

UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ
CÂMPUS CORNÉLIO PROCÓPIO
DIRETORIA DE GRADUAÇÃO E EDUCAÇÃO PROFISSIONAL
TECNOLOGIA EM ANÁLISE E DESENVOLVIMENTO DE SISTEMAS

MARIANNA JUNQUEIRA FERREIRA

**EXTRAÇÃO DE PRODUÇÃO BIBLIOGRÁFICA A PARTIR DO CURRÍCULO
LATTES**

TRABALHO DE CONCLUSÃO DE CURSO

CORNÉLIO PROCÓPIO
2015

MARIANNA JUNQUEIRA FERREIRA

EXTRAÇÃO DE PRODUÇÃO BIBLIOGRÁFICA A PARTIR DO CURRÍCULO LATTES

Trabalho de Conclusão de Curso de graduação, apresentado à disciplina Trabalho de Conclusão de Curso, do curso Tecnologia em Análise e Desenvolvimento de Sistemas da Universidade Tecnológica Federal do Paraná – UTFPR, como requisito parcial para a obtenção do título de Tecnólogo.

Orientador: Prof. Dr. Alessandro Botelho Bovo

CORNÉLIO PROCÓPIO

2015

Dedico este trabalho à minha mãe Regina, que sempre me deu exemplo de garra para conquistar tudo o que desejei e até mesmo o que não acreditava ser capaz e também pela paciência e carinho. À minha avó, Maria de Lourdes, que sempre esteve comigo, amparando e cuidando.

AGRADECIMENTOS

Agradeço a Deus pela vida e pela oportunidade de realizar mais uma conquista.

Aos meus amigos que me apoiaram para conseguir concluir este trabalho, fazendo acreditar que era possível.

À Secretaria do Curso, pela cooperação.

Gostaria de deixar registrado, também, o meu reconhecimento à minha família, pois acredito que sem o apoio deles, seria muito difícil vencer esse desafio.

Enfim, a todos os que, por algum motivo, contribuíram para a realização deste trabalho.

“Não sabendo que era impossível, ele foi lá e fez.” (Jean Cocteau)

RESUMO

FERREIRA, Junqueira Marianna. **Extração de Produção Bibliográfica a partir do Currículo Lattes**. 2015. 49 f. Trabalho de Conclusão de Curso – Tecnologia em Análise e Desenvolvimento de Sistemas. Universidade Tecnológica Federal do Paraná. Cornélio Procópio, 2015.

A Plataforma Lattes foi idealizada pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) com o objetivo de apoiar as atividades de Ciência, Tecnologia e Inovação (CT&I) do Brasil. O Currículo Lattes, seu principal componente, é considerado o padrão nacional de registro de toda atividade acadêmica de pesquisadores, e é adotado pela maioria das instituições de ensino e institutos de pesquisa. O CNPq disponibiliza o currículo Lattes de pesquisadores em formato XML. Este formato, por se tratar de um padrão para troca de dados entre sistemas computacionais, pode ser utilizado para se obter as informações sobre a produção CT&I de um determinado conjunto de pesquisadores. Assim, este trabalho apresenta o projeto e a implementação de um sistema que realiza a extração de informações sobre a produção científica e tecnológica constantes em arquivos XML's de currículos Lattes de um determinado grupo de pesquisadores e apresenta um conjunto de relatórios sobre essas informações. Os relatórios gerados podem ser utilizados, por exemplo, por gestores de Instituições Ensino e Pesquisa ou coordenadores de Programas de Pós-Graduação.

Palavras-chave: Plataforma Lattes, Currículo Lattes, Produção CT&I, Arquivo XML

ABSTRACT

FERREIRA, Junqueira Marianna. **Extração de Produção Bibliográfica a partir do Currículo Lattes**. 2015. 49 f. Trabalho de Conclusão de Curso – Tecnologia em Análise e Desenvolvimento de Sistemas. Universidade Tecnológica Federal do Paraná. Cornélio Procópio, 2015.

The Lattes Platform was designed by the The Brazilian National Council for Scientific and Technological Development (CNPq) in order to support the activities of Science, Technology and Innovation (ST&I) of Brazil. The Lattes curriculum, its main component, is considered the national registration standard of all academic activities of researchers, and is adopted by most educational institutions and research institutes. The CNPq provides the Lattes curriculum of researchers in XML format. This format, that is a standard for exchanging data between computer systems, may be used to obtain ST&I production data from a group of researchers. This work presents the design and implementation of a system that performs the extraction of ST&I information from those XML files of a particular group of researchers and presents reports on such information. These generated reports can be used, for example, by managers of educational institutions or coordinators of the graduate programs.

Keywords: Lattes Plataform, Lattes Curriculum, ST&I Production, XML File

LISTA DE FIGURAS

Figura 1 – Tela Busca de Currículo Lattes

Figura 2 – Exemplo de Currículo Lattes

Figura 3 – Currículos Lattes por Região do Brasil

Figura 4 – Diagrama de fluxo de informações considerado no scriptLattes

Figura 5 – Tipos de produção acadêmica extraídas do Currículo Lattes

Figura 6 – Ciclo da Metodologia Scrum

Figura 7 – Padrão MVC

Figura 8 – Aplicação do Padrão MVC

Figura 9 – Trecho XML de um Currículo Lattes

Figura 10 – Diagrama de Banco de Dados

Figura 11 – Tela Inicial do Sistema

Figura 12 – Tela Realizar Extração

Figura 13 – Selecionar XML

Figura 14 – Extrações Realizadas

Figura 15 – Relatório Extração Grupo PPGBIOINFO

Figura 16 – Tela Relatórios

Figura 17 - Relatório Produção Bibliográfica

Figura 18 - Relatório Pesquisador x Tipo

Figura 19 - Tela Ano – Produção Bibliográfica

Figura 20 - Relatório Produções por Ano

Figura 21 - Tela Tipos de Produção Bibliográfica

Figura 22 - Relatório Tipos de Produção Bibliográfica

LISTA DE TABELAS

Tabela 1 – Funcionalidades do Sistema

LISTA DE SIGLAS

UTFPR – Universidade Tecnológica Federal do Paraná

CNPQ – Conselho Nacional de Desenvolvimento Científico e Tecnológico

XML – *Extensible Markup Language*

UML – *Unified Modeling Language*

CAPES – Coordenação de Aperfeiçoamento de Pessoal de Nível Superior

UEL – Universidade Estadual de Londrina

CTC-ES - Conselho Técnico-Científico da Educação Superior

CONSCIENTIAS - Comunidade para Ontologias em Ciência, Tecnologia e
Informações de Aperfeiçoamento de Nível Superior

DTD - *Document Type Definition*

PPGBIOINFO - Programa de Pós-Graduação em Bioinformática

PL – Plataforma Lattes

ISSN- *International Standard Serial Number*

ISBN- *International Standard Book Number*

LMPL - Linguagem de Marcação da Plataforma Lattes

WEB – *World Wide Web*

PPG – Programa de Pós Graduação

CL – Currículo Lattes

HTML - *HyperText Markup Language*

SNPG – Sistema Nacional de Pós-Graduação

SGBD – Sistema Gerenciador de Banco de Dados

GNU – *General Public License*

IDE – Ambiente de Desenvolvimento Integrado

JDBC – *Java Database Connectivity*

JVM – *Java Virtual Machine*

IAPAR – Instituto Agronômico do Paraná

BD – Banco de Dados

CT&I – Ciência, Tecnologia e Inovação

OPAS - Organização Pan-Americana da Saúde

HTM – *Hypertext Markup*

CEP – Código de Endereçamento Postal

UF – Unidade da Federação

GPL – *General Public License*

CSS – *Cascading Style Sheets*

PDF – *Portable Document Format*

DOCX – Extensão Word

RTF – Rich Text Format

XLS – Extensão Excel

CSV – *Comma-separated Values*

MVC – *Model View Controller*

SUMÁRIO

1	INTRODUÇÃO	14
1.1	JUSTIFICATIVA	15
1.2	DELIMITAÇÕES DO TEMA	15
1.3	OBJETIVOS.....	16
1.1.3	OBJETIVO GERAL.....	16
1.1.4	OBJETIVOS ESPECÍFICOS.....	16
1.4	ORGANIZAÇÃO DO TRABALHO.....	17
2	FUNDAMENTAÇÃO TEÓRICA.....	18
2.1	PLATAFORMA LATTES.....	18
2.2	SCRIPTLATTES	21
3	FERRAMENTAS E TECNOLOGIAS UTILIZADAS	24
3.1	FERRAMENTAS.....	24
3.2	TECNOLOGIAS	25
4	DESENVOLVIMENTO	27
4.1	METODOLOGIA	27
4.2	ARQUITETURA DO SISTEMA.....	28
4.3	FUNCIONALIDADES	30
4.4	EXTRAÇÃO E ARMAZENAMENTO DA PRODUÇÃO BIBLIOGRÁFICA	31
4.5	APRESENTAÇÃO DO SISTEMA.....	35
5	CONSIDERAÇÕES FINAIS	46

1 INTRODUÇÃO

A Plataforma Lattes foi idealizada pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) com o objetivo de apoiar as atividades de Ciência, Tecnologia e Inovação (CT&I) do Brasil (CNPq, 2015). Sua finalidade era integrar os sistemas de informação das agências federais e promover a racionalização do processo de gestão de CT&I. Iniciada em 1999, com o lançamento do sistema Currículo Lattes, a Plataforma Lattes é atualmente formada de bases de dados, sistemas de informação, diretórios de serviços e portais Web. Trata-se de um recurso que forneceu uma nova maneira de ver as informações em Ciência, Tecnologia e Inovação no Brasil (SABBATINI, 2015). É utilizada por gestores, por técnicos de governo e pela comunidade científica em geral.

