

Extração de Dados Quantitativos do Currículo Lattes

Franklin Matheus Boijink¹, Fernando Sarturi Prass¹

¹Sistemas de Informação – Centro Universitário Franciscano
Rua dos Andradas, 1614 – 97.010.032 – Santa Maria – RS – Brasil

franklinboijink@unifra.edu.br, fernando.prass@unifra.edu.br

Abstract. *The extraction of academic indicators from the Lattes Platform is a process widely used by educational institutions. To streamline this process and reduce errors, it is recommended to automate it. This article presents a project developed in Python with Django framework which shows quantitative data of the curriculum, such as bibliographic productions and teacher training, data related to the professionals being assigned to a particular educational institution.*

Resumo. *A extração de indicadores acadêmicos da Plataforma Lattes é um processo amplamente utilizado pelas instituições de ensino. Para tornar este processo mais ágil, bem como obter uma redução de erros envolvidos, tem-se a necessidade de automatizá-lo. Este artigo apresenta um projeto desenvolvido em Python com framework Django, o qual apresenta dados quantitativos dos currículos, tais como as produções bibliográficas e capacitação docente, dados relacionados aos profissionais sendo eles atribuídos a determinada instituição de ensino.*

1. Introdução

A Plataforma Lattes consiste em uma base de dados de currículos acadêmicos, que armazena informações de grupos de pesquisa e de instituições de ensino em um único Sistema de Informação. A plataforma atua em algumas áreas, dentre elas: ações de planejamento, gestão, operacionalização do fomento do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), como em outras agências de fomento federal e estadual, nas fundações estaduais de apoio a ciência e tecnologia, das instituições de ensino superior e dos institutos de pesquisa [CNPq 2017 a].

Com o passar dos anos, a Plataforma Lattes tem sofrido alterações significativas, relacionadas principalmente a desempenho e usabilidade do sistema, tornando-a, desta forma, mais prática e simples de ser utilizada.

No último censo, realizado em 2016, houve um crescimento de aproximadamente 149% no número de grupos de pesquisa cadastrados, 251% no número de pesquisadores e de 278% de doutores, em relação aos números registrados no censo de 2002 [CNPq 2017 b].

O processo de extração de indicadores institucionais, quando realizado de forma manual, possui limitações relacionadas aos diversos passos que devem ser seguidos para a obtenção das informações desejadas. Para ter acesso a qualquer dado da Plataforma, é necessária a inserção do código de identificação ou do nome da pessoa, a seleção dos filtros desejados e a realização da inserção de um código de validação (*captcha*), para somente assim ter acesso as informações solicitadas.

O projeto aqui proposto busca realizar a extração desses dados de forma automatizada, tornando-a assim mais ágil e precisa, tendo em vista que o processo manual está sujeito a falhas. A realização de um *software* para a extração desses dados é proposta de duas maneiras: com acesso direto a base de dados da Plataforma Lattes, ou ao próprio arquivo Extensible Markup Language (XML) do currículo, realizando o processamento das informações através de um Parser C#, armazenando-as em uma base de dados MySQL, e posteriormente, apresentando essas informações em uma *interface*.

1.1. Objetivos Gerais

Este trabalho tem por objetivo o desenvolvimento de uma ferramenta para a extração e geração de dados quantitativos da Plataforma Lattes.

1.2. Objetivo Específico

Para atingir o objetivo proposto, os seguintes objetivos específicos devem ser seguidos:

- Elaborar uma proposta de *software*;
- Definir escopo e metodologia de desenvolvimento;
- Desenvolver uma *interface* para acessar os dados armazenados na base de dados MySQL e para a visualização desses dados de uma forma detalhada.

2. Referencial Teórico

Esta seção apresenta a revisão bibliográfica do trabalho, tendo como foco as tecnologias utilizadas para o desenvolvimento do *software*.

2.1. Plataforma Lattes

Desde o início dos anos 80, já havia interesse por parte do CNPq na criação de um formulário padrão para registro de currículos acadêmicos. No final dos anos 80, foi disponibilizada uma ferramenta utilizada em universidades e instituições de ensino que se comunicavam pela rede BITNET, com cerca de 30.000 currículos. No início dos anos 90, o CNPq desenvolveu um formulário eletrônico para realizar a captação de dados curriculares, onde os pesquisadores preenchiam o formulário e realizavam o envio de um disquete ao CNPq, que posteriormente, realizava o cadastro das informações. No final do ano de 1999, foi lançado e padronizado o Currículo Lattes, como sendo o formulário de currículos a ser utilizado no âmbito do Ministério da Ciência e Tecnologia e CNPq [CNPq 2017 a].

2.2. Extração de Dados

A Plataforma Lattes disponibiliza a extração de dados de sua base de duas maneiras, sendo elas:

- Extração de dados de forma manual, em que o usuário realiza todo o processo de inserção de dados nos campos desejados, juntamente com um código de acesso. A ferramenta realiza o processamento destes dados e os apresenta em sua *interface*. Ela permite a extração de currículos nos formatos: Document Type Definition (DTD), XML Schema Definition (XSD) (beta), Portable Document Format (PDF), PDF XSD (beta), e para os grupos de pesquisa nos formatos: XSD ou PDF.

- Extração de dados de forma automatizada, sendo necessária a realização de uma solicitação por parte da instituição de ensino, para, posteriormente receber a liberação de acesso a base via Web Service, que permite a extração destes dados no formato XML. [CNPq 2017 a].

2.3. Estruturação da base acadêmica

A organização do arquivo apresenta, inicialmente, a identificação e dados gerais, onde constam informações referentes a dados pessoais e formação acadêmica, a versão 1.4.1 do Currículo é a que está sendo utilizada neste trabalho.

Para a geração de indicadores institucionais, é fundamental a obtenção de informações referentes à formação acadêmica e titulação. Nela constam informações relacionadas aos níveis de formação, onde são apresentadas informações referentes à instituição de ensino, informações sobre o curso realizado, dentre outras. A Figura 1, apresenta o modelo de formação e titulação.

```
<TITULACAO-FORMACAO>
<GRADUACAO SEQUENCIA-FORMACAO="1" NIVEL="1" TITULO-DO-TRABALHO-DE-CONCLUSAO-DE-CURSO="" NOME-DO-ORIENTADOR="" CODIG
<MESTRADO SEQUENCIA-FORMACAO="11" NIVEL="3" CODIGO-INSTITUICAO="000000000000" NOME-INSTITUICAO="Universidade Federa
<PALAVRAS-CHAVE PALAVRA-CHAVE-1="Análise de Agrupamentos (Cluster Analysis)" PALAVRA-CHAVE-2="Mineracao de Dados
<AREAS-DO-CONHECIMENTO>
<AREA-DO-CONHECIMENTO-1 NOME-GRANDE-AREA-DO-CONHECIMENTO="CIENCIAS_EXATAS_E_DA_TERRA" NOME-DA-AREA-DO-CONHECII
<AREA-DO-CONHECIMENTO-2 NOME-GRANDE-AREA-DO-CONHECIMENTO="CIENCIAS_EXATAS_E_DA_TERRA" NOME-DA-AREA-DO-CONHECII
</AREAS-DO-CONHECIMENTO>
<SETORES-DE-ATIVIDADE SETOR-DE-ATIVIDADE-1="Atividades de Banco de Dados" SETOR-DE-ATIVIDADE-2="Informacao e Ges
</MESTRADO>
</TITULACAO-FORMACAO>
```

