

UNIVERSIDADE FEDERAL DO PAMPA

Murilo Dutra Meglioransi

Um Sistema para Coleta e Apresentação de
Indicadores Bibliométricos da Plataforma
Lattes

Alegrete
2019

Murilo Dutra Meglioransi

**Um Sistema para Coleta e Apresentação de
Indicadores Bibliométricos da Plataforma Lattes**

Trabalho de Conclusão de Curso apresentado
ao Curso de Graduação em Ciência da Com-
putação da Universidade Federal do Pampa
como requisito parcial para a obtenção do tí-
tulo de Bacharel em Ciência da Computação.

Orientador: Prof. Me. Diego Luis Kreutz

Coorientador: Dr. Fabio Natanael Kepler

Alegrete
2019

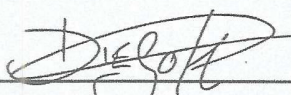
Murilo Dutra Meglioransi

Um Sistema para Coleta e Apresentação de Indicadores Bibliométricos da Plataforma Lattes

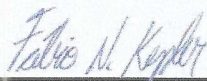
Trabalho de Conclusão de Curso apresentado ao Curso de Graduação em Ciência da Computação da Universidade Federal do Pampa como requisito parcial para a obtenção do título de Bacharel em Ciência da Computação.

Trabalho de Conclusão de Curso defendido e aprovado em 25 de junho de 2019.

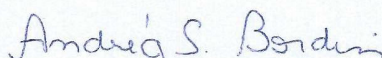
Banca examinadora:



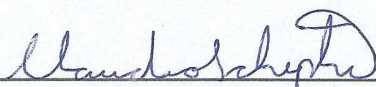
Prof. Me. Diego Luis Kreutz
Orientador
UNIPAMPA



Dr. Fabio Natanael Kepler
Coorientador
Unbabel, Portugal



Profa. Dra. Andréa Sabedra Bordin
UNIPAMPA



Prof. Dr. Claudio Schepke
UNIPAMPA

RESUMO

A Internet é o meio mais utilizado para a divulgação de publicações científicas. A razão disso é a proliferação acelerada de bases científicas como a IEEE Xplore, o Google Acadêmico e a Web of Science. No Brasil, a Plataforma Lattes agrupa uma das mais completas coleções de dados bibliográficos dos pesquisadores. Sistemas e pesquisas utilizam esses dados para gerar indicadores e estatísticas de produtividade e colaboração, o que é denominado de bibliometria. Os sistemas existentes oferecem indicadores quantitativos de produção pela *web* e permitem visualizar o desempenho de instituições e pesquisadores. Entretanto, alguns sistemas não podem ser acessados ou distribuídos livremente, outros carecem de alguma funcionalidade como a geração de redes de colaboração, perfil do pesquisador ou indicadores de uma instituição. Este trabalho implementa, entrega e avalia um processo automatizado de extração de dados e um sistema *web* que oferece indicadores bibliométricos de produção e colaboração, com foco em uma instituição e seus pesquisadores. O processo é constituído por duas partes: a coleta dos dados para o banco de dados e o desenvolvimento da aplicação *web* Openlattes. Utilizando os currículos dos docentes da UNIPAMPA, os dados foram armazenados com sucesso no MongoDB e, conseqüentemente, acessados dinamicamente pela aplicação *web*. O Openlattes apresentou indicadores interativos de produção e coautoria incluindo funcionalidades como, por exemplo, especificar para quais professores o indicador será gerado. A experiência do usuário do Openlattes foi avaliada como positiva e agradável.

Palavras-chave: Indicadores. Currículo Lattes. Bibliometria. Aplicação *Web*. Visualização de Dados. Produção Científica.

ABSTRACT

The Internet is the most widely used medium for the dissemination of scientific publications. The reason for this is an accelerated proliferation of scientific bases such as IEEE Xplore, Google Scholar and Web of Science. In Brazil, the Lattes Platform stores one of the most complete collections of researchers bibliographical data. Software systems and research use this data to generate indicators and statistics of productivity and collaboration, which is known as bibliometrics. Existing systems provide quantitative indicators of production over the web and allow visualizing the performance of institutions and researchers. However, some systems cannot be accessed or distributed freely, others lack some functionality such as the generation of collaboration networks, researcher's profile or indicators of an institution. This work implements, delivers and evaluates an automated process of data extraction and a web system that provide production and collaboration bibliometric indicators, focusing on an institution and its researchers. The process consists of two parts: collection of the data to the database and the development of the web application Openlattes. Using the teachers curricula of UNIPAMPA, the data was successfully stored in MongoDB and, consequently, accessed dynamically by the web application. Openlattes presented production and coauthorship interactive indicators including features such as, for example, specifying to which teachers the indicator will be generated. The Openlattes user experience was evaluated as positive and enjoyable.