Já o Currículo Lattes é considerado um padrão nacional no registro de toda atividade acadêmica de pesquisadores, e é adotado pela maioria das instituições de ensino, bem como institutos de pesquisa. É abrangente, extenso e detalha todos os aspectos relacionados a carreira do profissional em questão (CNPq, 2015). É possível visualizar cada currículo Lattes individualmente, em formato HTML, no sítio da Plataforma Lattes¹. Nesse mesmo sítio, também é possível fazer o *download* do mesmo currículo em formato XML. Além disso, o CNPq também faz acordos institucionais para que se possa realizar a extração, em lote, de vários currículos Lattes (em formato XML) de pesquisadores e estudantes de uma determinada instituição.

Independentemente da forma como se obtenha o XML do currículo Lattes, por se tratar de um formato padrão para troca de dados entre sistemas computacionais, pode-se utilizar esses arquivos para se obter as informações sobre a produção CT&I de um determinado conjunto de pesquisadores. Assim, pretende-se neste trabalho implementar um sistema *desktop* que realize a extração de informações sobre a produção científica e tecnológica constantes em arquivos XML-s de currículos Lattes, de um determinado grupo de pesquisadores, e apresente um conjunto de relatórios sobre essas informações. Os relatórios gerados podem ser utilizados, por exemplo,

¹ <http://lattes.cnpq.br>

por gestores de Instituições Ensino e Pesquisa ou coordenadores de Programas de Pós-Graduação.

1.1 JUSTIFICATIVA

O conjunto de XML-s utilizados para a extração de produções científicas e tecnológicas podem ser compostos, por exemplo, pelos arquivos dos currículos dos pesquisadores (e mesmo estudantes) de uma determinada Instituição de Pesquisa e Ensino, pelos pesquisadores de um determinado Grupo de Pesquisa ou pelos pesquisadores de um determinado Programa de Pós-Graduação.

No caso específico de um Programa de Pós-Graduação, esses relatórios sobre a produção científica e tecnológica dos membros do programa podem ser utilizados para gerar indicadores de produtividade que são utilizados como nota de corte para entrada ou permanência de membros no programa.

Ainda no contexto de Programa de Pós-Graduação, o uso de relatórios com informações sobre a produção científica e tecnológica dos membros do programa podem ser usados para auxiliar os coordenadores de curso no envio de informações à Capes, no processo de “Coleta” que a referida agência realiza todos os anos.

Por fim, pelo fato do sistema não ter regras específicas de uma área do conhecimento, ele pode ser utilizado em áreas totalmente distintas dentro das instituições, desde que seja utilizado como entrada o XML do Currículo Lattes.

1.2 DELIMITAÇÕES DO TEMA

Em relação à extração das informações do XML sobre produções científicas e tecnológicas, foi seguida a padronização do XML da Comunidade LMPL² pertencente ao CNPq, que desenvolve ontologias que fazem o intercâmbio de informações entre agências de fomento e instituições usuárias ligadas ao tema Ciência, Tecnologia,

² <http://impl.cnpq.br/impl/>

Inovação e Informações de Nível Superior, e por meio dessas ontologias, a coleta de informações torna-se mais ágil.

O sistema proposto neste trabalho foi testado com os XML-s dos currículos Lattes dos pesquisadores do Programa de Pós-graduação em Bioinformática (PPGBIOINFO)³ da Universidade Tecnológica Federal do Paraná (UTFPR), campus Cornélio Procópio. O PPGBIOINFO conta com a participação de pesquisadores da EMBRAPA-Soja, IAPAR, EMBRAPA-Café e UEL.

Um último aspecto a considerar em termos de delimitação do tema, trata de quais tipos de informações a versão atual do sistema permite extrair. Apesar de ser possível se extrair todos os tipos de produção científica e tecnológica (produção bibliográfica, técnica, etc.), por limitações de tempo optou-se apenas pela implementação do módulo que extrai as produções bibliográficas (artigos, livros, resumos, etc.).

1.3 OBJETIVOS

1.1.3 OBJETIVO GERAL

Desenvolver um sistema *desktop*, que permita realizar a extração de dados através de arquivos XML, sendo que cada XML pertence a um pesquisador.

1.1.4 OBJETIVOS ESPECÍFICOS

Extrair informações das produções bibliográficas de cada XML, de acordo com o(s) arquivo(s) que são selecionados pelo usuário, realizar a inserção das mesmas no Banco de Dados e permitir a visualização das extrações realizadas com seu nome e data respectivos.

Permitir que o usuário acesse relatórios em relação as informações de cada produção inserida e de seus respectivos pesquisadores de uma forma mais ágil.

³ <http://www.utfpr.edu.br/cornelioprocopio/ppgbioinfo>

1.4 ORGANIZAÇÃO DO TRABALHO

Além do capítulo de Introdução, este trabalho está organizado da seguinte forma: Capítulo 2: Fundamentação Teórica: apresenta duas plataformas que utilizam o Currículo Lattes como fonte de informações para o seu desenvolvimento e como cada uma atua no meio Científico e Tecnológico. O Capítulo 3, refere-se as ferramentas e tecnologias utilizadas durante o desenvolvimento do sistema, com uma explicação de como foram aplicadas. Capítulo 4: Desenvolvimento: apresenta a metodologia aplicada no desenvolvimento do sistema, o que ela disponibiliza e como foi utilizada, a arquitetura de sistema e suas camadas, as funcionalidades que estão disponíveis no sistema e o processo realizado para obter as informações necessárias de arquivos XML, que contemplam a extração e armazenamento.

2 FUNDAMENTAÇÃO TEÓRICA

A primeira parte deste capítulo envolve uma breve apresentação da Plataforma Lattes e de seu principal componente, o Currículo Lattes. Na segunda parte, será apresentada a descrição de uma ferramenta que também se utiliza de dados da Plataforma Lattes, o scriptLattes.

2.1 PLATAFORMA LATTES

A Plataforma Lattes (PL)⁴ foi concebida pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq)⁵ com a finalidade de apoiar as atividades de Ciência, Tecnologia e Inovação (CT&I) do Brasil (CNPq, 2015). Seu objetivo inicial era integrar os sistemas de informação das agências federais e promover a racionalização do processo de gestão de CT&I. Iniciada em 1999, com o lançamento do seu primeiro componente (o sistema Currículo Lattes), a PL é atualmente formada de bases de dados, sistemas de informação, diretórios de serviços e portais Web. Trata-se de um recurso que forneceu uma nova maneira de ver as informações em C&T no Brasil (SABBATINI, 2015). É utilizada por gestores, por técnicos de governo e pela comunidade científica em geral.

O Currículo Lattes é considerado um padrão nacional no registro de toda atividade acadêmica de pesquisadores, e é adotado pela maioria das instituições de ensino, bem como institutos de pesquisa. É abrangente, extenso e detalha todos os aspectos relacionados a carreira do profissional em questão (CNPq, 2015). É de fácil acesso ao usuário, pois “acordos institucionais têm sido realizados visando ampliar a visibilidade, utilidade e confiabilidade das informações constantes na Plataforma Lattes, e as instituições parceiras têm se beneficiado ao ampliar os recursos oferecidos aos seus usuários web” (CNPq, 2015).

Dentro deste contexto de disponibilização e padronização, o CNPq, juntamente com a Bireme/OPAS (Organização Pan-Americana da Saúde), criou no ano de 2002 uma versão em língua espanhola do Currículo Lattes. Nessa ocasião, foi instituída a

⁴ <http://lattes.cnpq.br/>

⁵ <http://www.cnpq.br/>

rede SCIENTE (STELA, 2002). Tal rede é formada por organizações nacionais de ciência e tecnologia, e outros organismos internacionais, visando promover a padronização e a troca de informação entre membros participantes da área científica em seus respectivos países.

A padronização das informações do Currículo Lattes foi realizada pela Comunidade para Ontologias em Ciência, Tecnologia e Informações de Aperfeiçoamento de Nível Superior (CONSCIENTIAS). A CONSCIENTIAS foi criada para desenvolver ontologias, em linguagem XML, que têm por finalidade principal o estabelecimento de uma forma comum de troca de informações entre agências de fomento e suas instituições usuárias.

É apresentada na Figura 1 a tela na qual se pode fazer a busca por currículos por meio do nome do pesquisador e/ou estudante.

The image shows the search interface of the Currículo Lattes system. At the top left is the CNPq logo. The main heading is 'Buscar Currículo Lattes (Busca Simples)'. Below this, there is a search bar with the text 'Alessandro Botelho Bovo'. To the right of the search bar is a link for 'Busca Avançada'. Underneath the search bar, there are several filter options: 'Nas bases' with 'Doutores' checked and 'Demais pesquisadores (Mestres, Graduados, Estudantes, Técnicos, etc.)' unchecked; 'Nacionalidade' with 'Brasileira' checked and 'Estrangeira' unchecked; and 'País de nacionalidade' set to 'Todos'. There are also buttons for 'Filtros' and 'Preferências'. At the bottom right, there is a blue 'Buscar' button.

Figura 1 – Tela Busca de Currículo Lattes
Fonte: <http://buscatextual.cnpq.br/buscatextual>

Na Figura 2 é apresentada a tela com um Currículo Lattes, como exemplo.



CNPq
Conselho Nacional de Desenvolvimento Científico e Tecnológico

Curriculo Lattes

Dados gerais | Formação | Atuação | +

Marianna Junqueira Ferreira

Endereço para acessar este CV: <http://lattes.cnpq.br/5457562394624969>

Última atualização do currículo em 19/03/2015

Possui ensino-medio-segundo-graupela Escola Rui Barbosa(2009). (Texto gerado automaticamente pela aplicação CVLattes)

Identificação

Nome Marianna Junqueira Ferreira

Nome em citações bibliográficas FERREIRA, M. J.

