

INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA DE SÃO PAULO
CAMPUS CAMPINAS

FRANCISCO GONÇALVES DA MOTA LONGHINI

SISTEMA DE GESTÃO DE INFORMAÇÕES DE CURRÍCULOS LATTES

CAMPINAS

2017

FRANCISCO GONÇALVES DA MOTA LONGHINI

SISTEMA DE GESTÃO DE INFORMAÇÕES DE CURRÍCULOS LATTES

Trabalho de Conclusão de Curso apresentado como exigência parcial para obtenção do diploma do Curso de Tecnologia em Análise e Desenvolvimento de Sistemas do Instituto Federal de Educação, Ciência e Tecnologia Campus Campinas.

Orientador: Prof. Dr. Andreiuid Sheffer Corrêa
Co-orientador: Prof. Me. Henrique Gonçalves Salvador

CAMPINAS

2017

L854s Longhini, Francisco Gonçalves da Mota
Sistema de gestão de informações de currículos Lattes / Francisco
Gonçalves da Mota Longhini. – Campinas, 2017.
41f. : il.

Orientador: Andreiwid Sheffer Corrêa.
Coorientador: Henrique Gonçalves Salvador.

Monografia (Graduação) – Instituto Federal de São Paulo – Câmpus
Campinas. Curso de Tecnologia em Análise e Desenvolvimento de Sistemas,
2017.

1. Currículos Lattes. 2. ETL. 3. Produção científica. 4. Gestão de
informações. I. Instituto Federal de São Paulo - Câmpus Campinas. Curso de
Tecnologia em Análise e Desenvolvimento de Sistemas. II. Título.

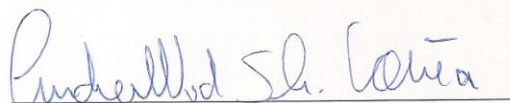
Francisco Gonçalves da Mota Longhini

Sistema de Gestão de Informações de Currículos Lattes

Trabalho de Conclusão de Curso apresentado como exigência parcial para obtenção do diploma do Curso de Tecnologia em Análise e Desenvolvimento de Sistemas do Instituto Federal de Educação, Ciência e Tecnologia de São Paulo Câmpus Campinas.

Aprovado pela banca examinadora em: 13 de DEZEMBRO de 2018.

BANCA EXAMINADORA



Prof. Dr. Andreiuid Sheffer Corrêa (orientador)

IFSP Câmpus Campinas

Prof.^a – Dra.. Bianca Pedrosa

IFSP Câmpus Campinas

Prof. Dr. Jarbas Lopes Cardoso Júnior

CTI – Renato Archer

AGRADECIMENTOS

Agradeço aos professores, meu orientador Dr. Andreiuid Sheffer Corrêa e co-orientador Msc. Henrique Gonçalves Salvador pela oportunidade deste trabalho, paciência e compreensão com as minhas limitações, pelos ensinamentos, confiança depositada, incentivo, amizade e por toda ajuda. Muito obrigado Andreiuid e Henrique.

A todos professores e colaboradores do IFSP – Campus Campinas, muito obrigado pelos ensinamentos, conversas, conselhos e por fazer parte desta minha jornada ao longo destes anos nesta instituição.

À Dra. Bianca Pedrosa e Dr. Jarbas Lopes Cardoso Júnior por aceitarem participar como examinadores do meu trabalho de conclusão de curso.

Aos meus pais José Eduardo e Luzia, e minha mulher Camila, pelo amor e apoio incondicional durante esta jornada.

A Deus e aos espíritos de luz.

RESUMO

No Brasil, dados referentes às pesquisas, produções científicas e currículos, há uma plataforma de cadastro nacional de currículos chamada Currículo Lattes, criada e padronizada em 1999 pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico – CNPq. A plataforma Lattes disponibiliza currículos em formato *HTML* (*HyperText Markup Language* - Linguagem de marcação de hipertexto) e em formato *XML* (*Extensible Markup Language* – Linguagem extensível de marcação genérica) de maneira individualizada, porém não disponibiliza informações agregadas ou estatísticas sobre qualquer conteúdo. O presente estudo teve por objetivo desenvolver uma ferramenta que armazene e exiba dados da plataforma Currículo Lattes, de forma a suportar pesquisa de publicações, relatórios e estatísticas. Para isto foi desenvolvido um procedimento, que lê os itens contidos no arquivo *XML* de cada pesquisador, em seguida armazena estas informações em um banco de dados relacional. Através do levantamento de trabalhos correlatos, questões levantadas por estas pesquisas, foram desenvolvidas funcionalidades como quantidade de produções bibliográficas elaboradas, perfil dos diferentes tipos de produção bibliográfica, regularidade e evolução de publicações ao longo dos anos e dados gerais sobre os pesquisadores, sendo possível exibir gráficos com informações visuais por categoria.

Palavras-chave: Currículos. Lattes. ETL. Produção científica. Gestão de informações.

ABSTRACT

In Brazil, for research data, scientific production and resums, there is a national curriculum registry platform called Curriculum Lattes, this platform was created and designed in 1999 by Conselho Nacional de Desenvolvimento Científico e Tecnológico – CNPq (National Council for Scientific and Technological Development). Lattes platform make resums available individually in *HTML* and *XML* format, but does not provide aggregate or statistical information on any content. The present study aimed to develop a system that stores data and a display it for the Curriculum Lattes platform, in order to research publications, information and statistics. A procedure was developed which reads the items contained in the *XML* file of each researcher resumé, then stores this information in a relational database. Through the survey of related works, questions raised by these researches, functionalities were developed as quantity of elaborated bibliographical productions, profile of the different types of bibliographic production, regularity and evolution of scientific publications over the years and general data about the researchers, being possible to display graphics with visual information by category.

Keywords: Resumé. Lattes. ETL. Scientific production. Information management.

LISTA DE QUADROS

Quadro 1 – Sistemas correlatos.....	19
-------------------------------------	----

LISTA DE FIGURAS

Figura 1 – Fluxo de serialização e de-serialização dos dados.....	24
Figura 2 – Fluxo ETL.....	25
Figura 3 – Diagrama entidade relacionamento do banco de dados.....	25
Figura 4 – Visão da tela de acesso ao sistema.....	26
Figura 5 – Visão da tela de cadastro de currículos.....	27
Figura 6 – Tela de visualização de pesquisadores.....	27
Figura 7 – Visão dos detalhes curriculares do pesquisador.....	28
Figura 8 – Visão de artigos publicados pelo pesquisador.....	29
Figura 9 – Quantidade total de produções bibliográficas e artigos publicados por ano....	30
Figura 10 – Estatística de proporção de publicações por perfil.....	31

LISTA DE SIGLAS

CNPq	Conselho Nacional de Desenvolvimento Científico e Tecnológico
PHP	<i>PHP: Hypertext Preprocessor</i>
SQL	<i>Structured Query Language</i>
HTML	<i>HyperText Markup Language</i> - Linguagem de marcação de hipertexto
ETL	<i>Extract, Transform and Load</i> – Extração, transformação e carga de dados
XML	<i>Extensible Markup Language</i> – Linguagem extensível de marcação genérica

SUMÁRIO

1	INTRODUÇÃO.....	10
2	JUSTIFICATIVA.....	11
3	OBJETIVOS.....	12
3.1	Objetivo geral.....	12
3.2	Objetivo específicos.....	12
4	FUNDAMENTAÇÃO TEÓRICA.....	13
4.1	Sistemas web.....	13
4.2	Engenharia web.....	16
4.3	Trabalhos correlatos.....	17
5	MÉTODOS.....	21
5.1	Levantamento e análise de requisitos.....	21
5.2	Banco de dados.....	21
5.3	Implementação e ferramentas utilizadas.....	22
6	RESULTADOS.....	23
6.1	Obtenção e estudo das fontes de dados.....	23
6.2	Planejamento do escopo da aplicação.....	23
6.3	Procedimento de extração, transformação e carga de dados (ETL).....	24
6.4	Modelo de dados.....	25
6.5	Construção de visões para acesso aos dados.....	26
7	CONCLUSÃO.....	31
8	TRABALHOS FUTUROS.....	32
	REFERÊNCIAS.....	33
	ANEXO A – Ordem de preenchimento do ETL e descrição do conteúdo.....	36
	ANEXO B – Diagramas UML.....	38

1 INTRODUÇÃO

Ao tratar-se de dados referentes às pesquisas, publicações científicas, projetos de pesquisa e currículos, no Brasil há uma plataforma de cadastro nacional de currículos denominada Currículo Lattes, criada e padronizada em 1999 pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) como sendo o sistema de currículo oficial incentivada pelo governo federal para manter informações sobre publicações, orientações, projetos de pesquisa, entre outros (DIGIAMPIETRI et al. 2012).

A plataforma Lattes disponibiliza, em formato *HyperText Markup Language (HTML)* - (Linguagem de marcação de hipertexto) e em formato *Extensible Markup Language (XML)* - (Linguagem extensível de marcação genérica), informações curriculares de maneira individualizada, porém, não disponibiliza informações agregadas ou estatísticas sobre qualquer conteúdo existente no sistema de Currículo Lattes.

Neste contexto, têm surgido ferramentas que visam extrair dados da base de dados da plataforma de Currículo Lattes, como exemplo o *ScriptLattes* (MENA-CHALCO; CESAR-JUNIOR, 2009) e a própria ferramenta de extração de dados da plataforma Lattes, a Lattes Extrator, porém esta ferramenta é de uso restrito e licenciado às instituições de ensino, que podem somente extrair dados inerentes aos pesquisadores, alunos e colaboradores da instituição que faz uso da ferramenta, cabe também a instituição realizar a gestão destes dados.

O presente trabalho visa o desenvolvimento de um sistema online que possibilite a importação (em formato *XML*) e exibição de currículos para auxiliar a gestão de informações curriculares e gerar informações e relatórios de currículos extraídas da base de dados da plataforma Lattes.

2 JUSTIFICATIVA

As instituições acadêmicas brasileiras exploram e utilizam o currículo Lattes no intuito de elaborar relatórios sobre suas produções científicas, supervisões, projetos de grupos de pesquisa, avaliação de programas de graduação, dentre outras atividades rotineiras na gestão de ensino, pesquisa e extensão. A plataforma Lattes disponibiliza alguns filtros de pesquisa de currículos (nome, assunto e instituição), e possibilidade de exportar o currículo no formato *XML*.

Mena-Chalco e César-Junior (2009) citam alguns exemplos de questões que poderiam ser respondidas se plataforma Lattes disponibilizasse mais funcionalidades estatísticas ou agregadas, como exemplo:

- Quantas produções bibliográficas, artísticas ou técnicas foram elaboradas?
- Qual é o perfil (em proporção de publicações) dos diferentes tipos de produção bibliográfica?
- Qual a regularidade e evolução das publicações ao longo dos anos?
- Como é a colaboração/cooperação entre pesquisadores?
- Quantas teses e dissertações foram concluídas?
- Qual a distribuição geográfica dos pesquisadores?
- Qual influência na formação científica de determinados pesquisadores?

