividades desenvolvidas na graduação. Com isto, este projeto possibilita o aprendizado dos discentes de conteúdos que, mesmo sendo de extrema relevância, não veriam nas disciplinas que compõem a grade curricular dos cursos da área de conhecimento.

Sob a última perspectiva, tecnológica, este trabalho justifica-se pela abordagem não costumeira nas Ciências Sociais Aplicadas, que gerará um produto que pode ser utilizado tanto em situações de análise de conteúdo, assim como nas perspectivas dos estudos de análise do domínio e da institucionalização social.

5. CRONOGRAMA DE EXECUÇÃO DO PROJETO

As atividades desenvolvidas neste projeto serão executadas de forma concomitante ao projeto MCTI/CNPQ N° 28/2018 - Universal: Web Semântica e Web Pragmática: estudo de representação e recuperação de informação em acervos informacionais digitais, representados no Quadro 1, cujos objetivos, fundamentação e justificativas foram apresentados neste documento.

Quadro 1 – Lista de Atividades

ATIVIDADE \ MÊS	01	02	03	04	05	06	07	08	09	10	11	12
Providências preliminares	X	X										
Levantamento Bibliográfico		X	X	X	X	X	X	X	X	X	X	X
Definição das etapas metodológica e planejamento com a equipe			X									
Mapeamento das fontes de pesquisa e coleta dos dados				X	X	X	X	X	X	X	X	X
Identificação de Contextos a Representar							X	X	X	X	X	X
Modelagem dos Contextos para a Base de Dados										X	X	X
Redação de trabalhos científicos para publicação							X	X	X	X	X	X
ATIVIDADE \ MÊS	13	14	15	16	16	18	19	20	21	22	23	24
Identificação de Contextos a Representar	X	X	X	X	X	X						
Modelagem dos Contextos para a Base de Dados	X	X	X	X	X	X	X	X	X	X	X	X
Visita técnica e reuniões nas Instituições parceiras		X	X	X	X	X	X					
Testes de exportação dos repositórios							X	X	X	X	X	X
Validação do mapeamento conceitual e exportações										X	X	X
Redação de trabalhos científicos para publicação	X	X	X	X	X	X	X	X	X	X	X	X
ATIVIDADE \ MÊS	25	26	27	28	29	30	31	32	33	34	35	36
Criação de servidor de testes	X	X										
Inserção de dados no servidor	X	X	X	X	X							
Testes de representação e recuperação no servidor	X	X	X	X	X	X	X	X	X			
Validação do comportamento do servidor							X	X	X	X	X	
Redação de trabalhos científicos para publicação	X	X	X	X	X	X	X	X	X	X	X	X
Redação do Relatório Final										X	X	X

Fonte: Elaborado pelo autor, 2019.

REFERÊNCIAS

ARBOIT, Aline Elis; BUFREM, Leilah Santiago; KOBASHI, Nair Yumiko. A institucionalização da Ciência da Informação no Brasil sob a ótica da evolução quantitativa dos cursos de graduação na área. Informação & Sociedade: estudos, João

Pessoa, v. 21, n. 1, p. 145-158, jan./abr. 2011. Disponível em: <<http://www.ies.ufpb.br/ojs2/index.php/ies/article/view/9584/5604>>. Acesso em: 21 abr. 2017

ARBOIT, Aline Elis; BUFREM, Leilah Santiago; GONZÁLEZ, José Antonio Moreira. A produção brasileira em Ciência da Informação no exterior como reflexo de institucionalização científica. *Perspectivas em Ciência da Informação*, v. 16, n. 3, 2011.

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS (ABNT). NBR 10520: Informação e documentação - Citações em documentos - Apresentação. Rio de Janeiro: ABNT, 2002.

BUFREM, L. S, et al. Modelizando práticas para a socialização de informações- a construção de saberes no ensino superior. *Perspectivas em Ciência da Informação*, Belo Horizonte, v.15, n.2, p.22-41, maio/ago. 2010.

CABRÉ, M. T. La terminología: teoría, metodología, aplicaciones. Barcelona: Antártida/Empúries, 1993.

CHEN, P. P-S. The Entity-Relationship Model-Toward a Unified View of Data. *ACM Transactions on Database Systems*, v.1, n.1, p. 9-36, 1976. Disponível em: <<http://www.csc.lsu.edu/news/erd.pdf>>. Acesso em 23 out. 2015.

CORTINA, A. O Príncipe de Maquiavel e seus leitores: uma investigação sobre o processo de leitura. São Paulo: Editora UNESP, 2000.

DE MOOR, A.; KEELER, M.; RICHMOND, G. Towards a pragmatic web, In: UTA, Priss et al.

Conceptual Structures: Integration and Interfaces. *Lecture Notes in Computer Science*, v. 2393, p. 235-249, 2002. Disponível em < <http://www.cspeirce.com/menu/library/aboutcsp/richmond/web.pdf>>. Acesso em 23 mar. 2014.

ECO, U. Lector in Fabula: la cooperación interpretativa en el texto narrativo. Barcelona: Lumen, 1987.

ECO, U. Os limites da interpretação. São Paulo: Pioneira, 2000.