Figura 2 – Exemplo de Currículo Lattes
Fonte: <http://lattes.cnpq.br/5457562394624969>

Também é possível utilizar as ferramentas disponibilizadas no site da PL para ver estatísticas referentes aos dados do Currículos Lattes. Por exemplo, é apresentada na Figura 3 a distribuição dos Currículos Lattes pelas regiões do Brasil.

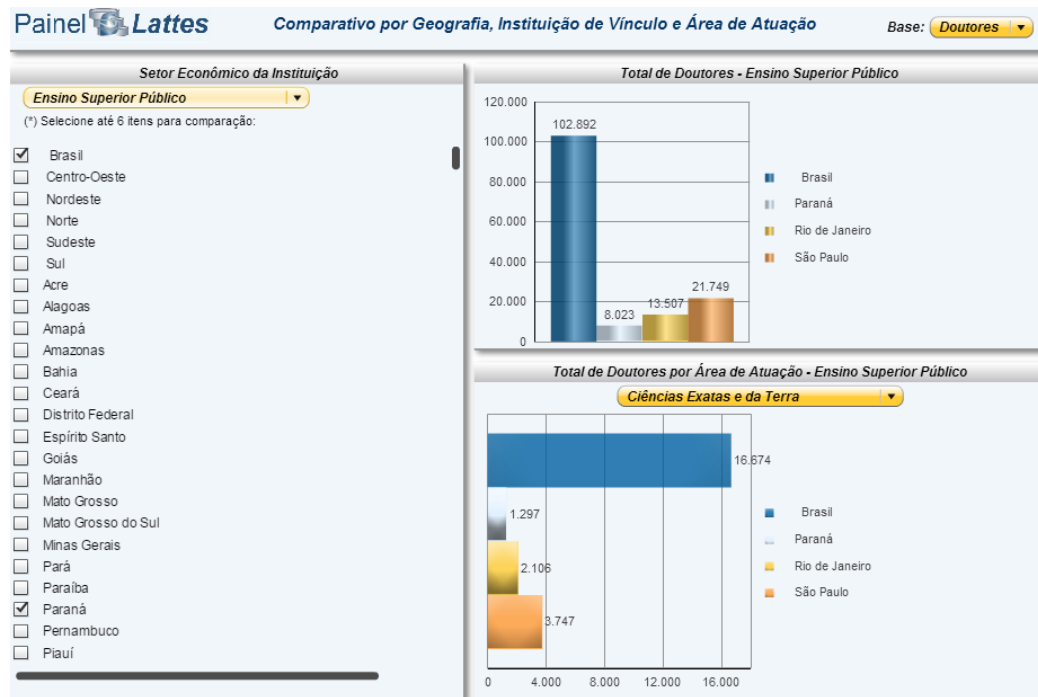


Figura 3 – Currículos Lattes por Região do Brasil
Fonte: <http://estatico.cnpq.br/painelLattes/mapa/>

Por fim, a Plataforma Lattes teve sua importância destacada em artigo publicado pela revista Nature (LANE, 2010). Segundo a autora, Julia Lane - que é diretora do programa Ciência e Política da Inovação, da Fundação Nacional da Ciência, dos Estados Unidos - é importante para o desenvolvimento de um país criar um sistema aberto, apropriado e consistente de medição da produtividade acadêmica.

2.2 SCRIPTLATTES

O scriptLattes⁶ é um software livre que permite a geração de relatórios a partir das informações do Currículo Lattes. Trata-se de uma ferramenta que realiza a extração e compilação automática de produções bibliográficas, técnicas e artísticas, orientações, projetos de pesquisa, prêmios e títulos, grafo de colaborações, e mapa de geolocalização de um conjunto de pesquisadores cadastrados na Plataforma Lattes (MENA-CHALCO; CESAR-JR, 2009; 2013).

Ainda segundo os mesmos autores, o scriptLattes faz o *download* dos arquivos HTML dos currículos Lattes de um determinado grupo de pessoas de interesse, extrai e compila as produções, e remove as produções duplicadas. O próximo passo envolve gerar relatórios com listas de produções e orientações, por tipo, e em ordem cronológica invertida. A ferramenta também permite a geração de redes de coautoria entre os membros do grupo e um mapa de geolocalização dos membros e alunos com orientação concluída.

Na Figura 4 é apresentado o diagrama de fluxo de informações da ferramenta ScriptLattes. Segundo (MENA-CHALCO, CESAR-JR, 2013), A captura dos dados da Plataforma Lattes para o programa ocorre por meio de uma lista de ID-s de currículo lattes que possuem 16 dígitos, estes são utilizados na identificação de cada currículo. Ainda para identificação existe um rótulo, utilizado na visualização do grafo de colaborações e um período de permanência no grupo, todas essas informações são extraídas da Plataforma Lattes. Quando se diz respeito a dados como: nome completo do pesquisador, nome em citações bibliográficas, endereço, entre outros, são extraídos através de um arquivo HTML. Em relação a troca de informações com a plataforma google maps, é realizada para obter, de forma automática parâmetros

⁶ <http://scriptlattes.sourceforge.net/>

como: CEP, UF, nome do país e essas informações são utilizadas na visualização de grafos de coautoria.

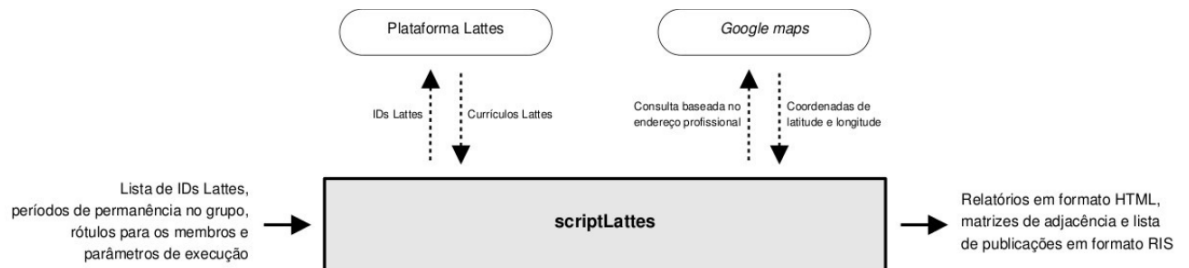


Figura 4 - Diagrama de fluxo de informações considerado no scriptLattes.
Fonte: (MENA-CHALCO; CESAR-JR, 2013)

Na Figura 5 são apresentados os tipos de produção acadêmica extraídos dos currículos Lattes.

A. Produção Bibliográfica
Artigos completos publicados em periódicos Livros publicados/organizados ou edições Capítulos de livros publicados Textos em jornais de notícias/revistas Trabalhos completos publicados em anais de congressos Resumos expandidos publicados em anais de congressos Resumos publicados em anais de congressos Artigos aceitos para publicação Apresentações de trabalho Demais Tipos de produção bibliográfica
B. Produção Técnica
Softwares com registro de patente Softwares sem registro de patente Produtos Tecnológicos Processos ou técnicas Trabalhos técnicos Demais Tipos de produção técnica Total de produção técnica
C. Produção Artística
D. Supervisões e orientações em andamento ou concluídas
Supervisão de Pós Doutorado

Tese de Doutorado Dissertação de Mestrado Trabalho de conclusão de curso de graduação Iniciação Científica Orientações de outra natureza
E. Projetos de Pesquisa
F. Prêmios e Títulos
G. Eventos (participação e organização)

Figura 5 – Tipos de produção acadêmica extraídas do Currículo Lattes

Fonte: (MENA-CHALCO; CESAR-JR, 2013)

O ScriptLattes está sob a licença GNU-GPL e pode ser baixado no seguinte sítio:
<http://scriptlattes.sourceforge.net/>.

3 FERRAMENTAS E TECNOLOGIAS UTILIZADAS

As ferramentas e tecnologias utilizadas no desenvolvimento do projeto estão descritas a seguir:

3.1 FERRAMENTAS

- **SQL SERVER**

O Microsoft SQL Server 2008 “é um sistema de gerenciamento de dados avançado e confiável, que oferece um conjunto rico de recursos, proteção aos dados e desempenho para clientes de aplicativos incorporados, aplicativos leves da Web e repositórios de dados locais” (MSDN, 2009). Essa ferramenta é usada como Sistema Gerenciador de Banco de Dados (SGBD), para armazenar todos os dados em relação a produção bibliográfica, pesquisadores, tipos de produção, ou seja, todos os dados que fazem parte de produção bibliográfica.

- **NET BEANS⁷**

O NetBeans IDE “permite o desenvolvimento rápido e fácil de aplicações desktop Java, móveis e Web e também aplicações HTML5 com HTML, JavaScript e CSS. O IDE também fornece um grande conjunto de ferramentas para desenvolvedores de PHP e C/C++” (NETBEANS, 2015). Ferramenta utilizada no desenvolvimento das classes em linguagem JAVA, as quais fazem a extração dos dados do XML e inserção dos mesmos no banco de dados.

- **iReport⁸**

“É Uma plataforma *open source*, que permite criar relatórios sofisticados com gráficos, imagens. É possível acessar os dados através de uma conexão JDBC e seus

⁷ <https://netbeans.org/>

⁸ <http://community.jaspersoft.com/>

relatórios podem ser publicados nos formatos: PDF, XML, DOCX, entre outros.” (JASPERSOFT, 2015). Utilizada na criação dos relatórios que o sistema disponibiliza ao usuário, sendo que esta ferramenta permite a conexão com o banco de dados utilizado no sistema.

- **VISIO⁹**

“Permite simplificar e transmitir informações complexas através de diagramas vinculados a dados, simplificando a diagramação.” (MSDN, 2015). A ferramenta foi utilizada na criação do fluxograma que representa o ciclo do sistema.