O sistema proposto neste trabalho deverá possibilitar ao usuário importar o currículo em formato *XML* (previamente exportado da plataforma Lattes pelo usuário) para auxiliar na gestão de informações curriculares, visualizar e gerar informações e relatórios de currículos Lattes, como exemplo: uma lista compilada de grupos de pesquisa de interesse, regularidade e evolução das publicações e perfil dos diferentes tipos de produções bibliográficas.

3 OBJETIVOS

Os objetivos deste trabalho são apresentados a seguir em duas subseções, objetivo geral contendo o escopo geral do trabalho e objetivos específicos contendo os pormenores para cumprimento deste trabalho

3.1 Objetivo geral

Desenvolver uma ferramenta que suporte os dados dos currículos Lattes de pesquisadores, professores, alunos e colaboradores de modo a suportar pesquisa de publicações, relatórios e estatísticas, como por exemplo: perfil de produção bibliográfica, evolução de publicações por período, quantidade de publicações e dados gerais de pesquisadores e publicações.

3.2 Objetivos específicos

Para o cumprimento do objetivo geral foram levantados os seguintes objetivos específicos:

- Levantar os requisitos do sistema;
- Desenvolver os métodos de extração e transformação de dados;
- Implementar o modelo de banco de dados;
- Desenvolvimento o servidor de aplicação e interface visual do sistema; e
- Propor conjunto de funcionalidades iniciais para extrair informações estatísticas.

4 FUNDAMENTAÇÃO TEÓRICA

Para o embasamento e referencial teórico deste trabalho foram realizadas pesquisas bibliográficas abrangendo sistemas web, engenharia de sistemas web e trabalhos correlatos, sendo apresentadas respectivamente a seguir em subseções de mesmo nome.

4.1 Sistemas web

O serviço conhecido por *World Wide Web*, ou apenas Web, facilitou a criação de um mundo virtual no qual é possível realizar todo tipo de tarefas (BIANCHINI; MALDONADO, 2008). Originalmente, a Web foi criada por Tim Berners-Lee como um projeto para o *CERN* (Organização Europeia para Investigação Nuclear) afim de facilitar o compartilhamento de informações entre grupos de pesquisa (W3C, 2005). Este advento permitiu o surgimento de novos sistemas de informações, combinando navegação por meio de um espaço heterogêneo utilizando operações de consulta que incidem diretamente na informação apresentada (BIANCHINI; MALDONADO, 2008).

No início havia apenas sistemas de hipermídia, compostos por documentos conectados entre si através de elos pré-definidos, porém, com pouca preocupação com planejamento ou projeto (BIANCHINI; MALDONADO, 2008).

O aumento de usuários e aplicações na internet propiciou a demanda para implementação de novas funcionalidades nos navegadores, permitindo a inclusão de elementos dinâmicos, como manipulação de eventos, processamento de formulários, performance de cálculos, entre outros (BOTWE, DAVIS, 2015). Existem sistemas na Web que são aplicações de grande porte, envolvendo interações sofisticadas de usuários a base de dados e, ainda, algumas aplicações podem ser consideradas de missão crítica (GRIFFITHS et al., 2002).

Diversas classificações para sistemas Web foram propostas ao longo do tempo, Powell (et al., 1998) propõe a seguinte classificação:

- *Site estático*: tipo mais simples de aplicação baseada na Web, corresponde a uma coleção de páginas estáticas. A partir de uma página é possível acessar as outras por meio de elos. Como exemplo: documentos, informações editadas e publicadas em formato *HTML*;

- *Site estático com formulários de entrada*: este tipo de site provê um nível de interação básico implementado por meio de formulários de entrada. Permitindo uma maior interação com os usuários. Exemplos: questionários, livros de visita, comentários, sugestões, entre outros;
- *Site com acesso a dados dinâmicos*: o usuário através de consultas pode ter acesso a base de dados. As respostas das consultas são geradas dinamicamente e geralmente apresentadas no formato de documentos *HTML*, *DHTML* ou *XML*;
- *Site criado dinamicamente*: criados nos casos em que o *site* deve fornecer conteúdo personalizado dependendo da interação do usuário;
- *Aplicação Web*: este tipo de aplicação baseada na Web herda atributos dos tipos relacionados anteriormente e acrescenta lógica de negócios, promove a interação do usuário com dados armazenados em bases de dados. Desse modo, fornecem a mesma funcionalidade de uma aplicação de software tradicional.

Segundo Christodoulou (2000) as aplicações Web podem ser divididas em duas categorias: aplicações hipermídias Web ou aplicações de software Web. Aplicação hipermídia Web consiste em uma aplicação que publica informações utilizando nós, *links*, âncoras, estruturas de acesso e disponibilizada através da Web. Aplicações de software Web é definida por uma aplicação de software convencional que depende da infraestrutura Web para sua execução.

Ginige e Murugesan (2001) relacionam as seguintes características de uma aplicação Web:

- Grande parte das aplicações Web são orientadas a documentos, pois são constituídas de documentos conhecidos por páginas Web, podendo ser estáticas quanto dinâmicas;
- Aplicações Web são focadas na aparência e exigem criatividade visual e incorporação de recursos multimídia (gráficos, imagens, áudio e vídeo), tanto na apresentação quanto na interface. Além disso, tais recursos podem acarretar implicações no desempenho da aplicação;
- A maioria das aplicações Web são dirigidas ao conteúdo, pois herdam características de sites Web, que tinham por objetivo originalmente apresentar informações. Portanto, as aplicações envolvem também o desenvolvimento de conteúdo;
- Multiplicidade de perfis de usuários. As aplicações Web podem atender a uma comunidade de usuários numerosa, com usuários de diferentes perfis, que possuem diferentes habilidades e capacidades, o que dificulta a interação homem-computador, a interface com o usuário e a apresentação de informações;

- O conjunto de profissionais envolvidos na construção e desenvolvimento de aplicações Web é variado, tanto em sua experiência, habilidades, conhecimento e compreensão do sistema, bem como em sua percepção da Web e da qualidade das aplicações envolvidas;
- As aplicações Web são executadas em um ambiente mais imprevisível do que o meio onde são executadas as aplicações de software tradicionais;
- Requisitos de segurança e privacidade em uma aplicação baseada na Web são mais exigidos do que em aplicações de software convencionais.

Conallen (2002) classificou as aplicações pelo fato de serem geradas dinamicamente ou não, para aquelas geradas de maneira dinâmica, Araújo (2001) adicionou uma nova classificação que são centradas em um banco de dados.

Segundo Finkelstein (2002), a aplicação Web consiste em todo tipo de aplicação existente para a Web, desde um site até um portal de comércio eletrônico com intenso processamento de informações.

Além destas classificações apresentadas, existe outro tipo de aplicação dentro do ambiente Web: os *Web Services*, cujo são componentes de software distribuídos na internet e que utilizam tecnologias abertas como *WSDL* (W3C, 2001), *XML* (W3C, 2006), *SOAP* (W3C, 2007) que dão suporte ao desenvolvimento rápido e de baixo custo de aplicações (PAPAZOGLU; GEORGAKOPOULOS, 2003). Esses serviços podem ser implementados, oferecidos e utilizados por diferentes empresas, provendo uma infraestrutura para colaboração e integração entre aplicações, desse modo, relacionam-se ao paradigma de *computação orientada a serviços*, que utilizam os *Web Services* como elementos essenciais no desenvolvimento de aplicações, envolvendo camada de serviços, funcionalidades e regras descritas na chamada arquitetura orientada a serviços (*SOA*) (BIANCHINI; MALDONADO, 2008).

Para um usuário que utiliza uma aplicação, os *Web Services* são transparentes, ou seja, são componentes utilizados internamente pelas aplicações e não possuem quaisquer interfaces gráficas (PAMPLONA, 2004). Já para os desenvolvedores, são como componentes de software, mas ao invés de serem utilizados localmente, os são de maneira distribuída, por meio do ambiente Web (BIANCHINI; MALDONADO, 2008).

4.2 Engenharia web

À medida que as aplicações foram crescendo o desempenho, confiabilidade, expectativas, demandas e qualidade depositadas em tais aplicações aumentou de modo significativo, resultando em um maior número de projeto envolvendo-as (BIANCHINI; MALDONADO, 2008), havendo então a necessidade da organização nos padrões de desenvolvimento.

A Engenharia de Software contribuiu com o conhecimento sobre como desenvolver aplicações Web, algumas preocupações adicionais inerentes ao desenvolvimento destas aplicações motivaram pesquisas relacionadas à engenharia de aplicações Web (*Web Engineering*), mantendo o objetivo de aplicar princípios de engenharia de software para desenvolver aplicações Web de qualidade (PRESSMAN, 2000). A Engenharia Web define-se como a aplicação de abordagens sistemáticas, disciplinadas e quantificáveis para o desenvolvimento, operação e manutenção de aplicações baseadas na Web (DESHPANDE et al., 2003).

Segundo Fraternalli e Paolini (2000), o projeto de uma aplicação Web deve considerar particularidades relacionadas às dimensões:

- Estrutural (conceitual): define a organização das informações a serem tratadas pela aplicação e os seus relacionamentos;
- Navegacional: representa como as informações serão acessadas através da aplicação;
- Apresentação: descreve como as informações e o acesso a essas serão apresentados ao usuário da aplicação.

Essas dimensões definem as diferentes visões para o projeto de aplicação, de acordo com Pastor (2004), o objetivo principal da Engenharia Web é desenvolver aplicações corretas, nas quais suas estruturas, funcionalidades, aspectos navegacionais e de interação como o usuário estejam representadas de forma apropriada.

Os processos que envolvem o desenvolvimento de aplicações Web devem produzir representações para projetos de aspectos de aplicações tradicionais, tais como, estrutura e funcionalidades; e também para aspectos orientados para a Web, como navegação e apresentação (com recursos Web) (CONTE et al., 2005).

Mendes (et al., 2004) discutem as principais diferenças entre desenvolvimento de software convencional e Web, Conte (et al., 2005) realiza uma revisão sistemática para processos de desenvolvimento para aplicações Web, no entanto, não identifica uma abordagem padronizada que promova e utilize o conhecimento adquirido em Engenharia de

Software em prol da construção de produtos de software Web. Vários métodos têm sido propostos, tais como, a *WebML* (CERI et al., 2000), o *OOHDM* (SCHWABE et al., 1996), as extensões para linguagem *UML* (CONALLEN, 2002), *OOWS* (FONS et al., 2003), *W2000* (BARESI et al., 2000), *ADM* (DIAZ et al., 2004), entre outros.