Key-words: Indicators. Lattes Curriculum. Bibliometrics. Web Application. Data Visualization. Scientific Production.

LISTA DE FIGURAS

Figura 1 – Estrutura simplificada do currículo em XML.	21
Figura 2 – Indicadores de um currículo Lattes.	23
Figura 3 – Uma das páginas com indicadores da UFABC gerados pelo scriptLattes.	24
Figura 4 – Uma parte do grafo ponderado da UFABC gerado pelo scriptLattes.	25
Figura 5 – Página de indicadores da UFMG gerados pelo Somos.	26
Figura 6 – Indicadores das produções da Stela Experta.	27
Figura 7 – Mapa de palavras-chave da Stela Experta.	28
Figura 8 – Indicadores bibliométricos do eLattes.	29
Figura 9 – Arquitetura e fluxo de dados do sistema.	32
Figura 10 – Diagrama de classe representando o esquema do Banco de Dados.	37
Figura 11 – Formato do indicador de produção retornado pela API.	39
Figura 12 – Formato do grafo retornado pela API.	40
Figura 13 – Como o indicador de produção é gerado.	41
Figura 14 – Como o grafo de coautoria é gerado.	41
Figura 15 – Formato de um grupo armazenado no navegador.	42
Figura 16 – Interface do usuário da página inicial do Openlattes.	44
Figura 17 – Tabela para visualizar as seleção salvas.	44
Figura 18 – Página dos indicadores das produções bibliográficas.	45
Figura 19 – Página dos indicadores das orientações.	46
Figura 20 – Tabela listando as produções da coluna 2011.	47
Figura 21 – Página dos indicadores das produções bibliográficas com filtros aplicados.	48
Figura 22 – Indicadores de produção classificado por professor.	49
Figura 23 – Rede de coautoria de todos os professores e tipos de produções bibliográficas.	50
Figura 24 – Rede de coautoria com filtros aplicados.	51
Figura 25 – Página para gerenciar grupos.	52
Figura 26 – Qualidade do Openlattes em relação a outros produtos.	53

LISTA DE TABELAS

Tabela 1 – Tipos de produção acadêmica extraídos do currículo em XML pelo scriptLattes.	21
Tabela 2 – Análise comparativa dos sistemas bibliométricos.	30
Tabela 3 – Resquisições e opções da API.	38
Tabela 4 – Requisições da API para alterar grupos.	40
Tabela 5 – Relação entre itens do questionário e qualidade avaliada.	49
Tabela 6 – Resultado da versão curta do UEQ.	52

LISTA DE SIGLAS

API Application Programming Interface

BD Banco de Dados

CPF Cadastro de Pessoas Físicas

CSS Cascading Style Sheets

CSV Comma Separated Values

HTML HyperText Markup Language

HTTP Hypertext Transfer Protocol

ID identificador

IES Instituição de Ensino Superior

JS JavaScript

JSON JavaScript Object Notation

PL Plataforma Lattes

SGBD Sistema Gerenciador de Banco de Dados

SPA Single-Page Application

UEQ User Experience Questionnaire

UFMG Universidade Federal de Minas Gerais

UFSCar Universidade Federal de São Carlos

UNIPAMPA Universidade Federal do Pampa

XML Extensible Markup Language

SUMÁRIO

1	INTRODUÇÃO	17
1.1	Objetivos	17
1.2	Organização do Trabalho	18
2	CONCEITOS BÁSICOS	19
2.1	Bibliometria	19
2.2	O Currículo Lattes	19
2.3	scriptLattes	20
3	TRABALHOS RELACIONADOS	23
3.1	Indicadores do Currículo Lattes	23
3.2	scriptLattes	23
3.3	Sistema Somos	24
3.4	Stela Experta	25
3.5	e-Lattes	27
3.6	Análise Comparativa	29
4	SOLUÇÃO PROPOSTA	31
4.1	O Sistema Proposto	31
4.2	A Arquitetura do Sistema	31
4.3	Extração de Dados	32
4.4	Aplicação <i>Web</i>	33
4.5	Armazenamento de Dados	33
5	IMPLEMENTAÇÃO	35
5.1	Principais Tecnologias	35
5.2	Estrutura do Banco de Dados	36
5.3	Extração de Dados	37
5.4	API	38
5.5	Indicador de Produção	40
5.6	Grafo de Coautoria	40
5.7	Armazenamento de Grupos no Navegador	41
6	RESULTADOS	43
6.1	Extração e Geração do Banco de Dados	43
6.2	Página Inicial	43
6.3	Lista e Seleção de Professores	43
6.4	Indicadores de Produção	45
6.5	Redes de Coautorias	47