Figura 1. Formação e titulação

Outra informação importante para a geração destes indicadores são as produções bibliográficas, é a seção na qual são descritas as publicações realizadas. Dentre elas destacam-se os trabalhos em eventos, artigos publicados, publicações em livros ou capítulos. Nelas estão descritas as informações referentes ao trabalho efetuado, os autores envolvidos na realização do mesmo, o local onde foram realizadas, dentre outras. Sua estrutura é apresentada na Figura 2.

```
<PRODUCAO-BIBLIOGRAFICA>
<TRABALHOS-EM-EVENTOS>
<TRABALHO-EM-EVENTOS SEQUENCIA-PRODUCAO="131">
<DADOS-BASICOS-DO-TRABALHO NATUREZA="RESUMO EXPANDIDO" TITULO-DO-TRABALHO="Implementacao Biometrica e e
<DETALHAMENTO-DO-TRABALHO CLASSIFICACAO-DO-EVENTO="LOCAL" NOME-DO-EVENTO="XIX SIMPOSIO DE ENSINO, PESQU
<AUTORES NOME-COMPLETO-DO-AUTOR="Jose Machado" NOME-PARA-CITACAO="MACHADO, Jose" ORDEM-DE-AUTORIA="1" N
<AUTORES NOME-COMPLETO-DO-AUTOR="Carla Freitas" NOME-PARA-CITACAO="FREITAS, Carla" ORDEM-DE-AUTORIA="2"
<PALAVRAS-CHAVE PALAVRA-CHAVE-1="Arduino" PALAVRA-CHAVE-2="Android" PALAVRA-CHAVE-3="" PALAVRA-CHAVE-4=
</TRABALHO-EM-EVENTOS>
</TRABALHOS-EM-EVENTOS>
<ARTIGOS-PUBLICADOS>
<ARTIGO-PUBLICADO SEQUENCIA-PRODUCAO="123" ORDEM-IMPORTANCIA="">
<DADOS-BASICOS-DO-ARTIGO NATUREZA="COMPLETO" TITULO-DO-ARTIGO="Análise e desenvolvimento de um Ambiente
<DETALHAMENTO-DO-ARTIGO TITULO-DO-PERIODICO-OU-REVISTA="Revista de Informatica Aplicada" ISSN="18095585
<AUTORES NOME-COMPLETO-DO-AUTOR="Carla Freitas" NOME-PARA-CITACAO="FREITAS, Carla" ORDEM-DE-AUTORIA="1"
<PALAVRAS-CHAVE PALAVRA-CHAVE-1="Infomatca na Educacao" PALAVRA-CHAVE-2="AVA - Ambiente Virtual de Apr
<AREAS-DO-CONHECIMENTO>
<AREA-DO-CONHECIMENTO-1 NOME-GRANDE-AREA-DO-CONHECIMENTO="CIENCIAS_EXATAS_E_DA_TERRA" NOME-DA-AREA-D
</AREAS-DO-CONHECIMENTO>
</ARTIGO-PUBLICADO>
</ARTIGOS-PUBLICADOS>
</PRODUCAO-BIBLIOGRAFICA>
```

Figura 2. Produções bibliográficas

2.4. Trabalhos Correlatos

Esta subseção apresenta os trabalhos correlatos, demonstrando suas similaridades, características e diferenças em relação a este trabalho.

2.4.1. Semantic Lattes: uma ferramenta de consulta de informações acadêmicas da base Lattes baseada em ontologias

O trabalho de [Costa; Yamate 2009] apresenta uma ferramenta para extração de dados do currículo Lattes baseadas em ontologias e que também utiliza conceitos e técnicas de Web Semântica para que suas buscas se tornem mais precisas. O modelo é dividido em três camadas:

- Camada de apresentação: responsável pela entrada das perguntas;
- Camada de aplicação: responsável pelo processamento dos dados recebidos;
- Camada de dados: responsável pelo armazenamento dos dados gerados.

A organização do sistema bem como de cada camada é representada na Figura 3.

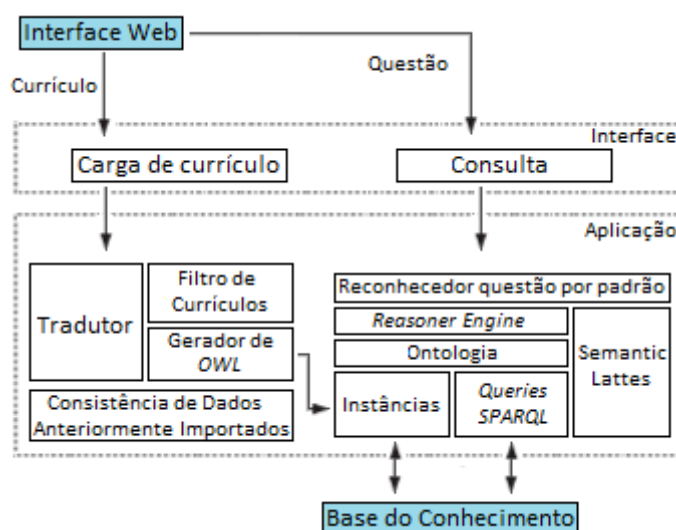


Figura 3. Representação da estrutura do Semantic Lattes. Adaptado de [Costa; Yamate 2009]

O funcionamento do Semantic Lattes ocorre da seguinte maneira: o sistema carrega o currículo no formato XML, o mecanismo tradutor identifica os elementos relevantes e gera um arquivo Web Ontology Language (OWL), que contém apenas as instâncias das classes que foram definidas na ontologia. Estas informações são incorporadas à base de dados na forma de triplas Resource Description Framework (RDF).

Esse projeto se assemelha ao proposto na geração de relatórios relacionados às pesquisas nas informações extraídas. Porém, a diferença que este trabalho apresenta tem relação com a geração de indicadores institucionais a partir dos dados extraídos.

2.4.2. ScriptLattes um sistema *open source* de extração de conhecimento da plataforma Lattes

O trabalho de [Mena-Chalco; Junior 2009] apresenta um sistema de código aberto, projetado com o intuito de fornecer relatórios acadêmicos de grupos a partir da extração de dados públicos disponíveis na plataforma Lattes. O sistema é composto por seis módulos, dentre eles destacam-se:

- A seleção de dados;
- O pré-processamento dos dados;

- O tratamento de redundância;
- A geração e gráficos de colaboração entre os membros do grupo;
- A geração e mapas de pesquisa tendo como base informações geográficas;
- A criação automática de relatórios de produção bibliográfica, técnica e artística e supervisão acadêmica.