Os processos aplicados na prática são em sua maioria ad-hoc, ou seja, não formalmente definidos e por muitas vezes melhores práticas de Engenharia de Software, que focam em garantir a qualidade de processo e de produto, não estão explícitas ou não são atendidas (STANDING, 2002). Os riscos de se utilizar processos de desenvolvimento ad-hoc são bem conhecidos (PRESSMAN, 2005). Segundo Abrahão et al. (2003), a avaliação da qualidade de aplicações Web tem sido avaliadas também de maneira ad-hoc, baseada primariamente em senso comum, intuição e conhecimento de desenvolvedores. De acordo com as melhores práticas de Engenharia de Software (PFLEEGER, 2004), os processos de desenvolvimento devem incluir atividades de garantia de qualidade.

Todavia, devemos considerar que diferentemente da maioria das aplicações de software tradicionais, que evoluem por meio de versões planejadas e espaçadas cronologicamente (ciclo de vida), as aplicações Web evoluem continuamente requerendo sempre modificações em sua estrutura e novo conteúdo, o que em muitos casos deve-se ao fato que não é possível definir o que a aplicação deve conter no início do seu processo de desenvolvimento, uma vez que há a variação de requisitos especialmente após ser colocada em uso (BIANCHINI; MALDONADO, 2008). O gerenciamento da evolução de uma aplicação Web é um dos maiores desafios técnicos, organizacional e gerencial para os desenvolvedores (MURUGESAN; GINIGE, 2005).

4.3 Trabalhos correlatos

Ao longo dos anos foram realizadas diversas pesquisas científicas envolvendo mineração de dados na plataforma Lattes. Martino et al. (2009) apresentou um estudo teste da ferramenta KAIROS para aplicação de técnicas de mineração de dados em textos científicos a fim de identificar especialistas e efetuar prospecção tecnológica. A ferramenta KAIROS minerou 1.100.000 produções científicas de 700.000 pesquisadores, Martino et al. (2009) posteriormente apresentaram estes resultados a 350 pesquisadores, 80% destes pesquisadores concordaram com as áreas promissoras (prospecção tecnológica) identificadas pela ferramenta e também obteve 94,3% de média de acerto sobre área de atuação dos pesquisadores.

Silva e Smit (2009) buscam avaliar e propor sugestões a melhoria e disponibilização das informações da plataforma Lattes, eles concluem que há o comprometimento em preenchimento de currículos comprometendo análises estatísticas.

Mena-Chalco e César-Júnior (2009) realizaram uma grande contribuição a comunidade de pesquisadores que trabalham com mineração de dados da plataforma Lattes com o desenvolvimento e disponibilização gratuita da ferramenta *scriptLattes*. A ferramenta, através de uma lista de identificadores presentes nos currículos, consegue gerar diversas páginas *HTML* organizando informações por categorias, mapa de distribuição geográfica de pesquisadores e grafo de coautorias.

Alves et al. (2011) desenvolveram uma ferramenta protótipo chamada SUCUPIRA na qual objetiva explorar os relacionamentos existentes entre os pesquisadores, buscando responder questões como: “Qual a relação entre pesquisadores ?”; “Em que região do país, estado e instituição os pesquisadores estão localizados ?”; “Quantas publicações em revistas e congressos ?”; e “Como é a regularidade e a evolução das publicações ?”.

Digiampetri e Silva (2011) desenvolvem um framework que utiliza agentes de busca para encontrar currículos de determinado pesquisador na plataforma Lattes de maneira automática e também provê uma interface gráfica que exibe a rede de coautoria de determinado pesquisador, analisando e cruzando informações sobre a produção destes pesquisadores através da análise de currículos.

Digiampetri et al. (2012) compilaram um banco de dados relacional com granularidade de campo a campo (não de publicação a publicação) e contendo mais de 1.000.000 de currículos minerados na plataforma Lattes e apresentam algumas características como relações entre currículos ligando orientados, orientadores e coautores de publicações científicas. Foram utilizadas ferramentas como *wget* e *parser* para obtenção de identificadores para download de currículos. Digiampetri et al. (2012) apresentaram diversos resultados com análises estatísticas como: distribuições de grandes áreas de atuação; principais áreas de atuação; período da última atualização do currículo em ano e meses; das publicações por tipo, número médio de autores; evolução no número publicações no tempo; de bolsas de estudo; orientações por categoria; e quantidade de orientações por tipo ao longo dos anos.

Digiampetri et al. (2014) buscaram analisar a atualização do currículo vitae na Plataforma Lattes e segmenta-los em áreas de conhecimento e formação acadêmica, os resultados foram apresentados relacionando a frequência de acordo com a segmentação proposta.

Amaral (et al., 2016) utilizaram da ferramenta *ScriptLattes* para identificar e analisar o perfil de pesquisadores de inteligência competitiva, foram analisados 1.434 currículos com dados referenciais de 223 artigos e 26 projetos de pesquisa em andamento. Os resultados apresentados foram a categorização da produção científica e dos projetos de inteligência competitiva, correlação de temas e representação do perfil de atuação e formação do pesquisador.

No mercado brasileiro existe a plataforma paga Stela Experta, desenvolvida pela empresa Tekis Tecnologias Avançadas Ltda., esta ferramenta tem como intuito fornecer informações qualitativas e quantitativas de currículos Lattes às instituições de ensino superior Brasileiras, suportando serviços de apoio estratégico nas áreas de pesquisa, extensão e gestão de ensino. A plataforma Stela Experta disponibiliza somente as informações de servidores, alunos e colaboradores da instituição de ensino contratante da plataforma.

Através do estudo em literatura das ferramentas desenvolvidas em trabalhos correlatos, o quadro 1 demonstra as peculiaridades entre cada aplicação.

Quadro 1 - Sistemas correlatos

Aplicação	Ano	Linguagem	Processo ETL próprio ?	Interface Web ?	Principal Propósito
<i>scriptLattes</i>	2009	Python	Sim, a aplicação é a ferramenta de extração.	Gera arquivos HTML, interface em linha de comando.	Extração de currículos lattes
KAIROS	2009	Sem menção	Utiliza <i>Agente Extrator</i> para extração	Sem menção	Prospecção tecnológica e identificação de especialistas
SUCUPIRA	2009	Java / Adobe Flash	Utiliza a ferramenta <i>Lattes Miner</i>	Sim	Visualização das redes sociais acadêmicas
DIGIAMPETRI e SILVA - Framework	2011	Java	Sim, através da Google API realiza a escavação por busca de currículos HTML	Sem menção	Visualização de rede sociais e de co-autorias.
Eckstein –	2016	Java	Sim, utiliza a	Sim	Aplicação

Aplicativo de gestão curricular			API JAXB		voltada a inteligência de negócios
Sistema de Gestão de Currículos Lattes	2017	PHP	Sim, utiliza componente de framework web – <i>Symfony</i> – <i>Serializer</i> <i>Component</i>	Sim	Relatórios e Estatísticas a partir de uma lista compilada de pesquisadores.

Fonte – Elaborado pelo autor (2017)

Dentre as pesquisas mencionadas, as ferramentas realizam a extração de dados, mas grande parte destas ferramentas cabe aos pesquisadores realizarem a análise e estatística dos dados, também não há quaisquer informações sobre: opção de acesso via interface web; disponibilização destas ferramentas de maneira aberta e gratuita; e menções sobre a manutenção destas ferramentas.

5 MÉTODOS

Para o desenvolvimento da ferramenta foram necessárias as etapas de levantamento e análise de requisitos, banco de dados e implementação, detalhadamente descritas a seguir em subseções.

5.1 Levantamento e análise de requisitos

O estudo de técnicas de *ETL* (*Extract, Transform and Load*, em português, Extração, Transformação e Carga de Dados) foi necessário para a transformação e persistência dos dados. Para obtenção dos dados (extração de currículos) o usuário deve utilizar plataforma Lattes, a qual oferece a opção de exportar currículos em formato *XML*. Foi necessário também, o estudo e entendimento sistemático da estrutura do arquivo *XML* para elaboração do design do banco de dados. Os resultados destes estudos e métodos de transformação e carga de dados estão descritos na seção de solução proposta.

Através do estudo de trabalhos correlatos foram identificadas algumas necessidades que foram levantadas pelos pesquisadores autores. As sugestões propostas para trabalhos futuros foram sistemas gratuitos e abertos que possam possibilitar a inclusão de currículos extraídos da base de dados da plataforma Lattes e possam exibir os dados presentes nestes currículos em forma de estatísticas e dados consolidados,

Este trabalho buscou atender algumas das necessidades levantadas em outros trabalhos correlatos, necessidades estas apresentaas na seção de “Justificativa”.

O sistema foi desenvolvido utilizando o padrão de projeto MVC (Modelo, Visão e Controlador), os artefatos produzidos (diagramas) podem ser consultados no final deste trabalho na seção de anexo B.

5.2 Banco de dados

O banco de dados utilizado foi o *MySQL Community Server* por ser gratuito e atender às necessidades do sistema, como o armazenamento dos dados providos do currículo *XML*. A estrutura do banco de dados foi elaborada com base no estudo da estrutura do currículo *XML* e é demonstrada na seção de solução proposta.

5.3 Implementação e ferramentas utilizadas

A linguagem de programação escolhida para o lado servidor do sistema foi *PHP*: (*Hipertext Preprocessor - PHP*), uma linguagem de *scripting* do lado do servidor, gratuita e de código aberto, geralmente usado para criar páginas web dinâmicas. Originalmente criado em 1994 para ajudar a simplificar as tarefas básicas do site, PHP desde então se tornou uma das linguagens de programação mais usadas em criação de sites, pode ser usado como uma linguagem de *scripting* de propósito geral para criar aplicativos de linha de comando e com interface de usuário juntamente com as páginas Web dinâmicas. *PHP* também é versátil e pode ser instalado e configurado em multiplataformas, mas também vem pré-instalado em muitas distribuições Linux. *PHP* permite que os usuários finais possam interagir mais com a Web e criar páginas com menos esforço. A escolha do *PHP* reside no fato de ser uma linguagem gratuita, consolidada no mercado, apoio da comunidade (material de suporte) e prover uma gama de bibliotecas,

Foi também escolhido um framework *PHP* Web denominado *Symfony*, este framework é um conjunto de componentes *PHP* desacoplados e reutilizáveis, a escolha pelo *Symfony* deve-se pela maturidade, documentação e forte comunidade deste framework que já foi utilizado em projetos como *Drupal*, *phpBB*, *Laravel*, *PIWIK* e *OroCRM*.

Para o lado cliente do sistema foi utilizado tecnologias como *HTML5*, *CSS3* (framework *Bootstrap*) e *Javascript*.