3.2 TECNOLOGIAS

- **Java**

É uma linguagem de programação orientada a objeto desenvolvida na década de 90. Os programas Java podem ser executados em qualquer sistema operacional, desde que a *Java Virtual Machine (JVM)* esteja instalada. Será utilizada para programar as classes responsáveis pela dinamicidade do sistema, as quais extraem as informações relacionadas com produção bibliográfica do XML.

- **XML**

A *Extensible Markup Language (XML)*, é uma linguagem de marcação para a criação de documentos com dados organizados hierarquicamente, tais como textos, banco de dados ou desenhos vetoriais. É classificada como extensível porque permite definir os elementos de marcação. O arquivo XML é o formato no qual os dados estão disponíveis.

⁹ <https://products.office.com/pt-br/Visio/microsoft-visio-2013-top-features-diagram-software>

- **JDOM¹⁰**

Biblioteca que fornece uma solução completa baseada em Java para acessar, manipular e exibir dados do XML, contribuiu com a extração dos dados do XML.

- **PLUGIN IREPORT DESIGNER¹¹**

O ireport é um *plugin* para gerar relatório visual para JasperReports. Podendo gerenciar gráficos, imagens, sub-relatórios e *crosstabs*. Suporta os formatos PDF, RTF, XML, XLS, CSV, HTM.

¹⁰ <http://www.jdom.org/>

¹¹ <http://plugins.netbeans.org/plugin/4425/ireport>

4 DESENVOLVIMENTO

Nesta seção, é apresentada a metodologia de gestão aplicada no desenvolvimento do trabalho, em seguida, a arquitetura de software na qual baseou-se para executar seu desenvolvimento e quais funcionalidades o sistema abrange, referenciando perfis de usuário, telas do sistema, entre outros.

4.1 METODOLOGIA

O processo de desenvolvimento adotado foi uma personalização do Scrum. O Scrum “é uma metodologia ágil que apresenta qualidade e produtividade para a gestão e o planejamento de projetos de software” (MUNDRA; MISRA; DHAWALE, 2013). Alguns motivos que contribuíram para a escolha desse processo é que, oferece dinamicidade, bastante objetivo e de fácil adaptação. O método ágil encontrado no Scrum é uma forma de desenvolvimento adaptativa e flexível, que permite o desenvolvimento de um projeto em diferentes ciclos ou iterações denominadas Sprints (CARVALHO; MELLO, 2009). Segundo Shwaber e Beedle (2002) a finalidade de uma Sprint é gerar resultados que irão agregar valor ao produto final desenvolvido.

O ciclo de vida adotado no Scrum é tipicamente iterativo. Inicialmente o *Product Owner (Dono do Produto)* seleciona e prioriza o trabalho a ser desenvolvido em uma Sprint. Como resultado as tarefas são detalhadas em uma lista e assume-se um ciclo de tempo para que sejam desenvolvidas. É altamente recomendado que a cada dia de trabalho ocorra uma pequena reunião a fim de controlar o andamento de seu desenvolvimento. Ao final de uma Sprint, que pode variar de 2 a 4 semanas, é apresentado o produto final, o qual é revisado e entregue, a Figura 6 exemplifica todo o ciclo do SCRUM. Neste projeto, foi adotada uma personalização dessa metodologia, ao invés de reuniões diárias, estas ocorreram a cada 15 dias, em cada reunião foi realizada a apresentação e correção, para o orientador, sobre o que foi desenvolvido.

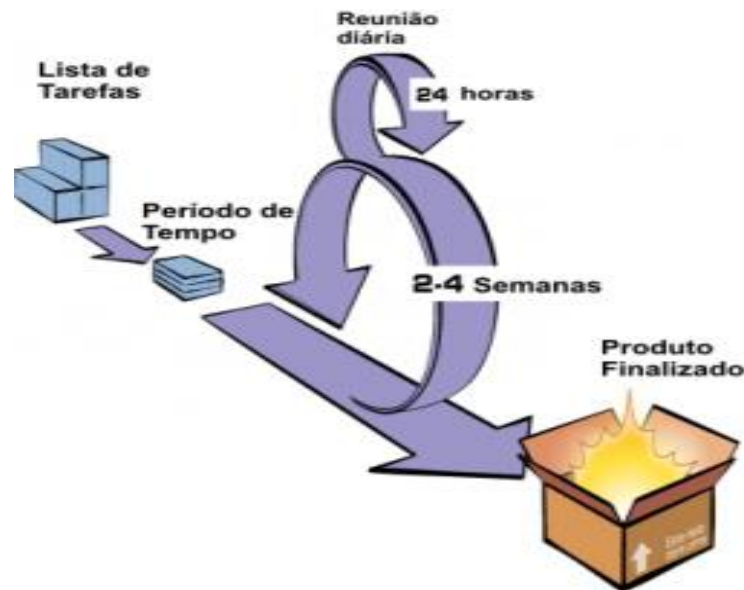


Figura 6 – Ciclo da Metodologia Scrum
 Fonte: <http://www.desenvolvimentoagil.com.br/scrum/>

4.2 ARQUITETURA DO SISTEMA

A arquitetura de software é uma descrição abstrata, na forma de modelos, que mostra diferentes visões do sistema em termos de partes que interagem entre si. Sendo que ela não é uma fase do desenvolvimento, mas o resultado das decisões de *design* sobre a estrutura e o comportamento do software.

O sistema utilizará a arquitetura *Model-View-Controller* (MVC). Essa arquitetura possui como objetivo fundamental, facilitar a manutenção devido a sua divisão em três partes e possibilitar que o código possa ter mais reuso, ou seja, possa ser reaproveitado sem replicações que não sejam necessárias. (OLIVEIRA JUNIOR; FORTES, 2007).

O objetivo deste padrão de desenvolvimento é uma separação total das três camadas da aplicação, que estão representadas na Figura 7 e explicadas a seguir:

Visão – representa o que será mostrado para o usuário ao término do processamento feito pelo controle e, se for o caso, pela camada de modelo.

- Inclui os elementos de exibição no cliente;
- É a camada de interface com o usuário;

- É usada para receber a entrada de dados e apresentar o resultado.

Modelo – Esta camada é o coração da aplicação, pois contém toda a parte de negócio, programação pesada, como acesso ao banco de dados, validação, fórmulas etc.

- Modela os dados e o comportamento por trás do processo de negócio;
- Preocupa-se apenas com o armazenamento, manipulação e geração de dados;
- É um encapsulamento de dados e de comportamento independente da apresentação.

Controle - Determina o fluxo da apresentação, servindo como uma camada intermediária entre a camada de apresentação e a de lógica.

- Controla e mapeia as ações

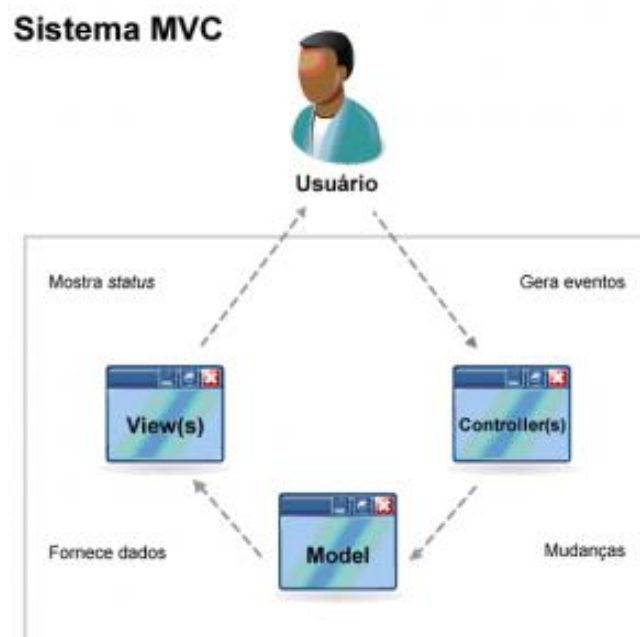


Figura 7 – Padrão MVC

Fonte: <http://blog.locaweb.com.br/tecnologia/aspnet-mvc-o-que-e-isso/>

A Figura 8 representa a estrutura do padrão MVC, a qual foi utilizada no desenvolvimento do sistema, seguida de uma breve explicação.

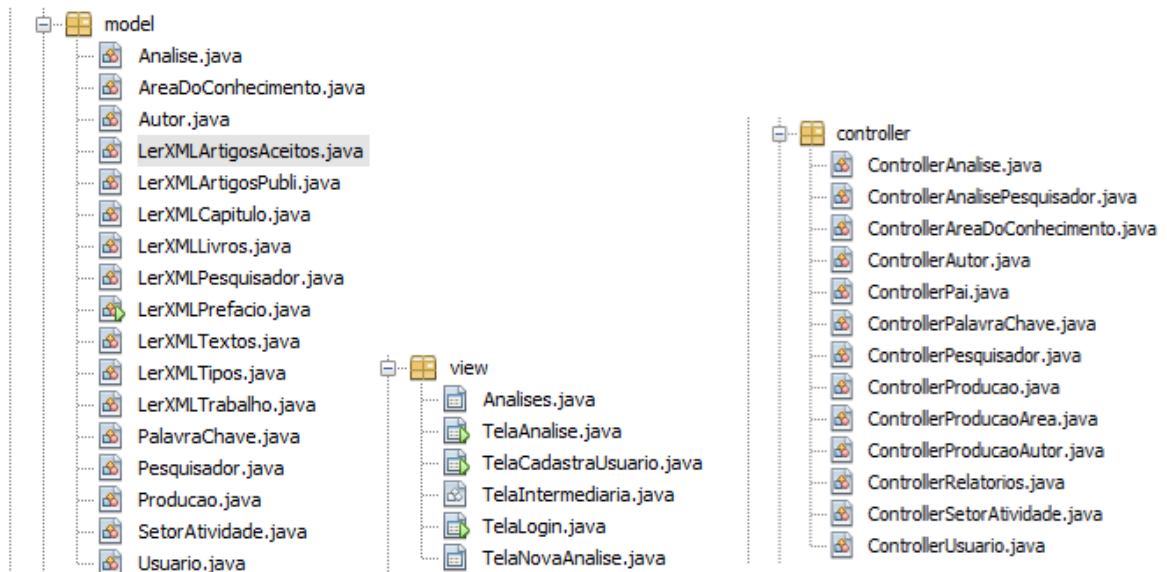


Figura 8 – Aplicação do Padrão MVC

O pacote *model*, contém as classes que dizem respeito à: escrita, validação e leitura dos dados que estão presentes no sistema, o pacote *view* é composto das classes que fazem a interação com o usuário, ou seja, aquelas que permitem digitar, escolher arquivos e acessar relatórios. O *controller*, faz a intermediação dos demais pacotes, ele fornece métodos, combinações para se exibir dados, os quais são retornados pelo *model*.