A organização estrutural destes módulos, é representada na Figura 4.

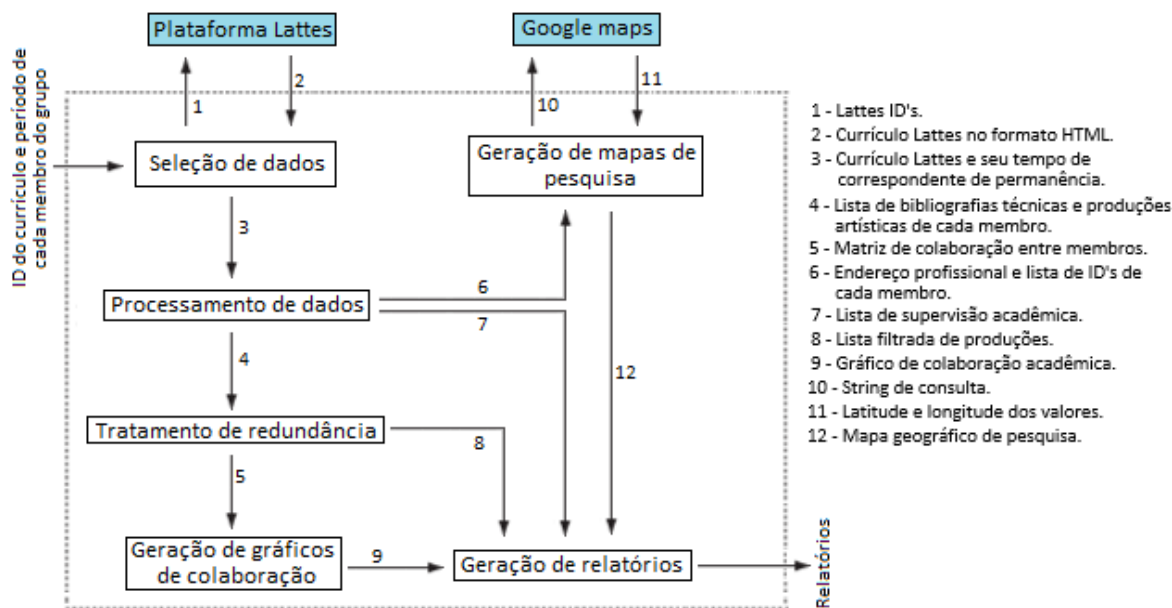


Figura 4. Representação estrutural do ScriptLattes. Adaptado de [Mena-Chalco; Junior 2009]

O projeto foi desenvolvido para fornecer respostas a perguntas, através da criação automática de relatórios, onde são informados os grupos de pesquisa previamente registrados na Plataforma Lattes. Ele realiza o *download* dos currículos através do site do CNPq, extrai as informações necessárias, realiza o controle de redundância nas produções científicas e gera relatórios de produção supervisão acadêmica e gráficos de colaboração bem como o mapa de pesquisa dos membros do grupo [Mena-Chalco; Junior 2009].

Dentre as similaridades encontradas com o trabalho aqui proposto destaca-se a geração de relatórios. A diferença para este trabalho está relacionada à geração de indicadores institucionais tanto individuais quanto por grupos de pesquisa, tendo em vista que o ScriptLattes realiza somente a geração de informações referentes a grupos de pesquisa.

3. Metodologias

Metodologias ágeis de desenvolvimento surgiram com o intuito de ter um maior foco nas pessoas ao invés dos processos. Esta metodologia está relacionada à redução de tempo gasto na realização de documentação do *software*, e com relação ao desenvolvimento de forma mais rápida [Beck 2001].

Foi nesse momento que surgiu o Manifesto Ágil. Dentre os reunidos estavam representantes do eXtreme Programming (XP), Scrum, Crystal, Adaptive Software

Development (ASD), Feature Driven Development (FDD), seu foco principal era encontrar um padrão para o desenvolvimento ágil de *software* [Highsmith 2001]. Técnicas do Scrum foram escolhidas como base para este trabalho, pois trata-se de uma metodologia de gestão de equipes, tendo em vista que o trabalho é parte de um projeto institucional já concluído, no qual são abordados temas relacionados à extração de dados da Plataforma Lattes, a parte específica deste projeto está destacada na Figura 7, no qual a *interface* é responsável pela apresentação de indicadores institucionais.

3.1. Scrum

Scrum é uma metodologia de desenvolvimento ágil de *software* em que os processos são divididos em ciclos geralmente mensais (Sprint), que representam um conjunto de atividades que devem ser realizadas. Cada Sprint dura geralmente entre duas a quatro semanas. Para melhor organização de cada Sprint pode ser utilizado o quadro de trabalho, também conhecido como Kanban, onde as tarefas são divididas em quatro estados, que são: a fazer, em andamento, em testes e concluído.

Dentro desta metodologia existem três papéis principais:

- **Product Owner:** pessoa responsável pelo retorno financeiro, aceitar ou recusar os resultados de cada Sprint, realizar a alteração de prioridades para cada Sprint, basicamente seria o gerente de projeto.
- **Scrum Master:** mantém o time unido, garante que o processo esteja sendo seguido, realiza o planejamento e participa de reuniões diárias para garantir que o processo seja seguido.
- **Scrum Team:** composto de cinco a nove membros seleciona o processo que deseja realizar dentro de cada Sprint [BRQ 2017].

A organização dos processos de realização do Scrum é descrita na Figura 5.

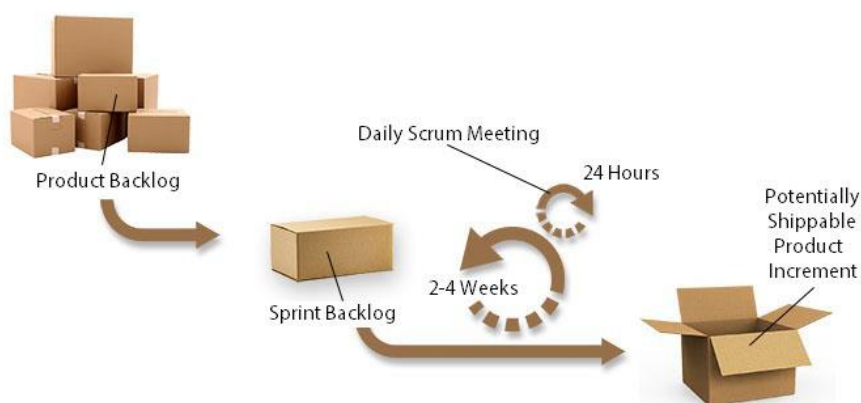


Figura 5. Representação da organização de cada processo. Adaptado de [BRQ 2017]

Técnicas desta metodologia foram escolhidas, pois este trabalho é uma contribuição a outro projeto já concluído, que consiste na extração de dados da Plataforma Lattes. Dentre as técnicas utilizadas para a realização deste trabalho, destacam-se as reuniões realizadas para exposição de dificuldades encontradas e avanços obtidos, a divisão das tarefas em *sprint*, bem como a utilização auxiliar de um quadro de trabalho (Kanban) para visualizar o andamento do projeto como um todo.