6 RESULTADOS

A construção da aplicação seguiu as etapas apresentadas na seção “Métodos”, a seguir é apresentada a solução proposta:

6.1 Obtenção e estudo das fontes de dados

Os dados curriculares para teste do sistema foram baixados através da Plataforma Lattes, através da ferramenta de exportação disponível na plataforma. Após, foi realizado o estudo da estrutura *XML* do currículo para concepção da fase *ETL* (exceto extração), e para elaboração da estrutura do banco de dados para armazenamento dos dados extraídos dos arquivos *XML*.

6.2 Planejamento do escopo da aplicação

Através do estudo dos dados obtidos, foram mapeadas e projetadas as seguinte funcionalidades e telas, sendo:

- *Login*: a tela de *login* é a primeira a ser apresentada ao usuário quando o mesmo acessa a página inicial do sistema.
- Cadastro de usuário: esta tela e funcionalidade permite o usuário criar uma conta no sistema, pode ser acessada via tela de *login* ao clicar em botão ‘Criar conta’. O acesso ao sistema somente poderá ser realizado mediante ao cadastro prévio do usuário.
- Cadastro de currículos - esta tela e funcionalidade permite ao usuário importar os currículos em formato *XML* (previamente exportados da plataforma Lattes). Pode-se realizar a importação de um ou vários arquivos de uma única vez.
- Visualização de currículos - esta tela e funcionalidade permite ao usuário visualizar todos os currículos importados e vinculados a sua conta no sistema.
- Visualização de currículo de pesquisador – esta tela e funcionalidade permite ao usuário visualizar as informações sobre dados gerais e produções bibliográficas do pesquisador cadastrado.
- Busca de produções bibliográficas – esta tela e funcionalidade permite ao usuário buscar através de um campo de pesquisa por produções bibliográficas através de palavras-chave presentes nos currículos cadastrados.

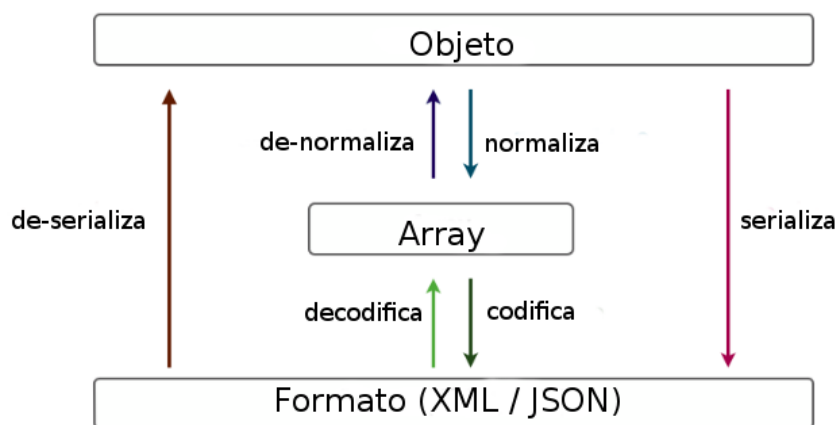
- Estatísticas de produções bibliográficas – esta tela e funcionalidade permite ao usuário visualizar as estatísticas propostas neste trabalho, de currículos cadastrados e de forma individualizada.

6.3 Procedimento de extração, transformação e carga de dados (ETL)

Após o processo de planejamento das funcionalidades e telas propostas, foi necessário a escolha dos métodos para a transformação e persistência dos dados do arquivo *XML* para o banco relacional (*MySQL*). Foi utilizada a linguagem *PHP*, *framework Symfony* e *ORM Doctrine*.

- A) *Symfony Framework - Serializer Component*: Para a transformação dos dados no arquivo *XML* foi utilizado o componente de serialização e de-deserialização do *framework Symfony*. Serializar significa criar uma representação em um determinado formato com um determinado objetivo, os dados em formato *XML* foram convertidos em representação de Objetos (Entidades) para persistência de dados. O fluxo de serialização é demonstrado na Figura 1.

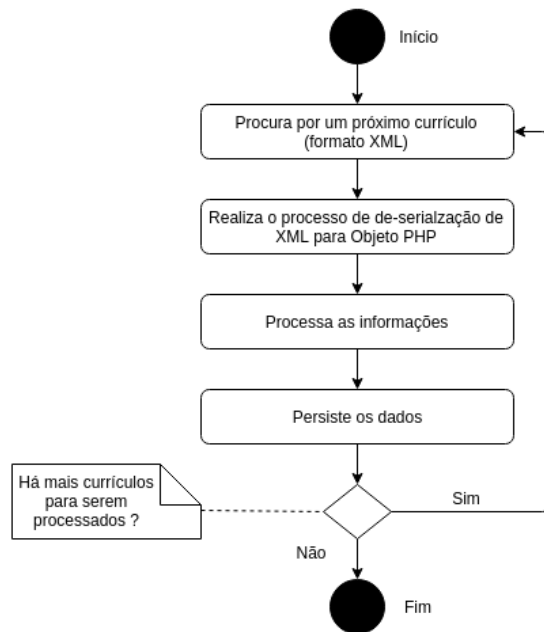
Figura 1 - Fluxo de serialização e de-deserialização dos dados.



Fonte: Adaptado de SYMFONY, 2017.

- B) *ORM Doctrine – API (Application programming interface – Interface de programação de aplicativos) PHP* utilizada para a persistência dos dados para o banco de dados *MySQL*.
- C) Sistema de Gestão de Currículos Lattes - o sistema faz uso de frameworks e *APIs* citadas para a realização da extração, transformação e carga dos currículos (ETL) conforme a Figura 2.

Figura 2 - Fluxo ETL

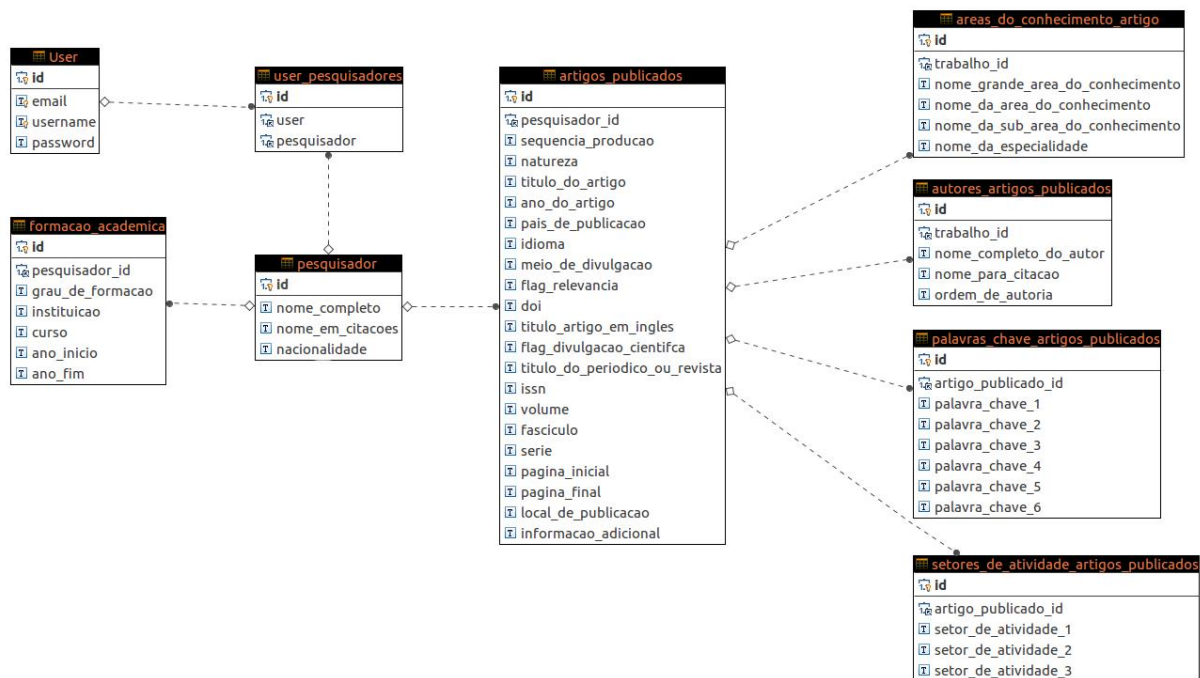


Fonte: Elaborado pelo autor (2017)

6.4 Modelo de Dados

Os dados provenientes dos currículos foram armazenados em tabelas ilustradas no modelo lógico, exibido na Figura 3.

Figura 3 - Modelo lógico do banco de dados.



Fonte: Elaborado pelo autor (2017)

Na Figura 3, há somente a representação de artigos publicados, porém, o a estrutura do modelo se repete para artigos aceitos para publicação, capítulos ou livros publicados, textos em jornal ou revista e trabalhos em eventos. O Anexo A na seção de anexo demonstra a ordem de preenchimento do *ETL* e conteúdo de cada.

6.5 Construção de visões para acesso aos dados

Após a etapa de *ETL*, foi implementado o *front-end* do sistema nas seguintes categorias: *login* (Figura 4), cadastro de currículos (Figura 5), visualização de pesquisadores (Figura 6), detalhes de dados gerais do pesquisador (Figura 7), detalhes de artigos publicados (Figura 8) e estatísticas de quantidade total de produções bibliográficas e artigos publicados por ano (Figura 9) e proporção de publicações por perfil.

Na visão de *login* (Figura 4), esta sendo a primeira tela que o usuário visualiza ao acessar o sistema, observa-se na Figura 4 duas marcações, A e B. Em A há a seção em que o usuário que já tiver uma conta criada no sistema irá entrar com o seu nome de usuário e senha, clicando no botão “Entrar” para acesso ao sistema. O sistema valida as informações e apresenta uma mensagem de erro ao usuário caso a senha esteja incorreta ou não possa localizar informações sobre os dados de acesso do usuário no banco de dados. Na marcação B, há a opção de criar uma conta, a qual o usuário deverá inserir informações para criar uma conta de acesso ao sistema.