4.3 FUNCIONALIDADES

Na Tabela 1 são apresentadas as funcionalidades do sistema:

ESSENCIAIS	DESEJÁVEIS
Extração das informações do XML	Remoção de produções redundantes
Inserção das informações do XML no BD	Visualização de redes de coautoria
Visualização de um conjunto de produções	
Gerar relatórios Produções Bibliográficas X Dados	

Tabela 1 – Funcionalidades do Sistema

Extração das informações do XML – Realiza a extração das informações de cada produção bibliográfica pertencentes aos arquivos XML, através de um conjunto de classes que realiza a leitura.

Inserção das informações do XML no BD – Faz a inserção de cada atributo da produção bibliográfica no banco de dados relacional.

Visualização de um conjunto de produções – Visualizar do nome das extrações realizadas e suas datas.

Gerar relatórios Produções Bibliográficas X Dados – Acesso a relatórios com as informações de todas as produções bibliográficas inseridas no sistemas.

4.4 EXTRAÇÃO E ARMAZENAMENTO DA PRODUÇÃO BIBLIOGRÁFICA

A abordagem deste trabalho é em relação a uma das *tags* que existem no XML dos docentes, conhecida como Produção Bibliográfica, segue uma explicação básica de como ela está dividida:

A <PRODUCAO-BIBLIOGRAFICA> agrupa informações a respeito das publicações realizadas pelos docentes. Encontra-se dividida nos tipos abaixo:

- TRABALHOS-EM-EVENTOS
- ARTIGOS-PUBLICADOS
- LIVROS-E-CAPITULOS
- TEXTOS-EM-JORNAIS-OU-REVISTAS
- DEMAIS-TIPOS-DE-PRODUCAO-BIBLIOGRAFICA
- ARTIGOS-ACEITOS-PARA-PUBLICACAO

O objetivo do sistema é ler cada uma dessas *tags* que estiverem presentes no arquivo XML, o qual o usuário pretende realizar uma análise. A seguir estão os conceitos relacionados a cada tipo de produção bibliográfica.

- Trabalhos em eventos: textos publicados em anais de eventos, vinculados a um evento específico.
- Artigos publicados: área destinada para inclusão de artigos científicos já publicados em revistas indexadas com ISSN.
- Livros e capítulos: área para inclusão de informações a respeito de livros ou capítulos de livros produzidos pelo usuário.
- Texto em jornal ou revista: qualquer publicação escrita que tenha sido publicada em meio jornalístico, como roteiros, ensaios, matérias, reportagens, relatos, depoimentos, entrevistas, resumos, resenhas, crônicas, contos, poemas e afins.
- Demais Tipos de produção bibliográfica: qualquer outra produção bibliográfica que não se enquadre nas opções anteriores, inclusive artigos publicados em periódicos sem ISSN.
- Artigos aceitos para publicação: artigos no prelo (*in-press*) que ainda não foram publicados. Os artigos neste item podem ser transferidos quando publicados. Devem ser indexadas com ISSN.

Os tipos de produção citados acima são compostos por *tags* internas, onde estão presentes os atributos que guardam as informações em relação as produções. E seguem um padrão de acordo com o tipo que estão relacionadas, ou seja: trabalho em eventos, artigos publicados, entre outros.

A Figura 9 representa um trecho de XML, mostrando o tipo *trabalho em eventos* com seus respectivos atributos e elementos, os quais pertencem a *produção bibliográfica*. Para realizar a leitura de cada produção, foi preciso implementar um conjunto de classes para realizar a extração das informações. Neste processo, foi utilizada a biblioteca JDOM, que permite acessar, manipular e exibir dados XML a partir do código JAVA. A biblioteca disponibiliza classes e métodos que ajudam na manipulação de informações no formato XML.


```

▼<PRODUCAO-BIBLIOGRAFICA>
▼<TRABALHOS-EM-EVENTOS>
▼<TRABALHO-EM-EVENTOS SEQUENCIA-PRODUCAO="6">
  <DADOS-BASICOS-DO-TRABALHO NATUREZA="COMPLETO" TITULO-DO-TRABALHO="Do DFD ao Use Case" ANO-DO-TRABALHO="2004" PAIS-DO-EVENTO="Brasil"
  IDIOMA="Português" MEIO-DE-DIVULGACAO="MEIO_DIGITAL" HOME-PAGE-DO-TRABALHO="" FLAG-RELEVANCIA="NAO" DOI="" TITULO-DO-TRABALHO-INGLES=""
  FLAG-DIVULGACAO-CIENTIFICA="NAO"/>
  <DETALHAMENTO-DO-TRABALHO CLASSIFICACAO-DO-EVENTO="REGIONAL" NOME-DO-EVENTO="I WorkComp Sul" CIDADE-DO-EVENTO="Florianópolis" ANO-DE-
  REALIZACAO="2004" TITULO-DOS-ANAIS-OU-PROCEEDINGS="NORRCOMP-SUL 2004: Technical Program" VOLUME="" FASCICULO="" SERIE="" PAGINA-INICIAL=""
  PAGINA-FINAL="" ISBN="" NOME-DA-EDITORIA="" CIDADE-DA-EDITORIA=""/>
  <AUTORES NOME-COMPLETO-DO-AUTOR="Renato Balancieri" NOME-PARA-CITACAO="BALANCIERI, R." ORDEM-DE-AUTORIA="2" NRO-ID-CNPQ="7536192699862782"/>
  <AUTORES NOME-COMPLETO-DO-AUTOR="Sandra Ferrari" NOME-PARA-CITACAO="FERRARI, S." ORDEM-DE-AUTORIA="3" NRO-ID-CNPQ="6899334557031469"/>
  <AUTORES NOME-COMPLETO-DO-AUTOR="Alessandro Botelho Bovo" NOME-PARA-CITACAO="BOVO, A. B." ORDEM-DE-AUTORIA="1"/>
  <PALAVRAS-CHAVE PALAVRA-CHAVE-1="Diagrama de Fluxo de Dados" PALAVRA-CHAVE-2="Diagrama de Use Case" PALAVRA-CHAVE-3="Engenharia de Software"
  PALAVRA-CHAVE-4="" PALAVRA-CHAVE-5="" PALAVRA-CHAVE-6=""/>
  ▼<AREAS-DO-CONHECIMENTO>
    <AREA-DO-CONHECIMENTO-1 NOME-GRANDE-AREA-DO-CONHECIMENTO="CIENCIAS_EXATAS_E_DA_TERRA" NOME-DA-AREA-DO-CONHECIMENTO="Ciência da Computação"
    NOME-DA-SUB-AREA-DO-CONHECIMENTO="Metodologia e Técnicas da Computação" NOME-DA-ESPECIALIDADE="Engenharia de Software"/>
  </AREAS-DO-CONHECIMENTO>
  <INFORMACOES-ADICIONAIS DESCRICAO-INFORMACOES-ADICIONAIS="" DESCRICAO-INFORMACOES-ADICIONAIS-INGLES=""/>
</TRABALHO-EM-EVENTOS>

```

Figura 9 - Trecho XML de um Currículo Lattes

As informações sobre as produções bibliográficas, depois que são extraídas do XML do CL, são armazenadas em um banco de dados relacional. Esse banco de dados possui o esquema apresentado na Figura 10. Esse esquema é composto pelas seguintes tabelas: Producao, SetorAtividade, PalavraChave, Autor, Pesquisador, AreaDoConhecimento e Analise. A seguir, apresenta-se uma descrição de cada tabela.

Producao – Tabela usada para armazenar os atributos das produções bibliográficas, como: título, ano, meio de divulgação, entre outros, que pertencem à duas *tags* internas: dados básicos e detalhamento.

SetorAtividade – Armazena atributos que pertencem a uma *tag* chamada setores-atividade, o nome dos atributos seguem o padrão: setor-atividade-1, setor-atividade-2, setor-atividade-3, setor-atividade-4, setor-atividade-5 e setor-atividade-6.

PalavraChave- Tabela utilizada para armazenar atributos que seguem o nome padrão: palavra-chave-1, palavra-chave-2, palavra-chave-3 e estão presentes na *tag* palavras-chave.

Autor – Armazena os atributos em relação a cada autor que está presente nas produções bibliográficas, como: nome completo, nome em citações bibliográficas, ordem de autoria, entre outros que pertencem a *tag* autores.

Pesquisador –Tabela utilizada para armazenar informações em relação ao pesquisador de cada arquivo XML, possui atributos como: nome em citações bibliográficas, nome completo, número identificador, país, sigla país, data atualização e hora atualização.