A estrutura geral do sistema é representada na Figura 7, onde são descritos os processos necessários para a apresentação dos dados em *interface*. Este trabalho apresenta um complemento para outro projeto institucional, desta forma optou-se por utilizar técnicas desta metodologia para dar seguimento ao projeto.

3.2. Ferramentas

No projeto como um todo foram utilizadas as linguagem de programação C# para a realização do Parser, Python com *framework* Django para o desenvolvimento da *interface*, bem como base de dados MySQL para armazenar as informações. Neste TFG, foi utilizado somente a linguagem de programação Python com *framework* Django, para o desenvolvimento da *interface* de geração de indicadores institucionais, juntamente com a base de dados MySQL, utilizada para armazenar as informações, sendo a mesma base de dados utilizada no projeto.

O Python é uma linguagem de programação amplamente utilizada em diversas áreas, com aplicação em diversos domínios, desde o desenvolvimento Web a aplicações científicas [Borges 2010].

O Django é um *framework* Python, de código aberto, com sua principal característica relacionada ao estímulo de desenvolvimento ágil, possui design limpo e objetivo [Django Software Foundation 2017].

O C# é uma linguagem de programação orientada a objetos, criada para desenvolver uma grande variedade de aplicativos, que são executados no .NET *framework* [Lotar 2010].

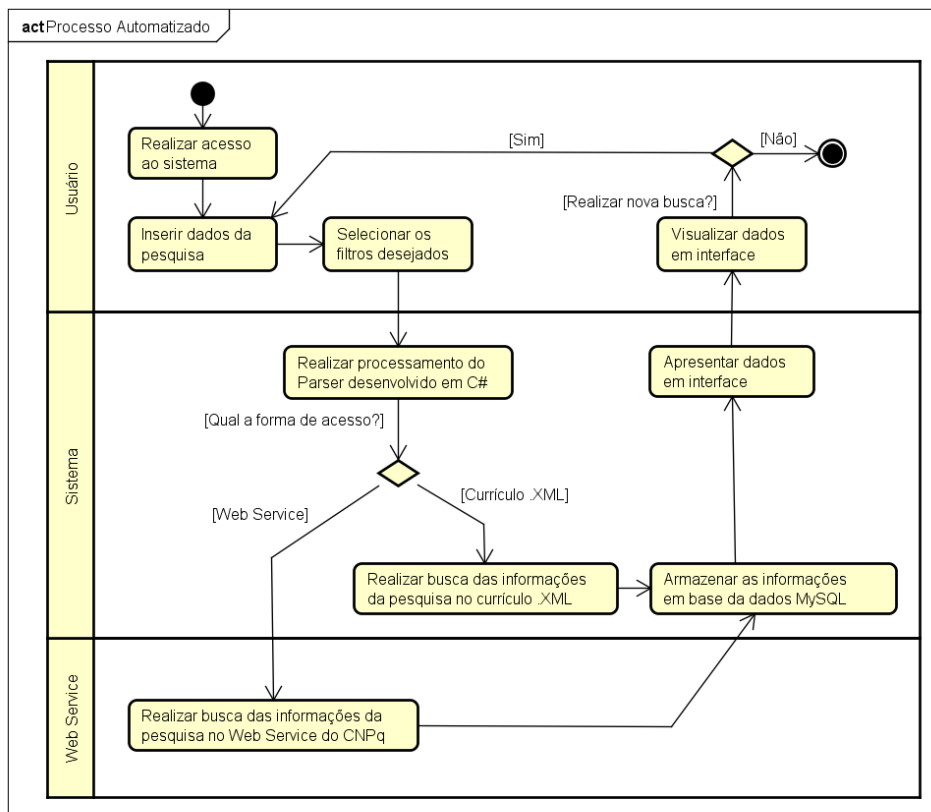
O MySQL é um Sistema Gerenciador de Banco de Dados, de código aberto, que utiliza a linguagem padrão Structered Query Language (SQL) para inserir, alterar ou excluir as informações em seu banco de dados [Sares 2001].

4. Projeto

Esta seção apresenta a proposta do sistema a ser desenvolvido, onde serão apresentadas suas funcionalidades, de que forma são armazenadas as informações na base de dados MySQL local, juntamente com a prototipação de telas e cronograma do projeto como um todo.

4.1. Funcionalidades do sistema

O processo para a obtenção dos dados, apresentado na Figura 6, se dá com o acesso do usuário à *interface* do sistema, o preenchimento dos campos de busca (CPF ou Nome e Data de Nascimento), busca dos dados via Web Service do CNPq, processamento e armazenamento dos dados em uma base de dados MySQL local e apresentação dos relatórios institucionais.



powered by Astah

Figura 6. Diagrama de atividades

O início do processo se dá na realização do Parser (leitura e processamento) dos dados curriculares, sejam eles obtidos por meio do Web Service ou através do arquivo XML do currículo. Após a realização da leitura e do processamento das informações solicitadas, as mesmas são armazenadas em uma base MySQL local.

A *interface* responsável pela apresentação desses dados, realiza o acesso a base de dados, obtém as informações necessárias e as apresenta em tela. A apresentação dessas informações ocorre de duas maneiras, sendo uma delas a exibição de relatórios simplificados e a outra relacionada em específico a este projeto, na qual são apresentados os indicadores institucionais, como destacado na Figura 7, estes indicadores são apresentados em uma *interface* de maneira gráfica ou em formato de relatórios.

Os relatórios simplificados exibem informações de maneira individual, essas informações estão relacionadas apenas ao indivíduo selecionado, não realizando desta forma a análise comparativa entre outros indivíduos do mesmo grupo. Para solucionar este problema, houve a necessidade da geração de indicadores institucionais que são descritos neste trabalho, os quais apresentam dados comparativos entre determinado grupo de indivíduos, selecionados de acordo com a necessidade do usuário, a Figura 7, apresenta uma visão ampla de todo o processo, ressaltando a representação de indicadores institucionais, a qual foi desenvolvida em Python com a utilização do *framework* Django.