Figura 4 – Visão da tela de acesso ao sistema



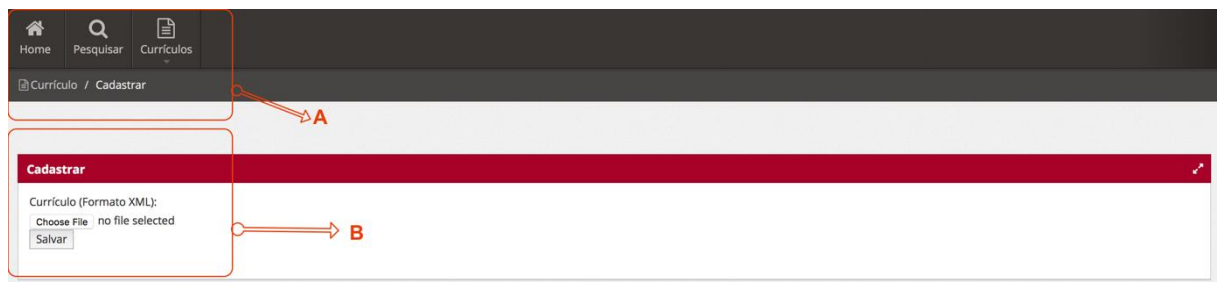
Fonte: Elaborado pelo autor (2017)

Na Figura 5, a visão de cadastro de currículos, o usuário poderá realizar o envio de currículo(s) em formato *XML* para o sistema processar e salvar as informações contidas nestes currículos enviados. Observa-se na Figura 5 duas marcações, A e B, a marcação A contém a barra de navegação do sistema contendo as opções: *Home*, *Pesquisar* e *Currículos*, para

acessar esta tela de cadastro de currículos, o usuário deve clicar em Currículos e em seguida em Cadastrar. Na marcação A também se observa abaixo do menu de navegação uma barra denominada *breadcrumb* com o propósito de apresentar ao usuário sua localização dentro do sistema.

A marcação B, na Figura 5, o usuário pode realizar o envio do(s) currículo(s) em formato *XML*, o sistema notifica o usuário caso ele venha tentar realizar o envio de arquivos que não sejam do formato *XML*.

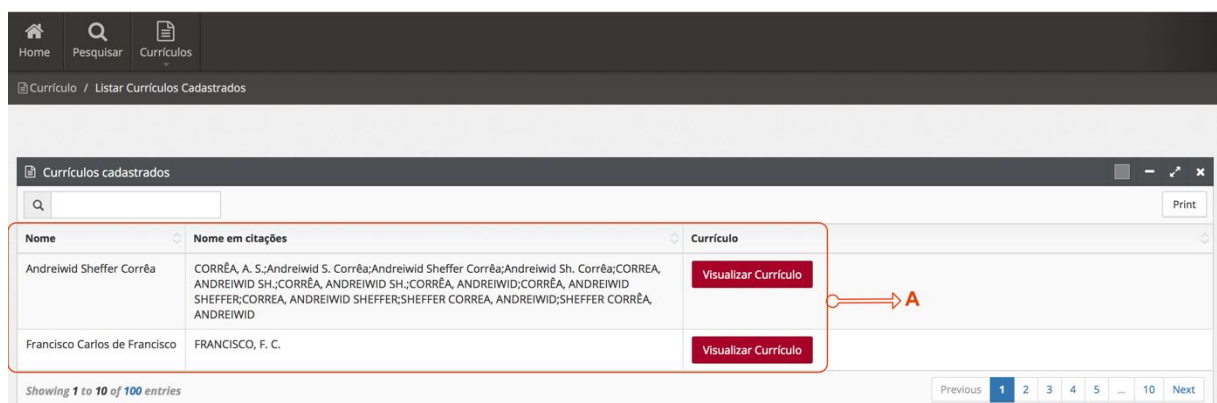
Figura 5 - Visão da tela de cadastro de currículos



Fonte: Elaborado pelo autor (2017)

Após o cadastro de currículos no sistema o usuário poderá visualizar os currículos cadastrados, classificados por pesquisador, na tela de visualização de pesquisadores conforme representado na Figura 6. A marcação A, na Figura 6, mostra a lista de pesquisadores já cadastrados no sistema, nesta lista pode-se verificar o nome do pesquisador, nome em citações de produções bibliográficas e a opção de clicar em um botão para a visualização detalhada do currículo do pesquisador.

Figura 6 - Tela de visualização de pesquisadores



Fonte: Elaborado pelo autor (2017)

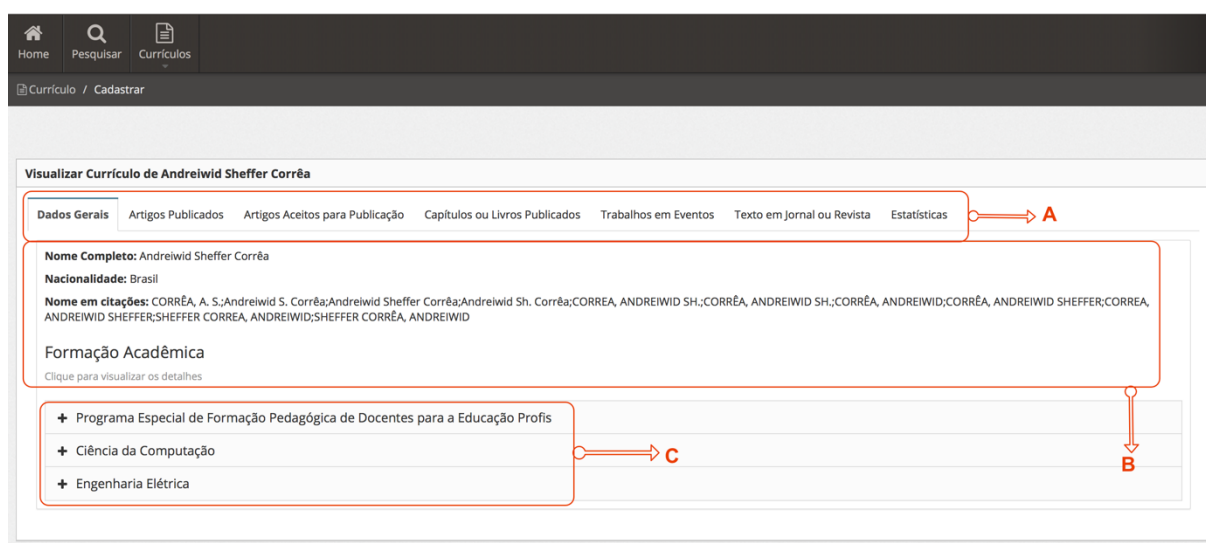
Quando o usuário desejar visualizar com detalhes o currículo do pesquisador cadastrado ele irá ser direcionado para a tela de visão dos detalhes curriculares, representado na Figura 7. Na Figura 7, temos três marcações, A; B; e C.

A marcação A representa abas contendo dados do currículo do pesquisador, sendo estas abas: Dados Gerais (apresenta informações gerais como nome, nacionalidade, e formação acadêmica); Artigos Publicados; Artigos Aceitos para Publicação; Capítulo ou Livros Publicados; Trabalhos em Eventos; Texto em Jornal ou Revista; e Estatísticas. Cada uma destas abas irá apresentar as informações contidas no currículo conforme o título da aba.

Na marcação B, será mostrado com detalhes as informações das abas, na Figura 7 temos a representação dos Dados Gerais do pesquisador.

A marcação C na Figura 7 representa cada item da seção da aba escolhida como exemplo na Figura 7 temos a aba de Dados Gerais, e cada formação acadêmica é inserida na seção C, ao clicar em cada uma da formação acadêmica ela se expande mostrando os detalhes.

Figura 7. Visão dos detalhes curriculares do pesquisador

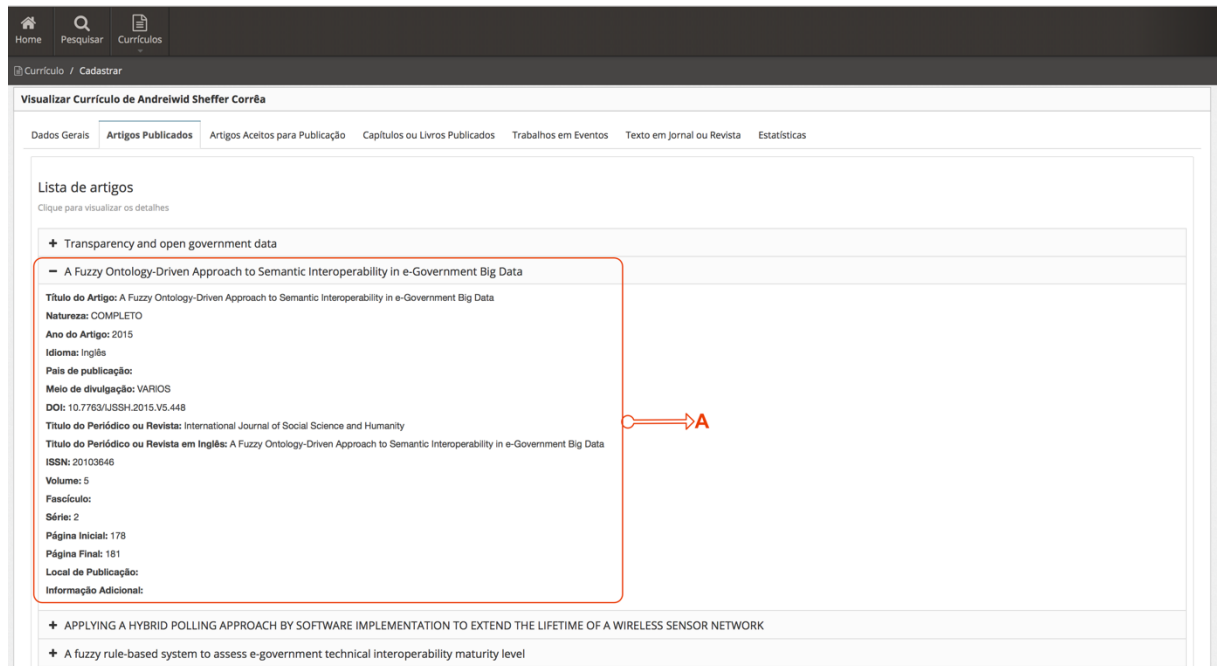


Fonte: Elaborado pelo autor (2017)

Na tela de detalhes curriculares do pesquisador, o usuário ao selecionar Artigos Publicados pode verificar uma lista de todos os artigos que o pesquisador possui cadastrado em seu currículo.

A marcação A da Figura 8, mostra que ao clicar no título do artigo que deseja visualizar mais detalhes, este se expande e exhibe os detalhes.

Figura 8 - Visão de artigos publicados pelo pesquisador



Fonte: Elaborado pelo autor (2017)

Se o usuário desejar visualizar as estatísticas geradas para o currículo do pesquisador que ele está visualizando, o usuário deve clicar na aba de “Estatísticas” (conforme visto na Figura 7, marcação A).

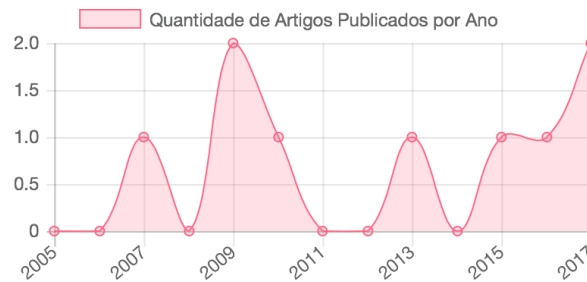
As estatísticas geradas e propostas neste trabalho se dividem em três categorias, “Quantidade total de produções bibliográficas”; “Artigos Publicados por ano”; e “Proporção de publicações por perfil”.