AreaDoConhecimento – Tabela utilizada para armazenar informações como: nome área do conhecimento, nome grande área do conhecimento, entre outras. Os atributos pertencem a uma *tag* interna chamada área do conhecimento.

Analise – Armazena as informações das extrações realizadas, como nome e data.

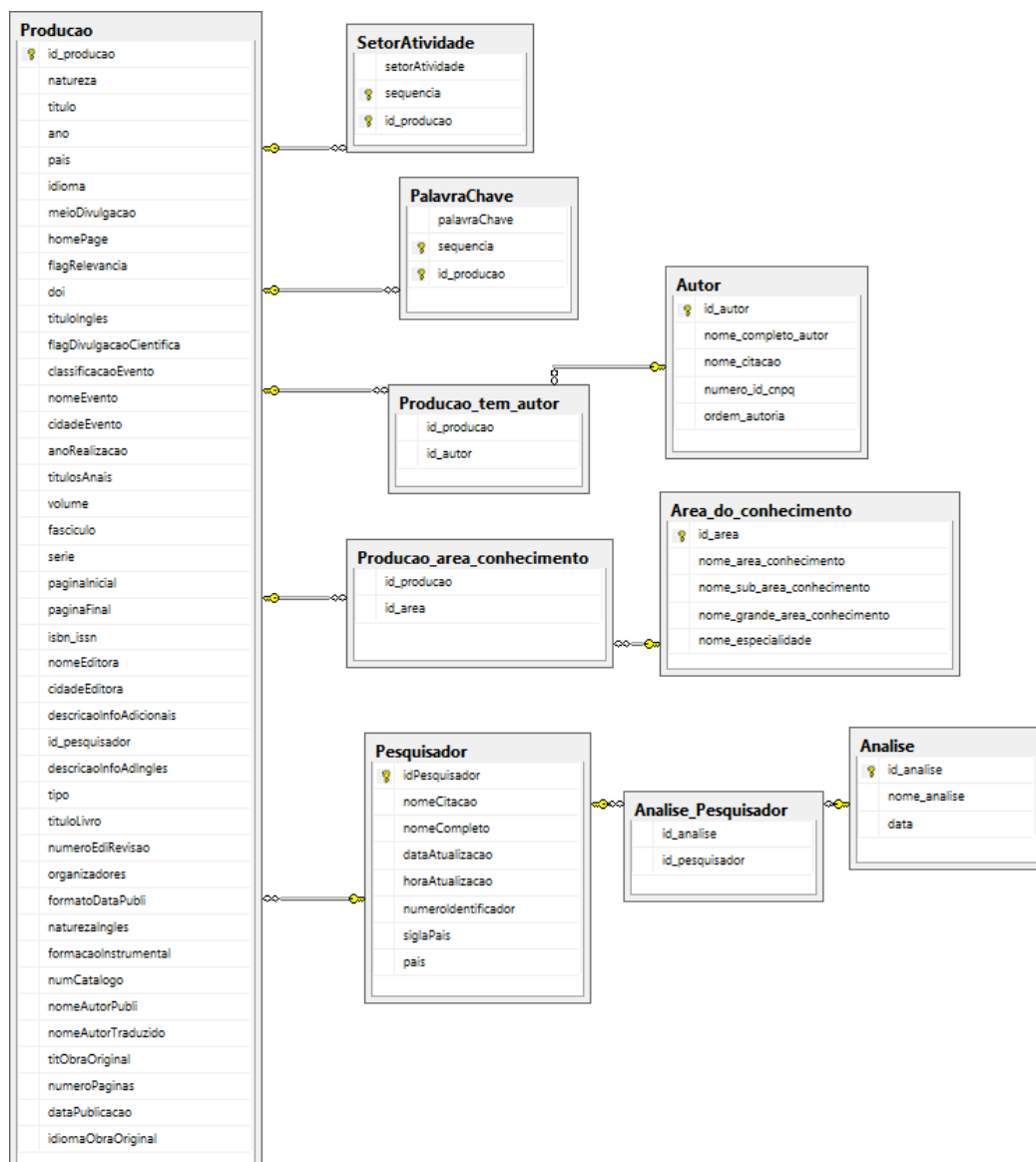


Figura 10 – Diagrama de Banco de Dados

4.5 APRESENTAÇÃO DO SISTEMA

Nesta seção são apresentadas as telas do sistema concluído, a disposição dos componentes e sua interface final.

A tela representada pela Figura 11, é a tela inicial do sistema, onde o usuário tem a opção de realizar uma nova extração de dados ou escolher relatórios para gerar, com informações em relação as produções já inseridas no sistema.

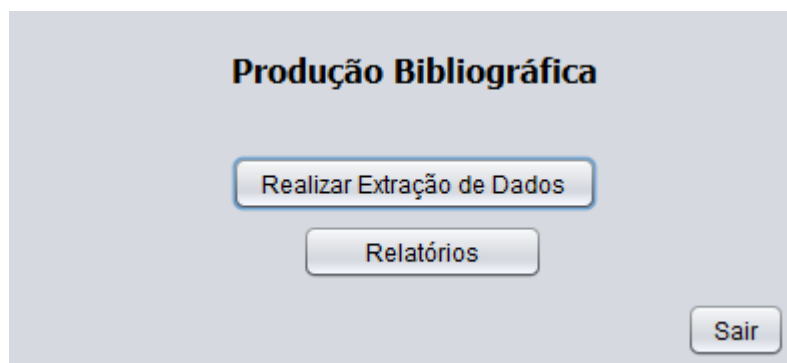


Figura 11 – Tela Inicial do Sistema

Se o usuário escolher a opção Realizar Extração de Dados, abrirá as telas a seguir, representadas pelas figuras 12 e 13:

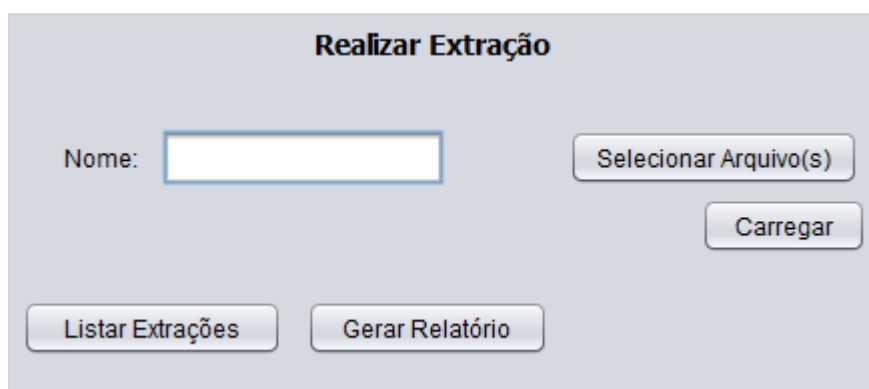


Figura 12 – Tela Realizar Extração

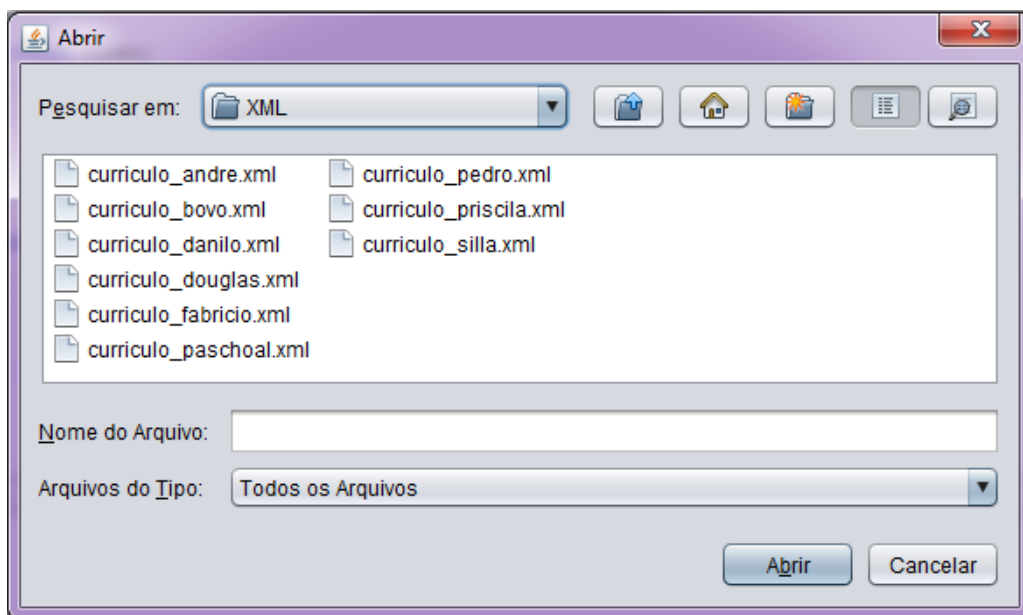
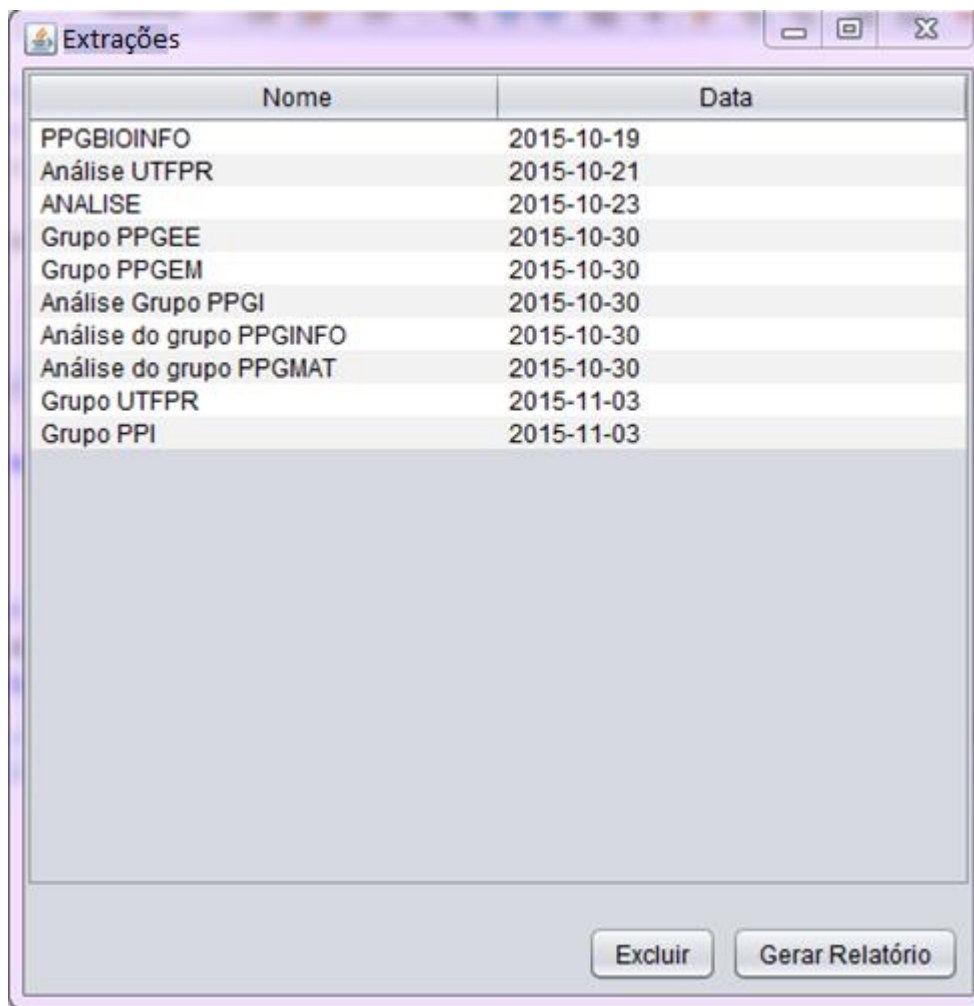