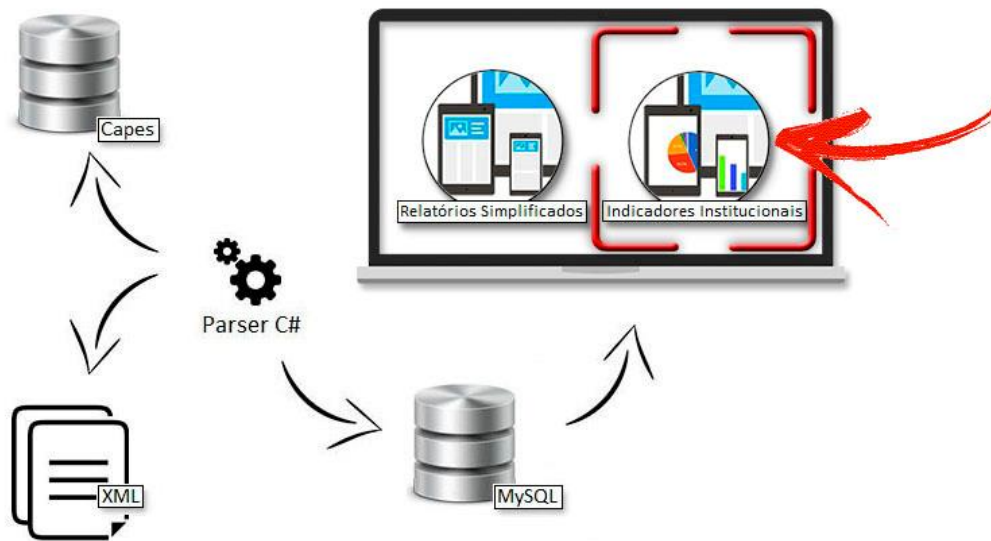


Figura 7. Estrutura do sistema

4.2. Armazenamento dos dados

O processo para armazenamento dos dados na base MySQL, depende de alguns passos a serem seguidos, dentre eles estão a busca das informações solicitadas, que são executadas através do Parser, em que o mesmo realiza o acesso ao XML do currículo, a leitura e a busca das informações (formação e publicações), e armazena as mesmas em suas respectivas tabelas na base de dados.

A Figura 8 apresenta o modelo desenvolvido para o armazenamento das titulações, como: título do trabalho, ano, curso e a instituição de ensino onde foram realizadas. Estes dados são obtidos a partir da estrutura XML apresentada na Figura 1.

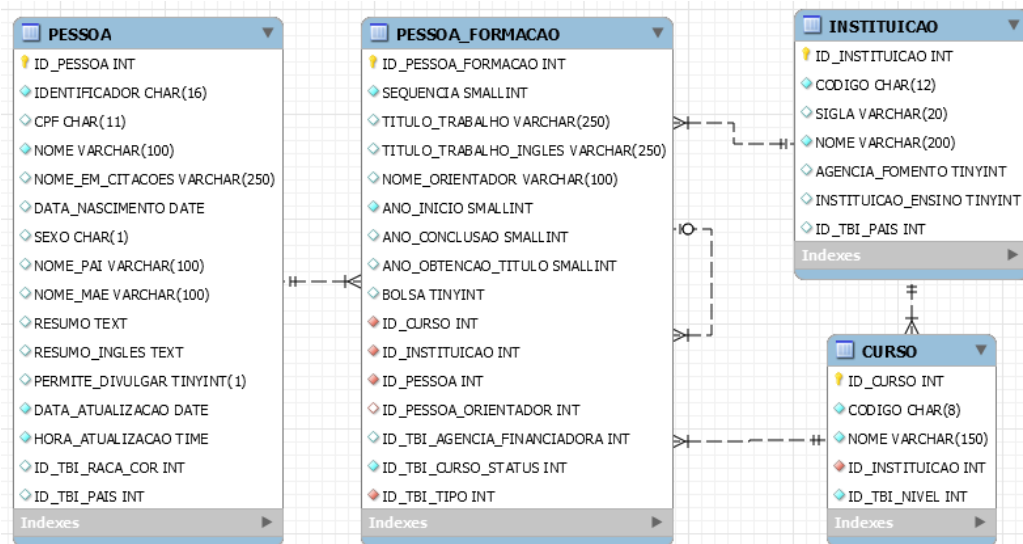


Figura 8. Estrutura da base de dados de titulação

As produções bibliográficas, como apresentadas na Figura 10, são armazenadas em suas respectivas tabelas. Cada dado está organizado de acordo com a identificação da produção realizada, em que cada uma destas publicações possui relação com os dados de

identificação do autor, o nome utilizado em citações e dados referentes à própria publicação, tais como: a sua titulação, o detalhamento da obra, o local onde foi realizada sua publicação, dentre outras. A obtenção destes dados ocorre a partir da estrutura XML, apresentada na Figura 2. Como representação da estrutura dos dados obtidos, processo que antecede seu armazenamento, os dados seguem o formato descrito na Figura 9, onde as informações obtidas tais como: TITULO-DO-TRABALHO, TITULO-DO-TRABALHO-INGLES, FLAG-RELEVANCIA, são armazenadas em seus respectivos campos na tabela PESSOA_PRODUCAO. As informações relacionadas ao autor, dentre elas o seu NOME-COMPLETO-DO-AUTOR e seu NOME-PARA-CITACAO, são armazenadas na tabela PESSOA nos campos a referentes.

```
<TRABALHO-EM-EVENTOS SEQUENCIA-PRODUCAO="">
  <DADOS-BASICOS-DO-TRABALHO NATUREZA="" TITULO-DO-TRABALHO="" ANO-DO-TRABALHO="" PAIS-DO-EVENTO="" IDIOMA="" MEIO-DE-DIVULGACAO="" HOME-PAGE-DO-TRABALHO="" FLAG-RELEVANCIA="" DOI="" TITULO-DO-TRABALHO-INGLES="" FLAG-DIVULGACAO-CIENTIFICA="" />
  <DETALHAMENTO-DO-TRABALHO CLASSIFICACAO-DO-EVENTO="" NOME-DO-EVENTO="" CIDADE-DO-EVENTO="" ANO-DE-REALIZACAO="" TITULO-DOS-ANAIS-OU-PROCEEDINGS="" VOLUME="" FASCICULO="" SERIE="" PAGINA-INICIAL="" PAGINA-FINAL="" ISBN="" NOME-DA-EDITORA="" CIDADE-DA-EDITORA="" NOME-DO-EVENTO-INGLES="" />
  <AUTORES NOME-COMPLETO-DO-AUTOR="" NOME-PARA-CITACAO="" ORDEM-DE-AUTORIA="" NRO-ID-CNPQ="" />
</TRABALHO-EM-EVENTOS>
```