Na Figura 9, podemos visualizar a quantidade total de produções bibliográficas e o gráfico gerado da quantidade de artigos publicados por ano ao longo do tempo.

Figura 9. Estatística de quantidade total de produções bibliográficas e artigos publicados por ano por pesquisador cadastrado.

Quantidade total de produções bibliográficas: 22

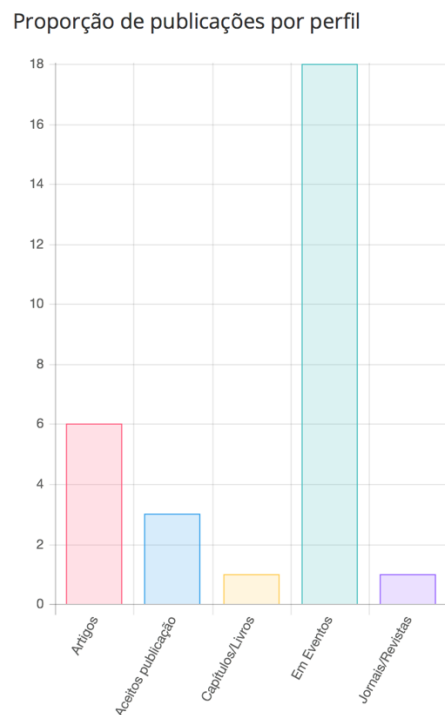
Artigos publicados por ano



Fonte: Elaborado pelo autor (2017)

A Figura 10 exibe o gráfico gerado pelo sistema da proporção de publicações por perfil de diferentes tipos de produção bibliográfica. Sendo exibida a quantidade de publicações e classificadas nas categorias propostas pela plataforma Lattes, assim o usuário pode visualizar e comparar a quantidade de publicações do pesquisador cadastrado.

Figura 10 - Estatística de proporção de publicações por perfil de produção bibliográfica.



Fonte: Elaborado pelo autor (2017)

7 CONCLUSÃO

O Sistema de Gestão de Informações de Currículos Lattes se propôs através do processamento dos dados contidos em currículos da plataforma Lattes disponibilizar informações em um painel. Para isto foi desenvolvido um procedimento que lê os itens contidos no arquivo *XML* de cada pesquisador, em seguida armazena estas informações em um banco de dados relacional.

Através do levantamento de trabalhos correlatos, questões levantadas por estas pesquisas, foram desenvolvidas funcionalidades como quantidade de produções bibliográficas elaboradas, perfil dos diferentes tipos de produção bibliográfica, regularidade e evolução de publicações ao longo dos anos e dados gerais sobre os pesquisadores, sendo possível exibir gráficos com informações visuais por categoria.

Ao possibilitar ao usuário do sistema montar uma lista de grupos de interesse e exibir os dados do currículo Lattes agregando as informações quantitativas de produção científica de cada currículo e exibindo de forma individualizada (em forma de gráficos e texto), o usuário pode obter uma compreensão, percepção ou clareza sobre as informações obtidas.

Espera-se que o sistema produzido neste trabalho ao ser disponibilizado de forma gratuita e de código aberto possa auxiliar gestores de ensino, pesquisadores e colaboradores na gestão das informações de currículos Lattes.

8 TRABALHOS FUTUROS

O sistema produzido poderá receber novas funcionalidades: geração de estatísticas por todos ou alguns currículos selecionados; novas estatísticas como localização geográfica e rede de colaboradores; classificação por tabela de áreas de conhecimento; sincronização de atualização de currículos; possibilidade de escolha do formato dos gráficos gerados; relatórios automatizados e manuais por período e produção científica.

Implementação de novos métodos ou integração de métodos existentes de extração de currículos na plataforma Lattes também será uma funcionalidade necessária para evoluir o sistema e melhorar a experiência para o usuário.

Será também criada uma área de sugestão de funcionalidades pelos usuários, pois como o sistema é de código aberto e gratuito, qualquer desenvolvedor de sistemas que desejar colaborar com a evolução do sistema poderá acatar (se assim desejar) a sugestão de usuários e começar a implementá-las, colaborando através de um repositório no qual o código será armazenado e disponibilizado.

REFERÊNCIAS

ABRAHÃO, S.; CONDORI-FERNÁNDEZ, N. OLSINA, L.; PASTOR, O. Defining and Validating Metrics for Navigational Models. **Proceeding of the Ninth International Software Metrics Symposium (METRICS'03)**, 2003.

ALVES, Alexandre Donizeti; YANASSE, Horacia Hideki. **SUCUPIRA: A System for Information Extraction of the Lattes Platform to Identify Academic Social Networks**. 2011.

ARAÚJO, A. C. M. **Framework de Análise e Projeto Baseado no RUP para o Desenvolvimento de Aplicações Web**. Recife, PE. Originalmente apresentada como dissertação de mestrado, Centro de Informática, Universidade Federal de Pernambuco, 2001.

BARESI, L., GARZOTTO, F., PAOLINI, P. From Web Sites Applications: New Issues for Conceptual Modeling. **Proceeding of the ER'00 Workshop on World Wide Web and Conceptual Modeling, Salt Lake City (USA)**, 2000.

BIANCHINI, S.; MALDONADO, J. C. **Avaliação de métodos de desenvolvimento de aplicações web**. São Paulo, SP. Originalmente apresentada como dissertação de mestrado, Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, 2008.

BOTWE, D. A.; DAVIS, J. G. A comparative study of web development technologies using open source and proprietary software. **International Journal of Computer Science and Mobile Computing**, v. 4, n. 2, p. 154-165, 2015.

CERI, S.; FATERNALI, P.; BONGIO, A. Web modeling language (webml): a modeling language for designing web sites. **Proceedings of the 9th international World Wide Web conference on Computer networks: the international journal of computer and telecommunications networking**. Amsterdam, Netherlands. North Holland Publishing Co., p. 137-157, 2000.

CHRISTODOULOU, S. P.; ZAFIRIS, P.A.; PAPATHEODOROU, T.S., WWW2000: The developer's view and a practitioner's approach to web engineering. **Proceedings 2nd ICSE Workshop Web Eng.**, p. 75-92, 2000.

CONALLEN, J. **Building Web Applications with UML**. 2º Edição. Editora: Addison-Wesley, 2002.

CONTE, T.; MENDES, E.; TRAVASSOS, G. H. Processo de desenvolvimento para aplicações web: Uma revisão sistemática. **XI Simpósio Brasileiro de Sistemas Multimídia e Web**. Poços de Caldas - MG, 2005.

DIAZ, P.; MONTERO, S.; AEDO, I. Modelling hypermedia and web applications: the Ariadne Development Method. **Information Systems**, 2004.

DESHPANDE, Y.; MURUGESAN, S.; GINIGE, S.; HANSEN, S.; SCHWABE, D.; GAEDKE, M.; WHITE, B. Web engineering. **ArXiv Computer Science e-prints**, 2003.

DIGIAMPIETRI, Luciano Antonio; DA SILVA, Ernando Eduardo. A framework for social network of researchers analysis. **Iberoamerican Journal of Applied Computing**, v. 1, n. 1, 2011.

DIGIAMPIETRI, Luciano. et al. Minerando e caracterizando dados de currículos lattes. In: **Brazilian Workshop on Social Network Analysis and Mining** (BraSNAM). 2012.

DIGIAMPIETRI, Luciano et al. Análise da atualização dos Currículos Lattes. **IV Encontro Brasileiro de Bibliometria e Cientometria**, 2014.

ECKSTEIN, L.F.P.; TEIXEIRA, F.O.M.; LOUTFI, M.S.; ALVARES, R.V. **Um aplicativo para gestão curricular docente no ensino superior: um estudo de caso envolvendo informações da plataforma Lattes**.

FINKELSTEIN, A. C. W.; SAVIGNI, A.; KAPPEL, G.; RETSCHITZEGGER, W.; KIMMERSTORFER, E.; SCHWINGER, W. Ubiquitous web application development - a framework for understanding. **6th world Multiconference on Systemics, Cybernetics and Informatics**. Orlando, Florida - US, 2002.

FONS, J.; PELECHANO, V.; ALBERT, M.; PASTOR, O. Development of Web Applications from Web Enhanced Conceptual Schemas. **22nd International Conference on Conceptual Modeling (ER)**, 2003.

FRATERNALI, P.; PAOLINI, P. Model-driven development of web applications: the autoweb system. **ACM Transactions on Office Information Systems**, 2000.

GINIGE, A.; MURUGESAN, S. Web engineering: An introduction. **IEEE Multimedia**, 2001b, p. 14-18, 2001.

GRIFITHS, G.; HEBBRON, B. D.; LOCKYER, M. A.; OATES B. J. A simple method & tool for web engineering. **Proceedings of the 14th international conference on software engineering and knowledge engineering**, New York - NY, USA: ACM Press, p. 755-762, 2002.

MARTINO, R. N.; OLIVEIRA, Jonice; SOUZA, J. M. Mineração de Dados Científicos para Prospecção Tecnológica e Identificação de Especialistas. In: **V Workshop em Algoritmos e Aplicações de Mineração de Dados**. Ceará. 2009.

MENA-CHALCO, J.P; JUNIOR, C.; MARCONDES, R.. ScriptLattes: an open-source knowledge extraction system from the Lattes platform. **Journal of the Brazilian Computer Society**, v. 15, n. 4, p. 31-39, 2009.

MENDES, E.; MOSLEY, N.; COUNSELL, S. Investigating Web size metrics for early Web cost estimation. **The Journal of System and Software**, Artigo impresso, 2004.

MURUGESAN, S.; GINIGE, A. **Web engineering: Introduction and perspectives**. editora: Idea Group Publishing, cap. 1, p. 23, 2005.

PAMPLONA, V. F. **Web services, construindo, disponibilizando e acessando Web Services via J2SE e J2ME**. 2004. Disponível em <http://www.javafree.org/content/view.jf?idContent=4>>. Acesso em 15/03/2017.

PASTOR, Oscar. Fitting the pieces of the Web engineering puzzle. **Invited Talk, XVIII Simpósio Brasileiro de Engenharia de Software (SBES)**, Brasília, 2004.

PAPAZOGLU, M. P.; GEROGAKOPOULOS, D. Service-Oriented Computing. **Communications of the ACM**, v. 46, n. 10, p. 24-28, 2003.