EDMONDS, B. The Pragmatic Roots of Context. In: PROC. OF THE 2ND INTERNATIONAL AND INTERDISCIPLINARY CONFERENCE ON MODELING AND USING CONTEXT. Berlin; Heidelberg; New York, v. 1688, 1999. Anais... v. 1688, p. 119-132, 1999.

FRANCELIN, M. M. Ordem dos conceitos na organização da informação e do conhecimento. Tese (Doutorado em Ciência da Informação). Universidade de São

Paulo: PPGCI, São Paulo, 2010.

HABIBI, M. Java regular expressions. New York: Apress Media LLC, 2003.

HJØRLAND, Birger. Domain analysis in Information Science: Eleven approaches – traditional as well as innovative. *Journal of Documentation*, v.58, n.4, p.422-462, 2002.

_____. Domain analysis in Information Science. In: *ENCYCLOPEDIA OF LIBRARY AND INFORMATION SCIENCE*. New York: Dekker, 2004. p.1-7.

KOBASHI, Nair Yumiko. Estudos de institucionalização social e cognitiva da pesquisa científica no Brasil: reflexões sobre um programa de pesquisa. In: LARA, Marilda Lopes Ginez; FUJINO, Asa; NORONHA, Daisy Pires (Org.) *Informação e Contemporaneidade: perspectivas*. Recife: NÉCTAR, 2007. p. 191-204. Disponível em: <<http://www.pos.eca.usp.br/sites/default/files/file/cienciaInformacao/informacaoContemporaneidade.pdf>>. Acesso em: 30 abr. 2017.

MARTINS, Gracy Kelli. *Institucionalização cognitiva e social da Organização e Representação do Conhecimento na Ciência da Informação no Brasil*. 2014. 184f. TESE (Doutorado em Ciência da Informação) – Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília, 2014.

WHITLEY, Richard. Cognitive and social institutionalization of scientific specialities and research areas. In: WHITLEY, Richard. *Social processes of scientific development*. London: Routledge and Kegan, 1974. p. 69-95.

_____. *The Intellectual and Social Organisation of the Sciences*. Oxford: Clarendon Press, 2000.

THE OPEN GROUP. *Regular Expressions*. 1997. Disponível em: <<http://pubs.opengroup.org/onlinepubs/007908799/xbd/re.html> > Acesso em 15 mar. 2017.

Dados do Plano de Trabalho	
Título do Plano de Trabalho:	Representação de Referências em RegEx - Periódicos
Modalidade de bolsa solicitada:	PIBITI
Projeto de Pesquisa vinculado:	Institucionalização Cognitiva e Social da Pesquisa Científica no Brasil: extração automática e análise de citação para identificação de redes de colaboração

1. OBJETIVOS

Estudar as expressões regulares e seu uso; identificar a expressão regular que represente as referências de Periódicos, conforme ABNT 10520:2002; desenvolver um software que identifique adequadamente as referências e seus elementos.

2. METODOLOGIA

Analisar a norma ABNT 10520:2002 e separar cada elemento individual que compõe as referências de Periódicos e seus elementos (autor, título do trabalho, nome do periódico, etc); representar cada elemento identificado via RegEx; codificação dos elementos anteriormente representados em código que permita sua execução de forma isolada.

3. CRONOGRAMA DE ATIVIDADES

Atividade	Mês											
	1	2	3	4	5	6	7	8	9	10	11	12
Estudo para aprofundamento das Expressões Regulares	X	X	X									
Identificação dos Elementos das Referências Conforme ABNT	X	X	X	X	X	X						
Representação dos Elementos das Referências em RegEx	X	X	X	X	X	X	X	X	X			
Testes na ferramenta										X	X	X
Redação de artigos para publicação				X	X	X	X	X	X	X	X	X
Elaboração e envio do Relatório Final individual												X

Dados do Plano de Trabalho	
Título do Plano de Trabalho:	Representação de Referências em RegEx - Eventos
Modalidade de bolsa solicitada:	PIBITI
Projeto de Pesquisa vinculado:	Institucionalização Cognitiva e Social da Pesquisa Científica no Brasil: extração automática e análise de citação para identificação de redes de colaboração

1. OBJETIVOS

Estudar as expressões regulares e seu uso; identificar a expressão regular que represente as referências de Eventos, conforme ABNT 10520:2002; desenvolver um software que identifique adequadamente as referências e seus elementos.

2. METODOLOGIA

Analisar a norma ABNT 10520:2002 e separar cada elemento individual que compõe as referências de Eventos e seus elementos (autor, título do trabalho, local do evento, etc); representar cada elemento identificado via RegEx; codificação dos elementos anteriormente representados em código que permita sua execução de forma isolada.

3. CRONOGRAMA DE ATIVIDADES

Atividade	Mês											
	1	2	3	4	5	6	7	8	9	10	11	12
Estudo para aprofundamento das Expressões Regulares	X	X	X									
Identificação dos Elementos das Referências Conforme ABNT	X	X	X	X	X	X						
Representação dos Elementos das Referências em RegEx	X	X	X	X	X	X	X	X	X			
Testes na ferramenta										X	X	X
Redação de artigos para publicação				X	X	X	X	X	X	X	X	X
Elaboração e envio do Relatório Final individual												X