Figura 9. Formato de obtenção dos dados

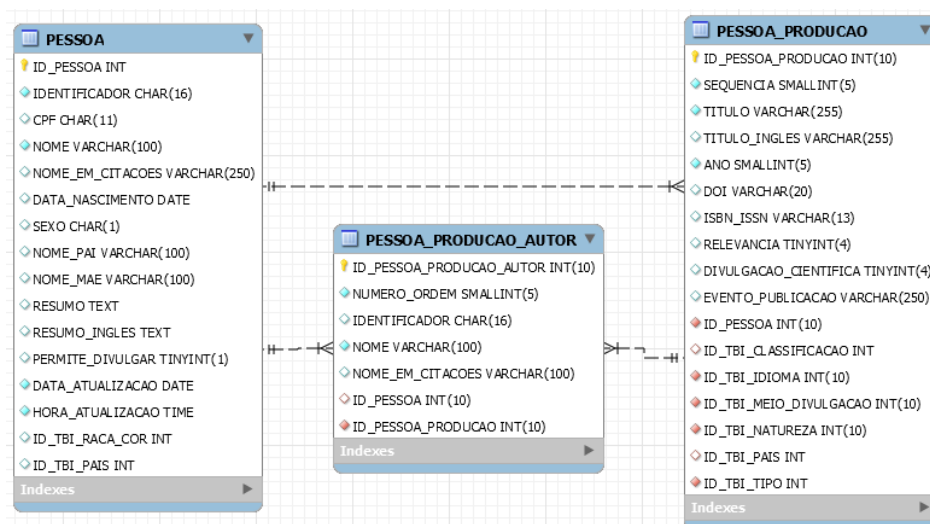


Figura 10. Estrutura da base de dados de produções bibliográficas

5. Resultados

Nesta seção são apresentados os resultados obtidos na construção do *software* de extração de dados quantitativos do Currículo Lattes com foco na apresentação de indicadores institucionais, onde as mesmas estão divididas em subgrupos.

5.1. Detalhamento de funcionalidades

Esta subseção apresenta as funcionalidades disponíveis no sistema de acordo com a Figura 7. A listagem de produções bibliográficas é acessada através de sua seleção na tela principal do sistema, nela constam informações relacionadas à quantidade de produções realizadas, organizadas por tipos de produção dentro de um período de tempo previamente selecionado. Para efetuar a listagem das informações, foram utilizadas estruturas de repetição, onde estas realizam a chamada de outra classe que contém trechos de código

necessários para a seleção dos dados na base de dados, a Figura 11 apresenta o trecho de código relacionado à exibição das informações de acordo com a listagem apresentada na Figura 14.

Para realizar a listagem das produções, um laço de repetição foi utilizado (linha 44), onde são preenchidas as informações da tabela. A estruturação da tabela está subdividida em três partes, sendo a primeira responsável pela listagem dos nomes das produções ({{producao.1}}, linha 49), onde cada uma destas listagens possui um *RadioButton* para capturar seu identificador ({{producao.0}}, linha 48), identificador necessário para realizar a listagem de informações referentes a seleção efetuada na tela seguinte. Já a segunda parte do código é responsável pela listagem do total de produções enquadradas em sua categoria, um novo laço de repetição foi necessário (linha 53), bem como uma comparação de identificadores (linha 54), caso os identificadores forem iguais, o somatório total de produções é apresentada (linha 56). A terceira parte do código esta relacionada ao descritivo dos totais de produção separados pelo seu ano de publicação. Para a realização deste trecho, mais um laço de repetição foi necessário (linha 67), onde dentro dele é realizada a comparação para verificar se a identificação da produção é a mesma das outras informações retornadas (linha 68), caso elas forem iguais, estas informações são representadas na *interface*, onde o trecho ({{somatorio.3}}, linha 70), apresenta a quantidade de produções, e o trecho ({{somatorio.2}}, linha 71) apresenta o seu ano de referência.

A quantidade de produções relacionadas ao seu tipo é apresentada na Figura 14. Caso não haja produções em determinado ano, esta informação não é apresentada na *interface*, apresentando apenas o somatório de produções nos anos onde elas ocorreram. Tornando desta forma a *interface* mais simplificada, retirando informações nulas.

```
44  {% for producao in listProducao %}
45      <tr id="trTitulo">
46          <td class="bg-primary">
47              <input title="" name="rbTipoProducao" required type="radio" value="{{ producao.0 }}">
48              <b>{{ producao.1 }}</b>
49          </td>
50          <td class="bg-primary">
51              {% for listSomatorioTotalProd in listSomatorioTotalProd %}
52              {% if producao.0 == listSomatorioTotalProd.0 %}
53              <p style="...">
54              <i>Total de Produções: {{ listSomatorioTotalProd.1 }}</i>
55              </p>
56              {% endif %}
57              {% endfor %}
58          </td>
59      </tr>
60
61      <tr style="...">
62          <td bgcolor="#e1e4e8"><b>Quantidade</b></td>
63          <td bgcolor="#e1e4e8"><b>Ano</b></td>
64      </tr>
65      {% for somatorio in listSomatorio %}
66      {% if producao.0 == somatorio.0 %}
67      <tr style="...">
68          <td>{{ somatorio.3 }}</td>
69          <td>{{ somatorio.2 }}</td>
70      </tr>
71      {% endif %}
72      {% endfor %}
73      {% empty %}
74      <h4 style="...">Nenhum registro encontrado</h4>
75      {% endfor %}
```

Figura 11. Listagem de produções bibliográficas

Para que seja possível a visualização das informações relacionadas às produções bibliográficas individuais, é necessário que o usuário realize a seleção do profissional desejado, o trecho de código responsável parte desta listagem é apresentado na Figura 12, onde é apresentado o modelo de exibição do descritivo geral, apresentado previamente no resumo de produções de acordo com a Figura 16.

O trecho de código apresentado na Figura 12, demonstra o modelo de exibição do descritivo das produções realizadas pelo profissional selecionado. Para efetuar a representação destas informações tornou-se necessário a utilização de um laço de repetição (linha 83), as produções estão separadas por tipo ({{producao.1}}, linha 87), desta forma caso haja mais de uma produção de um determinado tipo, estas informações são agrupadas dentre de seu respectivo tópico. O descritivo destas produções esta subdividido em três colunas, sendo a primeira responsável pela exibição dos autores envolvidos na publicação ({{producao.2}}, linha 93), a titulação do trabalho ({{producao.3}}, linha 93), o local de sua publicação ({{producao.4}}, linha 94) e o ano a ela referente ({{producao.5}}, linha 94). A segunda coluna possui informações sobre o país de publicação (linha 97), bem como o âmbito do evento (linha 98). A terceira coluna contempla informações relacionadas ao modelo de publicação (linha 101) e ordem de autoria (linha 102).

```

83  {% for producao in listPessoaProducao %}
84      {% ifchanged %}
85          <tr id="trPub">
86              <td colspan="3" id="tdPub" class="bg-primary">
87                  <b> {{ producao.1 }} </b>
88              </td>
89          </tr>
90      {% endifchanged %}
91      <tr>
92          <td>
93              {{ producao.2 }}, <b>{{ producao.3 }}</b>,
94              {{ producao.4 }}, {{ producao.5 }}
95          </td>
96          <td class="col-sm-2">
97              <p>{{ producao.6|default_if_none:"" }} </p>
98              <p>{{ producao.7|default_if_none:"" }} </p>
99          </td>
100         <td class="col-sm-2">
101             <p>{{ producao.8|default_if_none:"" }} </p>
102             <p>{{ producao.9|add:" Autor"|default_if_none:"" }} </p>
103         </td>
104     </tr>
105     {% empty %}