PFLEEGER, S. **Engenharia de Software: Teoria e Prática**, São Paulo: Prentice Hall, 2ª edição, 2004.

POWELL, T.; JONES, D.; CUTTS, D. **Web site engineering: beyond web page design**. Prentice Hall, 1998.

PRESSMAN, R. S. What a tangled Web we have. **IEEE Software**, (Jan - Feb), p.18 - 21, 2000.

PRESSMAN, R. S. **Software engineering - A Practioner's Approach**. 6 ed. McGraw-Hill, 2005.

SCHWABE, D.; ROSSI, G.; BARBOSA, S. Systematic Hypermedia Design with OOHDM. **ACM Conference on Hypertext**. 1996.

SILVA, Fábio Mascarenhas; SMIT, Johanna Wilhelmina. Organização da informação em sistemas eletrônicos abertos de Informação Científica & Tecnológica: análise da Plataforma Lattes. **Perspectivas em ciência da informação**, v. 14, n. 1, p. 77-98, 2009.

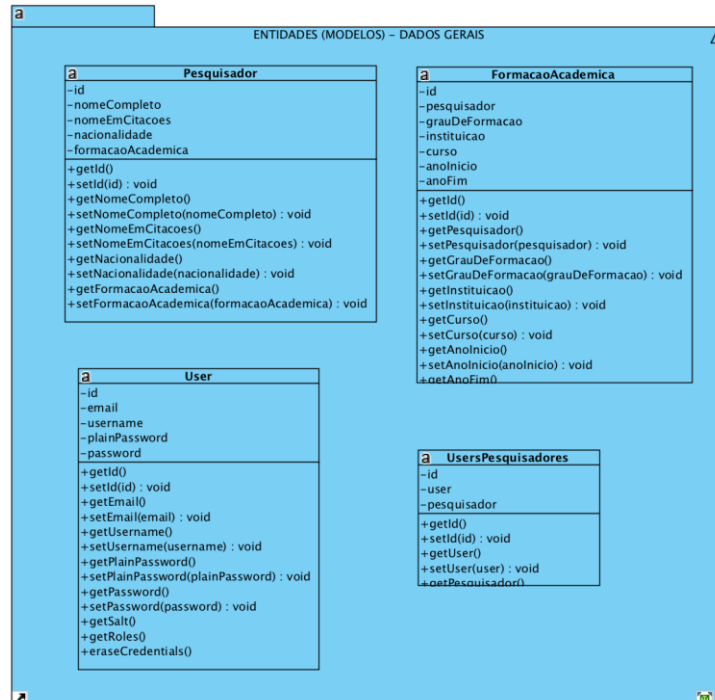
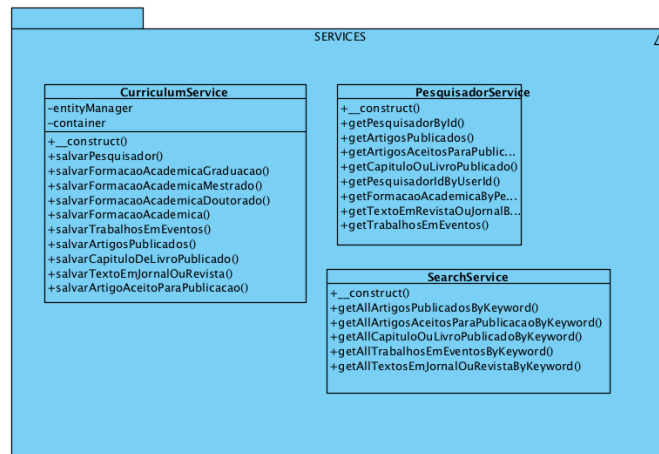
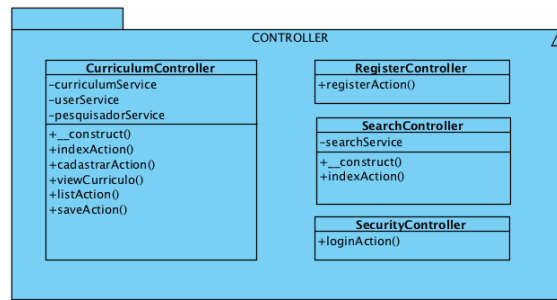
STANDING, C. Methodologies for developing Web applications. **Information and Software Technology**, v. 44, n.3, 2002.

ANEXO A - ORDEM DE PREENCHIMENTO DO ETL E DESCRIÇÃO DE CONTEÚDO.

Ordem de preenchimento	Tabela	Descrição
-	user	Dados sobre o usuário do sistema
-	user_pesquisadores	Armazena a associação entre usuário do sistema e pesquisadores salvos em seu perfil.
1	pesquisador	Dados gerais do pesquisador.
2	formacao_academica	Dados sobre a formação acadêmica do pesquisador.
3	artigo_aceito_para_publicacao	Dados sobre todos os artigos aceitos para publicação.
4	autores_artigo_aceito_para_publicacao	Dados sobre os autores de artigos aceitos para publicação.
5	areas_do_conhecimento_artigo_aceito_para_publicacao	Dados sobre as áreas de conhecimento de artigos aceitos para publicação.
6	setores_de_atividade_artigo_aceito_para_publicacao	Dados sobre os setores de atividade de artigos aceitos para publicação.
7	palavras_chave_artigo_aceito_para_publicacao	Armazenamento de palavras-chave para cada artigo aceito para publicação.
8	artigos_publicados	Dados sobre todos os artigos publicados.
9	autores_artigos_publicados	Dados sobre os autores de artigos publicados.
10	areas_do_conhecimento_artigo	Dados sobre as áreas de conhecimento de artigos publicados.
11	setores_de_atividade_artigos_publicados	Dados sobre os setores de atividade de artigos publicados.
12	palavras_chave_artigos_publicados	Armazenamento de palavras-chave para cada artigos

		publicados.
13	capitulo_de_livro_publicado	Dados sobre todos os capítulos ou livros publicados.
14	autores_capitulo_de_livro_publicado	Dados sobre os autores de capítulos de livros publicados.
15	areas_do_conhecimento_capitulo_de_livro	Dados sobre as áreas de conhecimento de capítulo de livros publicados.
16	setores_de_atividade_capitulo_de_livro_publicado	Dados sobre os setores de atividade de capítulo de livros publicados.
17	palavras_chave_capitulo_de_livro_publicado	Armazenamento de palavras-chave para capítulos de livros publicados.
18	texto_em_jornal_ou_revista_publicado	Dados sobre todos os textos em jornais ou revistas.
19	autores_texto_em_jornal_ou_revista	Dados sobre os autores de textos em jornais ou revistas.
20	areas_do_conhecimento_texto_em_jornal_ou_revista	Dados sobre as áreas de conhecimento de textos em jornais ou revistas.
21	setores_de_atividade_texto_em_jornal_ou_revista	Dados sobre os setores de atividade de textos em jornais ou revistas.
22	palavras_chave_texto_em_jornal_ou_revista	Armazenamento de palavras-chave para textos em jornais ou revistas.
23	trabalhos_em_eventos	Dados sobre todos os textos em trabalhos em eventos.
24	autores_trabalhos_em_eventos	Dados sobre os autores de trabalhos em eventos.
25	areas_do_conhecimento_trabalho_em_evento	Dados sobre as áreas de conhecimento de trabalhos em eventos.
26	setores_de_atividade_trabalho_em_evento	Dados sobre os setores de atividade de trabalhos em eventos.
27	palavras_chave_trabalho_em_evento	Armazenamento de palavras-chave para trabalhos em eventos.

ANEXO B – DIAGRAMAS UML.



ArtigoAceitoParaPublicacao
-id -pesquisador -sequenciaProducao -natureza -tituloDoArtigo -anoDoArtigo -paisDePublicacao -idioma -meioDeDivulgacao -flagRelevancia -doi -tituloDoArtigoEmIngles -flagDivulgacaoCientifica -tituloDoPeriodicoOuRevista -issn -volume -fasciculo -serie -paginaInicial -paginaFinal -localDePublicacao -autores -areasDoConhecimento -setoresDeAtividade -informacaoAdicional -palavrasChave +getId() : void +setId(id) : void +getPesquisador() : void +setPesquisador(pesquisador) : void +getSequenciaProducao() : void +setSequenciaProducao(sequenciaProducao) : void +getNatureza() : void +setNatureza(natureza) : void +getTituloDoArtigo() : void +setTituloDoArtigo(tituloDoArtigo) : void +getAnoDoArtigo() : void +setAnoDoArtigo(anoDoArtigo) : void +getPaísDePublicacao() : void +setPaísDePublicacao(paisDePublicacao) : void +getIdiom() : void +setIdiom(idiom) : void +getMeioDeDivulgacao() : void +setMeioDeDivulgacao(medioDeDivulgacao) : void +getFlagRelevancia() : void +setFlagRelevancia(flagRelevancia) : void +getDoi() : void +setDoi(doi) : void +getTituloDoArtigoEmIngles() : void +setTituloDoArtigoEmIngles(tituloDoArtigoEmIngles) : void +getFlagDivulgacaoCientifica() : void +setFlagDivulgacaoCientifica(flagDivulgacaoCientifica) : void +getTituloDoPeriodicoOuRevista() : void +setTituloDoPeriodicoOuRevista(tituloDoPeriodicoOuRevista) : void +getIssn() : void +setIssn(issn) : void +getVolume() : void +setVolume(volume) : void +getFasciculo() : void +setFasciculo(fasciculo) : void +getSerie() : void +setSerie(serie) : void +getPaginaInicial() : void +setPaginaInicial(paginaInicial) : void +getPaginaFinal() : void +setPaginaFinal(paginaFinal) : void +getLocalDePublicacao() : void +setLocalDePublicacao(localDePublicacao) : void +getAutores() : void +setAutores(autores) : void +getAreasDoConhecimento() : void +setAreasDoConhecimento(areasDoConhecimento) : void +getSetoresDeAtividade() : void +setSetoresDeAtividade(setoresDeAtividade) : void +getInformacaoAdicional() : void +setInformacaoAdicional(informacaoAdicional) : void +getPalavrasChave() : void +setPalavrasChave(palavrasChave) : void +_construct()