Figura 13 – Selecionar XML's dos Currículos

Na tela 12, a funcionalidade é realizar uma nova extração, onde é colocado um nome para a mesma, por exemplo, Grupo PPGBIOINFO e possui a opção de selecionar o(s) arquivo(s) XML dos pesquisadores que fazem parte do programa no qual deseja-se fazer uma extração das produções bibliográficas. A seleção dos arquivos está representada na figura 13, após isso, o usuário clica no botão carregar, que faz com que os dados das produções bibliográficas sejam inseridos no banco de dados. Outra funcionalidade representada na Figura 12, é o botão para listar extrações já inseridas no sistema, que está representada na Figura 14.



Nome	Data
PPGBIOINFO	2015-10-19
Análise UTFPR	2015-10-21
ANALISE	2015-10-23
Grupo PPGE	2015-10-30
Grupo PPGE	2015-10-30
Análise Grupo PPGI	2015-10-30
Análise do grupo PPGINFO	2015-10-30
Análise do grupo PPGMAT	2015-10-30
Grupo UTFPR	2015-11-03
Grupo PPI	2015-11-03

Figura 14 – Extrações realizadas

A figura 14, mostra todas as extrações inseridas no sistema, com seu nome e data respectivos, o usuário tem a opção de excluir uma extração ou gerar um relatório em relação a mesma, basta selecionar a extração e pressionar o botão gerar relatório, assim trará apenas o nome dos pesquisadores responsáveis pelos XML-s que foram selecionados no momento da nova extração de dados. Exemplo:

Pesquisadores da Extração**Grupo PPGBIOINFO**

06/11/2015 01.23

Andre Yoshiaki Kashiwabara

Alessandro Botelho Bovo

Danilo Sipoli Sanches

Douglas Silva Domingues

Fabricio Martins Lopes

Alexandre Rossi Paschoal

Pedro Henrique Bugatti

Priscila Tiemi Maeda Saito

Carlos Nascimento Silla Junior

Figura 15 – Relatório Extração Grupo PPGBIOINFO

Outra funcionalidade existente na tela inicial (Figura 11) do sistema são os relatórios, que está representada pela figura a seguir:



Figura 16 – Tela Relatórios

Na figura 16, estão presentes as opções de relatórios. A opção Produções Bibliográficas traz um relatório com todas as informações, como: nome do pesquisador, título da produção, tipo, ano, natureza, país, idioma, entre outros, de cada produção inserida durante as extrações. Segue a figura do relatório citado acima:

04/11/2015 00.15

Pesquisador: Andre Yoshiaki Kashiwabara

Extração	Data
Analise	2015-11-04
TIPO PRODUÇÃO:	Artigo Publicado
Título: Characterization of SCAR markers of spp. of domestic fowl and construction of a public relational database (The SCARdb)	
Natureza: COMPLETO	Ano: 2004
Meio Divulgação: MEIO_DIGITAL	Pais:
Flag Relevância: SIM	Idioma: Inglês
Home Page: [http://cluster.fmvz.usp.br/eimeriaScardb/]	DOI: 10.1016/j.femsle.2004.07.034
Título Inglês: Characterization of SCAR markers of spp. of domestic fowl and construction of a public relational database (The SCARdb)	Flag Divulgação Científica: NAO
Classificação Evento:	Nome Evento:
Cidade Evento:	
Títulos Anais: FEMS Microbiology Letters	
Ano Realização:	Volume: 238
Página Inicial: 183	Fasciculo: 1
	Série:
Nome Editora:	Página Final: 188
Cidade Editora:	ISBN/ISSN 03781097

AUTORES

FERNANDEZ, S ,FERNANDEZ, S ,KATSUYAMA, A ,KATSUYAMA, A ,Andre Yoshiaki Kashiwabara ,KASHIWABARA, A ,
MADEIRA, A ,MADEIRA, A ,DURHAM, A ,DURHAM, A ,GRUBER, A ,GRUBER, A ,

ÁREAS DO CONHECIMENTO

Bioquímica , CIENCIAS_BIOLÓGICAS ,Biologia Molecular ,

PALAVRAS-CHAVE

Computação Paralela ,Performance Attributes ,Coccidiosis ,Diagrama de Fluxo de Dados ,Mineração de Texto ,Text Mining ,Text Mining ,Colaboração científica ,Protein-protein interactions ,Descoberta de Conhecimento em Textos ,Engenharia do Conhecimento ,Processamento de Imagens ,Parallel implementation of algorithms ,Domestic fowl ,Diagrama de Use Case , Reconhecimento de Entidades ,Protein-protein interactions ,Protein-protein interactions ,Análise de redes sociais ,Text Mining ,Temporal Knowledge Discovery in Texts ,Nuvens de Termos ,Imagens Médicas ,Eimeria , Engenharia de Software ,Gestão do Conhecimento ,Concept Profile ,Concept

SETORES ATIVIDADE

Page 43 of 96

Figura 17 – Relatório Produção Bibliográfica

Outra opção de relatório é o Quantidade Tipo x Pesquisador, que traz a quantidade por tipo de produção de cada pesquisador, a figura a seguir representa um exemplo do relatório citado acima:

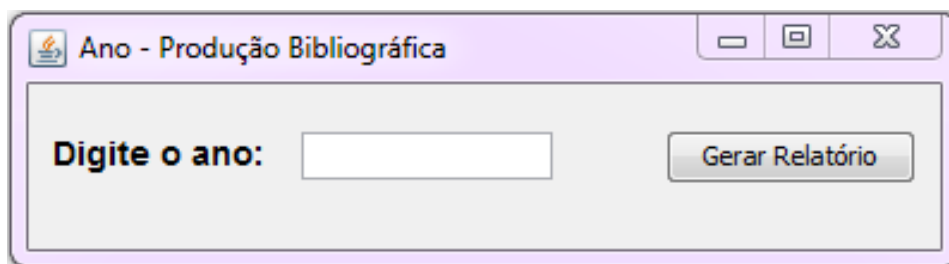
Pesquisador x Tipo				
Pesquisador	Citação	Número Identificador	Tipo Produção	Quantidade por Tipo
Daniilo Sipoli Sanches	SANCHES, D. S.;SANCHES, D.S.;SANCHES,	637765727439814	Artigo Publicado	6
Daniilo Sipoli Sanches	SANCHES, D. S.;SANCHES, D.S.;SANCHES,	637765727439814	Trabalho em Eventos	29
Alessandro Botelho Bovo	BOVO, A. B.	509629414107288	Artigo Publicado	2
Alessandro Botelho Bovo	BOVO, A. B.	509629414107288	Capitulo de Livro Publicado	1

28/10/2015 23.48

Page 1 of 6

Figura 18 – Relatório Pesquisador x Tipo

Quando o usuário escolher a opção Produção/Ano, presente na tela de relatórios (Figura 16), abrirá uma tela que permite digitar um ano, para que o relatório traga apenas as produções realizadas no ano escolhido.



A imagem mostra uma janela de aplicativo com o título "Ano - Produção Bibliográfica". No interior da janela, há o texto "Digite o ano:" à esquerda de um campo de entrada de texto. À direita do campo de entrada, há um botão com o texto "Gerar Relatório".

Figura 19 – Tela Ano – Produção Bibliográfica

Abaixo segue um exemplo do relatório citado, o qual é gerado pela tela acima, trazendo todas as produções publicadas no ano escolhido.