```

Figura 12. Descritivo das produções bibliográficas

Para efetuar o descritivo das produções apresentadas na Figura 16, torna-se necessário a utilização de trechos de código SQL, a Figura 13 apresenta o código necessário para efetuar a listagem do descrito de produções realizadas pelo profissional. A listagem das produções está vinculada a seleção de um período previamente efetuado, (informações capturadas na linha 101 e linha 102), bem como a seleção do profissional desejado (linha 103), estas informações são fundamentais para a geração destes indicadores. O comando necessário para efetuar o descritivo destas produções é representado na consulta (linhas 105 até linha 115), onde a mesma apresenta os dados necessários para efetuar a listagem das informações relacionadas ao descritivo das produções bibliográficas.

```

100 def get_relatorio_by_pessoa(request):
101     ano_min = request.POST.get('minDate')
102     ano_max = request.POST.get('maxDate')
103     id_prod_nome = request.POST.get('rbProdPessoa')
104     cursor = connection.cursor()
105     cursor.execute('SELECT p.NOME , tbi_tipo.DESCRICAO TIPO PRODUCAO, fcGetAutoresProducao(pp.ID_PESSOA PRODUCAO), '
106     'pp.TITULO, pp.EVENTO_PUBLICACAO, pp.ANO, tbi_classif.DESCRICAO as CLASSIFICACAO, tbi_pais.DESCRICAO as PAIS
107     ', tbi_nat.DESCRICAO as NATUREZA, fcGetPosicaoAutorByProducao(pp.ID_PESSOA PRODUCAO, p.ID_PESSOA) POSICAO '
108     'FROM PESSOA PRODUCAO pp '
109     'INNER JOIN PESSOA p ON p.ID_PESSOA = pp.ID_PESSOA '
110     'INNER JOIN TABELA BASICA ITEM tbi_tipo ON tbi_tipo.ID TABELA BASICA ITEM = pp.ID TBI TIPO '
111     'LEFT JOIN TABELA BASICA ITEM tbi_classif ON tbi_classif.ID TABELA BASICA ITEM = pp.ID TBI CLASSIFICACAO '
112     'LEFT JOIN TABELA BASICA ITEM tbi_pais ON tbi_pais.ID TABELA BASICA ITEM = pp.ID TBI PAIS '
113     'LEFT JOIN TABELA BASICA ITEM tbi_nat ON tbi_nat.ID TABELA BASICA ITEM = pp.ID TBI NATUREZA '
114     'WHERE pp.ANO BETWEEN %s AND %s AND %s AND pp.ID_PESSOA = %s '
115     'ORDER BY tbi_tipo.DESCRICAO, p.NOME, pp.ANO, pp.TITULO', [ano_min, ano_max, id_prod_nome])
116     relatorio_individual = cursor.fetchall()
117     cursor.close()
118     return relatorio_individual

```

Figura 13. Comando para apresentação do descritivo de produções individuais

5.2. Produções bibliográficas

O sistema apresenta uma *interface* para análise de dados relacionados a informações do Currículo Lattes. Dentre as informações listadas estão: dados quantitativos de publicações realizadas, agrupadas de acordo com seus respectivos modelos, a especificação individual de cada uma delas, bem como a capacitação docente, que está subdividida em cursos, apresentado o somatório dos profissionais enquadrados em determinado curso, juntamente com o descritivo individual de cada um dos profissionais previamente selecionados.

A geração de relatórios é importante para o acompanhamento das atividades realizadas pelo profissional dentro da área acadêmica, bem como em sua área de atuação. A Figura 14, apresenta os modelos de produções bibliográficas, onde se torna possível a visualização de dados como a quantidade de produções por tipo realizadas em um determinado período de tempo, o qual deve ser previamente selecionado (a figura apresenta um intervalo de três anos).

● Artigo Aceito para Publicação		Total de Produções: 80
Quantidade	Ano	
6	2015	
10	2016	
64	2017	
● Artigo Publicado		Total de Produções: 845
Quantidade	Ano	
341	2015	
337	2016	
167	2017	

Figura 14. Produções bibliográficas

Após a seleção do campo desejado, o sistema apresenta uma *interface* com informações a ela relacionadas. A Figura 15, apresenta informações referentes às produções bibliográficas por tipo, onde são descritas a quantidade de publicações de acordo com o período de tempo selecionado, atribuindo a elas o nome do profissional. Esta por sua vez, fornece ao usuário a opção de realizar a seleção do nome desejado, para que o mesmo apresente as informações a ele referentes. A Figura 16, apresenta estas informações.

● Adriana Dall'Asta Pereira		Total de Produções: 8
Quantidade	Ano	
2	2015	
3	2016	
3	2017	
● Alencar Kolinski Machado		Total de Produções: 19
Quantidade	Ano	
8	2015	
6	2016	
5	2017	

Figura 15. Produções bibliográficas por tipo

As informações relacionadas a publicações trazem dados relevantes para a geração de indicadores institucionais. Para a visualização destes dados, o usuário deve previamente selecionar o profissional desejado, após este processo, o sistema carrega uma *interface* onde são apresentadas as informações relacionadas. Nela se torna possível à visualização dos dados de forma mais ampla, onde são apresentadas todas as publicações efetuadas de acordo com o período de tempo selecionado. A Figura 16 apresenta o modelo de apresentação dos dados relacionados às publicações realizadas, informações sobre: a titulação do artigo, autores envolvidos na publicação, local de publicação, a modalidade do artigo e o ano de sua publicação nela são descritas. Esta *interface* é baseada em uma das *interfaces* presentes no projeto, alterações visuais foram realizadas, mas as informações nela contidas são semelhantes.

Patricia Gomes				
Resumo				
Tipo	Natureza	Classificação	País	Total
Artigo Aceito para Publicação	Não Informado			1
Artigo Publicado	Completo			22
Trabalho em Evento	Completo	Local	Brasil	2
Trabalho em Evento	Resumo	Internacional	Brasil	5
Artigo Aceito para Publicação				
DALCIN, A. J. F.; SANTOS, C. G.; GUNDEL, S. S.; ROGGIA, I.; RAFFIN, R. P.; OURIQUE, A. F.; DOS SANTOS, R. C. V.; GOMES, P., Anti biofilm effect of dihydromyricetin-loaded nanocapsules on urinary catheter infected by <i>Pseudomonas aeruginosa</i> , COLLOIDS AND SURFACES B-BIOINTERFACES, 2017				Não Informado Último Autor
Trabalho em Evento				
SANCHOTENE, R. C.; ROGGIA, I.; DALCIN, A. J. F.; NASCIMENTO, K.; CRUZ, I. B. M.; SAGRILLO, M. R.; OURIQUE, A. F.; GOMES, P., Análise da citotoxicidade de lipossomas contendo guaraná através da expressão antioxidante não enzimática de NPSH, VIII Workshop em Nanociências, 2017		Nacional Brasil		Resumo Último Autor

Figura 16. Produções bibliográficas por pessoa

5.3. Capacitação docente

A capacitação docente é outro dado importante para a análise de níveis institucionais. Na Figura 17 são apresentadas informações relacionadas aos cursos bem como o somatório total de profissionais que nele atuam. Com esta informação torna-se possível a visualização dos dados de uma forma mais ampla quando relacionado a quantidade de profissionais que atuam em determinado curso.