ArtigosPublicados
-id -pesquisador -sequenciaProducao -natureza -tituloDoArtigo -anoDoArtigo -paisDePublicacao -idioma -meioDeDivulgacao -flagRelevancia -doi -tituloDoArtigoEmIngles -flagDivulgacaoCientifica -tituloDoPeriodicoOuRevista -issn -volume -fasciculo -serie -paginaInicial -paginaFinal -localDePublicacao -autores -areasDoConhecimento -setoresDeAtividade -informacaoAdicional -palavrasChave +getId() : void +setId(id) : void +getPesquisador() : void +setPesquisador(pesquisador) : void +getSequenciaProducao() : void +setSequenciaProducao(sequenciaProducao) : void +getNatureza() : void +setNatureza(natureza) : void +getTituloDoArtigo() : void +setTituloDoArtigo(tituloDoArtigo) : void +getAnoDoArtigo() : void +setAnoDoArtigo(anoDoArtigo) : void +getPaísDePublicacao() : void +setPaísDePublicacao(paisDePublicacao) : void +getIdiom() : void +setIdiom(idiom) : void +getMeioDeDivulgacao() : void +setMeioDeDivulgacao(medioDeDivulgacao) : void +getFlagRelevancia() : void +setFlagRelevancia(flagRelevancia) : void +getDoi() : void +setDoi(doi) : void +getTituloDoArtigoEmIngles() : void +setTituloDoArtigoEmIngles(tituloDoArtigoEmIngles) : void +getFlagDivulgacaoCientifica() : void +setFlagDivulgacaoCientifica(flagDivulgacaoCientifica) : void +getTituloDoPeriodicoOuRevista() : void +setTituloDoPeriodicoOuRevista(tituloDoPeriodicoOuRevista) : void +getIssn() : void +setIssn(issn) : void +getVolume() : void +setVolume(volume) : void +getFasciculo() : void +setFasciculo(fasciculo) : void +getSerie() : void +setSerie(serie) : void +getPaginaInicial() : void +setPaginaInicial(paginaInicial) : void +getPaginaFinal() : void +setPaginaFinal(paginaFinal) : void +getLocalDePublicacao() : void +setLocalDePublicacao(localDePublicacao) : void +getAutores() : void +setAutores(autores) : void +getAreasDoConhecimento() : void +setAreasDoConhecimento(areasDoConhecimento) : void +getSetoresDeAtividade() : void +setSetoresDeAtividade(setoresDeAtividade) : void +getInformacaoAdicional() : void +setInformacaoAdicional(informacaoAdicional) : void +getPalavrasChave() : void +setPalavrasChave(palavrasChave) : void +_construct()

CapituloDeLivroPublicado
-id -pesquisador -sequenciaProducao -tipo -tituloDoCapituloDoLivro -ano -paisDePublicacao -idioma -meioDeDivulgacao -flagRelevancia -doi -tituloDoCapituloDoLivroEmIngles -flagDivulgacaoCientifica -tituloDoLivro -numeroDeVolumes -paginaInicial -paginaFinal -isbn -organizadores -numeroDeEdicaoRevisao -numeroDeSerie -cidadeDaEditora -nomeDaEditora -autores -areasDoConhecimento -setoresDeAtividade -palavrasChave -informacaoAdicional +_construct() : void +getId() : void +setId(id) : void +getPesquisador() : void +setPesquisador(pesquisador) : void +getSequenciaProducao() : void +setSequenciaProducao(sequenciaProducao) : void +getTipo() : void +setTipo(tipo) : void +getTituloDoCapituloDoLivro() : void +setTituloDoCapituloDoLivro(tituloDoCapituloDoLivro) : void +getAno() : void +setAno(ano) : void +getPaísDePublicacao() : void +setPaísDePublicacao(paisDePublicacao) : void +getIdiom() : void +setIdiom(idiom) : void +getMeioDeDivulgacao() : void +setMeioDeDivulgacao(medioDeDivulgacao) : void +getFlagRelevancia() : void +setFlagRelevancia(flagRelevancia) : void +getDoi() : void +setDoi(doi) : void +getTituloDoCapituloDoLivroEmIngles() : void +setTituloDoCapituloDoLivroEmIngles(tituloDoCapituloDoLivroEmIngles) : void +getFlagDivulgacaoCientifica() : void +setFlagDivulgacaoCientifica(flagDivulgacaoCientifica) : void +getTituloDoLivro() : void +setTituloDoLivro(tituloDoLivro) : void +getNumeroDeVolumes() : void +setNumeroDeVolumes(numeroDeVolumes) : void +getPaginaInicial() : void +setPaginaInicial(paginaInicial) : void +getPaginaFinal() : void +setPaginaFinal(paginaFinal) : void +getIsbn() : void +setIsbn(isbn) : void +getOrganizadores() : void +setOrganizadores(organizadores) : void +getNumeroDeEdicaoRevisao() : void +setNumeroDeEdicaoRevisao(numeroDeEdicaoRevisao) : void +getNumeroDeSerie() : void +setNumeroDeSerie(numeroDeSerie) : void +getCidadeDaEditora() : void +setCidadeDaEditora(cidadeDaEditora) : void +getNomeDaEditora() : void +setNomeDaEditora(nomeDaEditora) : void +getAutores() : void +setAutores(autores) : void +getAreasDoConhecimento() : void +setAreasDoConhecimento(areasDoConhecimento) : void +getSetoresDeAtividade() : void +setSetoresDeAtividade(setoresDeAtividade) : void +getPalavrasChave() : void +setPalavrasChave(palavrasChave) : void +getInformacaoAdicional() : void +setInformacaoAdicional(informacaoAdicional) : void

TextoEmJornalOuRevistaPublicado
-id -pesquisador -sequenciaProducao -natureza -tituloDoTexto -anoDoTexto -paisDePublicacao -idioma -meioDeDivulgacao -flagRelevancia -doi -tituloDoTextoEmIngles -flagDivulgacaoCientifica -tituloDoJornalOuRevista -issn -formatoDataDePublicacao -dataDePublicacao -volume -paginaInicial -paginaFinal -localDePublicacao -autores -areasDoConhecimento -setoresDeAtividade -palavrasChave -informacaoAdicional +_construct() : void +getId() : void +setId(id) : void +getPesquisador() : void +setPesquisador(pesquisador) : void +getSequenciaProducao() : void +setSequenciaProducao(sequenciaProducao) : void +getNatureza() : void +setNatureza(natureza) : void +getTituloDoTexto() : void +setTituloDoTexto(tituloDoTexto) : void +getAnoDoTexto() : void +setAnoDoTexto(anoDoTexto) : void +getPaísDePublicacao() : void +setPaísDePublicacao(paisDePublicacao) : void +getIdiom() : void +setIdiom(idiom) : void +getMeioDeDivulgacao() : void +setMeioDeDivulgacao(medioDeDivulgacao) : void +getFlagRelevancia() : void +setFlagRelevancia(flagRelevancia) : void +getDoi() : void +setDoi(doi) : void +getTituloDoTextoEmIngles() : void +setTituloDoTextoEmIngles(tituloDoTextoEmIngles) : void +getFlagDivulgacaoCientifica() : void +setFlagDivulgacaoCientifica(flagDivulgacaoCientifica) : void +getTituloDoJornalOuRevista() : void +setTituloDoJornalOuRevista(tituloDoJornalOuRevista) : void +getIssn() : void +setIssn(issn) : void +getFormatoDataDePublicacao() : void +setFormatoDataDePublicacao(formatoDataDePublicacao) : void +getDataDePublicacao() : void +setDataDePublicacao(dataDePublicacao) : void +getVolume() : void +setVolume(volume) : void +getPaginaInicial() : void +setPaginaInicial(paginaInicial) : void +getPaginaFinal() : void +setPaginaFinal(paginaFinal) : void +getLocalDePublicacao() : void +setLocalDePublicacao(localDePublicacao) : void +getAutores() : void +setAutores(autores) : void +getAreasDoConhecimento() : void +setAreasDoConhecimento(areasDoConhecimento) : void +getSetoresDeAtividade() : void +setSetoresDeAtividade(setoresDeAtividade) : void +getPalavrasChave() : void +setPalavrasChave(palavrasChave) : void +getInformacaoAdicional() : void +setInformacaoAdicional(informacaoAdicional) : void

TrabalhosEmEventos
-id -pesquisador -sequenciaProducao -natureza -tituloDoTrabalho -anoDoTrabalho -paisDoEvento -idioma -meioDeDivulgacao -flagRelevancia -doi -tituloDoTrabalhoEmIngles -flagDivulgacaoCientifica -classificacaoDoEvento -nomeDoEvento -cidadeDoEvento -anoDeRealizacao -tituloDoAnaisOuProceedings -volume -fasciculo -serie -paginaInicial -paginaFinal -gen -nomeDaEditora -cidadeDaEditora -autores -areasDoConhecimento -setoresDeAtividade -palavrasChave +_construct() : void +getId() : void +setId(id) : void +getPesquisador() : void +setPesquisador(pesquisador) : void +getSequenciaProducao() : void +setSequenciaProducao(sequenciaProducao) : void +getNatureza() : void +setNatureza(natureza) : void +getTituloDoTrabalho() : void +setTituloDoTrabalho(tituloDoTrabalho) : void +getAnoDoTrabalho() : void +setAnoDoTrabalho(anoDoTrabalho) : void +getPaísDoEvento() : void +setPaísDoEvento(paisDoEvento) : void +getIdiom() : void +setIdiom(idiom) : void +getMeioDeDivulgacao() : void +setMeioDeDivulgacao(medioDeDivulgacao) : void +getFlagRelevancia() : void +setFlagRelevancia(flagRelevancia) : void +getDoi() : void +setDoi(doi) : void +getTituloDoTrabalhoEmIngles() : void +setTituloDoTrabalhoEmIngles(tituloDoTrabalhoEmIngles) : void +getFlagDivulgacaoCientifica() : void +setFlagDivulgacaoCientifica(flagDivulgacaoCientifica) : void +getClassificacaoDoEvento() : void +setClassificacaoDoEvento(classificacaoDoEvento) : void +getNomeDoEvento() : void +setNomeDoEvento(nomeDoEvento) : void +getCidadeDoEvento() : void +setCidadeDoEvento(cidadeDoEvento) : void +getAnoDeRealizacao() : void +setAnoDeRealizacao(anoDeRealizacao) : void +getTituloDoAnaisOuProceedings() : void +setTituloDoAnaisOuProceedings(tituloDoAnaisOuProceedings) : void +getVolume() : void +setVolume(volume) : void +getFasciculo() : void +setFasciculo(fasciculo) : void +getSerie() : void +setSerie(serie) : void +getPaginaInicial() : void +setPaginaInicial(paginaInicial) : void +getPaginaFinal() : void +setPaginaFinal(paginaFinal) : void +getGen() : void +setGen(isbn) : void +getNomeDaEditora() : void +setNomeDaEditora(nomeDaEditora) : void +getCidadeDaEditora() : void +setCidadeDaEditora(cidadeDaEditora) : void +getAutores() : void +setAutores(autores) : void +getAreasDoConhecimento() : void +setAreasDoConhecimento(areasDoConhecimento) : void +getSetoresDeAtividade() : void +setSetoresDeAtividade(setoresDeAtividade) : void +getPalavrasChave() : void +setPalavrasChave(palavrasChave) : void