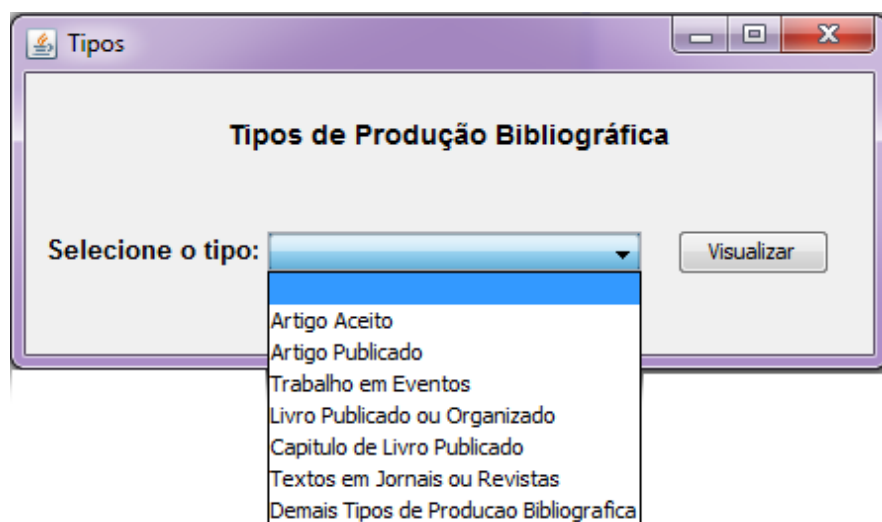


Figura 21 – Tela Tipos de Produção Bibliográfica

Ao escolher algum tipo de produção, o usuário irá clicar em visualizar para trazer o relatório com as produções do tipo escolhido, com o nome do pesquisador e suas respectivas produções.

TIPO PRODUÇÃO BIBLIOGRÁFICA**Artigo Publicado**

05/11/2015 17.35

Pesquisador: Priscila Tiemi Maeda Saito**Título:** Otimização de Algoritmos de Processamento de Imagens Médicas Utilizando a Computação Paralela**Pesquisador:** Priscila Tiemi Maeda Saito**Título:** An Analysis of Parallel Approaches for a Mobile Robotic Self-localization Algorithm.**Pesquisador:** Priscila Tiemi Maeda Saito**Título:** Modelo de Paralelismo para Processamento de Imagens Médicas**Pesquisador:** Priscila Tiemi Maeda Saito**Título:** An Active Learning Paradigm Based on a Priori Data Reduction and Organization**Pesquisador:** Priscila Tiemi Maeda Saito**Título:** Robust active learning for the diagnosis of parasites**Pesquisador:** Priscila Tiemi Maeda Saito**Título:** Choosing the Most Effective Pattern Classification Model under Learning-Time Constraint**Pesquisador:** Andre Yoshiaki Kashiwabara**Título:** Characterization of SCAR markers of spp. of domestic fowl and construction of a public relational database (The SCARdb)

Page 1 of 10

Figura 22 – Relatório Tipos de Produção Bibliográfica

5 CONSIDERAÇÕES FINAIS

Na busca pela aprendizagem de novos conceitos, experiências e claro, visando uma solução para uma realidade corriqueira da nossa Universidade, entre outras, foi possível enxergar claramente que um sistema como esse, ajudaria na execução de algumas atividades relacionadas a programas de pós-graduação. Porém foi observado durante o desenvolvimento, um certo grau de complexidade no tratamento do XML com a linguagem JAVA, justamente pela pouca experiência com programação, porém há inúmeros recursos disponíveis para a implementação do mesmo, seja por pesquisas, trabalhos relacionados a este tema, bibliotecas que se tornam essenciais no seu desenvolvimento.

Fica clara a total importância que a Plataforma Lattes tem nas áreas de pesquisa e educação do nosso País, ela realmente pode ser chamada ou vista como um grande banco de dados, onde inúmeros usuários podem ter acesso, sem restrição, querendo apenas deixar público, as informações relevantes em relação aos pesquisadores, docentes do mundo inteiro, o que a torna cada vez mais procurada pelas partes interessadas. Possui acordos institucionais com Universidades que são definidos como uma parceria para ampliar cada vez mais os benefícios oferecidos aos usuários da web.

A Plataforma é um exemplo, que mostra para os grupos de pesquisa de outros países, que é possível unir inúmeros dados importantes em relação aos seus pesquisadores, os quais trazem novidades, tecnologia, esperança a tudo que está relacionado com a sociedade e através do conhecimento, que está descrito em algumas linhas nos Currículos Lattes, podendo ser passado adiante, ocasionando melhorias em qualquer área de pesquisa e inovação no País.

Cada vez mais cresce o número de pessoas interessadas em desenvolver algum sistema voltado para esta área, ou que “trabalhe” junto com a própria plataforma, pois abre um leque de oportunidades e melhorias.

As tecnologias utilizadas durante o desenvolvimento do sistema, foram de suma importância, pois proporcionaram um grande aprendizado em relação a linguagem e aos benefícios que a mesma proporciona, tornando possível o desenvolvimento do que se planejou, mesmo sem ter alguma experiência.

Este trabalho é um projeto piloto e ficou aberto para possíveis melhorias e novas implementações em relação ao conteúdo do XML, como por exemplo, fazer a leitura de outras produções, como: técnica e artística, gerar indicadores relevantes e mostrar redes de coautoria.

Com o desenvolvimento deste trabalho, foi possível enxergar outras oportunidades de melhorias e aperfeiçoamentos para trabalhos futuros, como: retirada de produções repetidas, gerar redes de co-autorias e aperfeiçoamento de relatórios, pois existem inúmeros dados que podem ser trabalhados para ajudar que as informações cheguem de uma maneira mais prática a quem sempre trabalha com Currículo Lattes e seu XML.

REFERÊNCIAS

CARVALHO, Bernardo V.; MELLO, Carlos H. P.; **Revisão, Análise E Classificação Da Literatura Sobre O Método De Desenvolvimento De Produtos Ágil Scrum**. In: SIMPÓSIO DE ADMINISTRAÇÃO DA PRODUÇÃO, LOGÍSTICA E OPERAÇÕES INTERNACIONAIS, 12., São Paulo, **Anais...** 2009. p. 1-16.

CONSCIENTIAS. **Comunidade Conscientias**. Disponível em: < <http://impl.cnpq.br/impl/>>. Acesso em: 21/10/2015.

CNPQ. **Conselho Nacional de Desenvolvimento Científico e Tecnológico**. Disponível em: < <http://www.cnpq.br/>>. Acesso em: 15/08/2015.

CLARO D. B. e SOBRAL J. B. M. **PROGRAMAÇÃO EM JAVA**. Copyleft Pearson Education. Florianópolis, SC., Daniela Barreiro; SOBRAL, João Bosco Manguera. **Programação em JAVA**.

HEITLINGLER, Paulo. **O Guia Prático da XML**. Centro Atlântico, Lda, 2001.

IREPORTDESIGNER. **JasperSoft Community**. Disponível em: < <http://community.jaspersoft.com/project/ireport-designer/>>. Acesso em: 08/09/2015.

MENA-CHALCO, J. P. CESAR-JR, C. **scriptLattes: An open-source knowledge extraction system from the Lattes platform**. Journal of the Brazilian Computer Society, vol. 15, n. 4, páginas 31--39, 2009.

MENA-CHALCO, J. P. CESAR-JR, C. **Prospecção de dados acadêmicos de currículos Lattes através de scriptLattes**. Capítulo do livro Bibliometria e Cientometria: reflexões teóricas e interfaces São Carlos: Pedro & João, páginas 109-128, 2013.

JDOM. **JDOM**. Disponível em: < <http://www.jdom.org/>>. Acesso em: 04/06/2015.

JDOM. **JDOM API Specification**. Disponível em: < <http://www.jdom.org/docs/apidocs/>>. Acesso em: 11/07/2015.

JASPERSOFT. **JasperSoft Community**. Disponível em: < <http://community.jaspersoft.com/>>. Acesso em: 10/09/2015.

MUNDRA, Ashish.; MISRA, Sanjay.; DHAWALE, Chitra A. **Practical Scrum-Scrum Team: Way to Produce Successful and Quality Software**. In: INTERNATIONAL CONFERENCE COMPUTATIONAL SCIENCE AND ITS APPLICATIONS, 13, 2013. Ho Chi Minh City. Anais... Ho Chi Minh: IEEE, 2013. p. 119-123.

MISTRY, Ross; MISNER, Stacia. **Introducing Microsoft SQL Server 2008 R2**. ed. Redmond Washington: Microsoft Express, 2009.

MSDN. **Microsoft SQL Server**. Disponível em: <<https://msdn.microsoft.com/pt-br/library/bb545450.aspx>>. Acesso em: 12/08/2015.

NETBEANS. **NetBeans**. Disponível em: <<https://netbeans.org/>>. Acesso em: 02/06/2015.

NETBEANS. **iReport Plugin**. Disponível em: <<http://plugins.netbeans.org/plugin/4425/ireport>>. Acesso em: 26/08/2015.

OLIVEIRA JUNIOR, Edson A.; FORTES, Renata P. **Arquitetura de Software na Web Atual: Processamento no Servidor**. 2007.

PLATAFORMA LATTES. **Plataforma Lattes**. Disponível em: <<http://lattes.cnpq.br/>>. Acesso em: 07/06/2015.

Programa de Pós-Graduação em Bioinformática – PPGBIOINFO. **Corpo Docente**. Disponível em: <<http://www.utfpr.edu.br/cornelioprocopio/cursos/mestrados-doutorados/Ofertados-neste-Campus/ppgbioinfo/corpo-docente>>. Acesso em: 01/09/2015.

STELA, Grupo. **Histórico SCienTI**. Revista SCienTI, v. 1, n. 1, p. 6-7, dez. 2002

SCHWABER, Ken; BEEDLE, Mike. **Agilè Software Development with Scrum**. New Jersey: Prentice Hall, 2002.