Cursos	Profissionais
○ Medicina	2
○ Engenharia Ambiental e Sanitária	1
○ Nutrição	3
○ Odontologia	5

Figura 17. Capacitação docente por cursos

Após o usuário realizar a seleção do curso desejado, o sistema apresenta uma nova *interface* onde são apresentados os nomes dos profissionais que se enquadram na seleção realizada, bem como o descritivo de seus níveis de titulação, realizando um somatório de profissionais agrupados de acordo com seu maior nível de titulação (exemplo: mestrado, doutorado e afins). Após seleção do profissional desejado, seu descritivo é efetuado, onde são apresentadas informações relacionadas à capacitação docente por pessoa, que contém dados como o seu histórico acadêmico de titulação, tal como título do trabalho, grau de titulação, ano de início e ano de término. A Figura 18, representa o descritivo sobre a capacitação docente relacionada ao profissional selecionado, nela constam informações como o resumo e descritivo sobre as titulações efetuadas.

Juliana Saibt Martins			
Resumo			
Possui graduação em Fisioterapia pela Universidade Federal de Santa Maria (UFSM), mestrado e doutorado em Ciências Biológicas (Bioquímica) pela Universidade Federal de Santa Maria (UFSM) e Universidade Federal do Rio Grande do Sul (UFRGS), respectivamente. Exerce atividades docentes no Curso de Fisioterapia do Centro Universitário Franciscano e no Mestrado em Ciências da Saúde e da Vida. Atualmente compõe o Núcleo Docente Estruturante (NDE) do Curso de Fisioterapia e coordena o curso de Mestrado em Ciências da Saúde e da Vida. Possui experiência na área de Fisioterapia Neurofuncional, Bioquímica e Farmacologia. Atua principalmente nos seguintes temas: Fisioterapia na Saúde da Criança e no Cuidado ao Envelhecimento, Qualidade de vida, Promoção da Saúde, Neuroproteção e Neurotoxicidade. Integrante do grupo de pesquisa ?Neuroproteção? (dgp.cnpq.br/dgp/espelhogrupo/2882316182129676) e líder do grupo de pesquisa ?Promoção da saúde e tecnologias aplicadas à fisioterapia? (dgp.cnpq.br/dgp/espelhogrupo/4416922610310469).			
Titulação			
Título do Trabalho	Titulação	Ano de Início	Ano de Conclusão
"Enfoque educativo da Fisioterapia para uma consciência anti-tabágica em pré-adolescentes."	GRADUACAO	1994	1998
O estudo da plasticidade nervosa como uma contribuição à prática fisioterapêutica.	ESPECIALIZACAO	2000	2001
Efeito da diacereína sobre a febre induzida por diferentes pirógenos em ratos jovens.	MESTRADO	2004	2005
Atividade antipirética e antiinflamatória de derivados 5-trifluormetil-4,5-diidro-1H-1-carboxiamida pirazol em ratos.	DOCTORADO	2006	2010

Figura 18. Capacitação docente por pessoa

6. Conclusão

Este trabalho apresentou uma proposta de desenvolvimento de uma ferramenta para automatização do processo de geração de indicadores institucionais, com o intuito de realizar a construção de uma ferramenta para tornar este processo mais ágil e seguro, tendo em vista que o processo realizado de forma manual está sujeito a falhas. O foco principal na implementação deste sistema possui relação com o auxílio a obtenção desses dados, agilizando e reduzindo possíveis erros envolvidos no processo de obtenção desses indicadores. Como resultados obtidos na geração de indicadores institucionais, estão as capacitação docente e produções bibliográficas, sendo estas listagens relacionadas a vida acadêmica do profissional, como suas áreas de atuação, níveis de formação, dentre outras.

Quando relacionado aos trabalhos correlatos, buscou-se a identificação de metodologias utilizadas no desenvolvimento, para desta forma tornar possível a implementação do projeto. Em geral os trabalhos possuem similaridades, porém o

ScriptLattes é o que mais se assemelha, quando relacionado a geração de relatórios institucionais, uma vez que o Semantic Lattes trabalha com ontologias.

O projeto foi desenvolvido seguindo técnicas da metodologia Scrum, com o uso das linguagens de programação Python e *framework* Django. Como projetos futuros, pretende-se:

- Implementar a segurança no acesso a estas informações, bem como realizar a segurança contra ataques ao banco de dados.
- O complemento de alguns relatórios, atribuindo novas listagens de informações, tais como: a organização das produções bibliográficas por áreas de conhecimento e a representação gráfica dos dados obtidos nos demais relatórios.
- A exportação destes dados em PDF.

Referências

- Beck, K. (2001). Manifesto para o desenvolvimento ágil de software. <http://www.manifestoagil.com.br/>, [accessed on Mar 21].
- Borges, L. E. (2010). *Python para Desenvolvedores*. 2º ed. Rio de Janeiro: .
- BRQ (2017). Metodologias Ágeis – SCRUM - BRQ. <http://www.brq.com/metodologias-ageis/>, [accessed on Mar 21].
- CNPq a (2017). Como os dados são obtidos - Plataforma Lattes - CNPq. <http://lattes.cnpq.br/web/dgp/como-os-dados-sao-obtidos/>, [accessed on Mar 20].
- CNPq b (2017). Censo atual - Plataforma Lattes - CNPq. <http://lattes.cnpq.br/web/dgp/centso-atual/>, [accessed on Apr 1].
- Costa, A. P. Da; Yamate, F. S. (2009). Semantic Lattes: uma ferramenta de consulta de informações acadêmicas da base Lattes baseada em ontologias.
- Django Software Foundation (2017). A estrutura da Web para perfeccionistas com prazos | Django. <https://www.djangoproject.com/>, [accessed on Apr 1].
- Highsmith, J. (2001). História: O Manifesto Ágil. <http://agilemanifesto.org/history.html>, [accessed on Mar 21].
- Lotar, A. (2010). *Como programar com ASP.NET e C#*. 2º ed. São Paulo: Editora Novatec.
- Mena-Chalco, J. P.; Junior, R. M. C. (dec 2009). scriptLattes: an open-source knowledge extraction system from the Lattes platform. *Journal of the Brazilian Computer Society*, v. 15, n. 4, p. 31–39.
- Sares, W. (2001). *MySQL Conceitos e Aplicações*. São Paulo: Editora Érica.