6.6	Página de Administração	47
6.7	Avaliação da Solução Proposta	48
7	CONSIDERAÇÕES FINAIS	55
7.1	Trabalhos Futuros	55
	REFERÊNCIAS	57
	APÊNDICES	61
	APÊNDICE A – LISTA DE TAREFAS	63
	APÊNDICE B – QUESTIONÁRIO	65

1 INTRODUÇÃO

A Internet é uma ferramenta importante para a comunicação científica, pois é através dela que pesquisadores comumente divulgam suas pesquisas e encontram trabalhos relacionados. Como resultado, as bases de dados científicas se tornaram essenciais para o armazenamento e catalogação *online* do conhecimento. A IEEE Xplorer, a ACM Digital Library, o Google Acadêmico, a Web of Science e o Scopus são exemplos de importantes bases científicas da atualidade.

A Plataforma Lattes (PL)¹, tornou-se uma importante base de dados da produção acadêmica brasileira, porque armazena essencialmente informações dos currículos dos pesquisadores. O currículo Lattes inclui identificação, formação acadêmica, áreas de atuação, prêmios, produções bibliográficas, entre outras informações (AMORIN, 2003). Em particular, a produção bibliográfica é um dos conjuntos de dados mais relevantes porque permite gerar indicadores de produção científica (GREGOLIN et al., 2005).

Hoje, a maior deficiência é a falta de acesso simples e flexível aos indicadores de produção dos pesquisadores e das instituições. Considerando a Universidade Federal do Pampa (UNIPAMPA) como caso de uso, não há uma forma simples e direta de ter acesso aos indicadores de produção dos pesquisadores da instituição. Uma das alternativas é utilizar o scriptLattes (MENA-CHALCO; CESAR-JR, 2013), que gera gráficos de produção e grafos de colaboração. Entretanto, organiza seus indicadores em um conjunto de páginas estáticas, dificultando eventuais manutenções.

A própria PL oferece, de forma limitada, alguns indicadores. Porém, os indicadores estão restritos ao escopo de um currículo (CNPQ, 2018). Outros sistemas, como o Somos (FUNDEP, 2018) e a Stela Experta (STELATEK, 2018), não apresentam redes de colaborações e possuem restrições de uso ou de acesso.

Com o intuito de contribuir para mudar o cenário atual, este trabalho propõe a criação de um processo bem definido e automatizado de coleta dos dados da PL e um sistema *web* que permita consultas flexíveis, dinâmicas e visualização gráfica dos dados bibliométricos de uma instituição (exemplo: uma Universidade). Esse sistema pode ser usado pelos gestores da instituição no apoio a tomada de decisões, além de permitir que alunos, professores, pesquisadores e até mesmo o público geral utilize-o para se informar sobre a performance e a colaboração científica da instituição.

1.1 Objetivos

O objetivo geral é criar um sistema *web* que facilite a disponibilização e o acesso a indicadores bibliométricos de instituições de ensino e pesquisa a partir dos dados providos pela PL. Os objetivos específicos são:

1. Automatizar a extração de informações dos currículos Lattes.

¹ Disponível em <<http://lattes.cnpq.br/>>.

2. Armazenar os dados em um Banco de Dados (BD).
3. Desenvolver uma aplicação *web* que suporte visualização interativa de indicadores.
4. Integrar o processo de extração de dados com a aplicação *web* em um único sistema.
5. Avaliar a aplicação *web* em relação a experiência do usuário, utilizando a UNIPAMPA como um caso de uso.

1.2 Organização do Trabalho

Os próximos capítulos estão organizados da seguinte forma. O Capítulo 2 apresenta os conceitos fundamentais para o entendimento do trabalho. O Capítulo 3 descreve e compara os trabalhos relacionados. O Capítulo 4 apresenta características e o funcionamento da solução proposta. O Capítulo 5 apresenta a implementação do sistema proposto. O Capítulo 6 apresenta os resultados obtidos utilizando os currículos dos docentes da UNIPAMPA. Por fim, o Capítulo 7 discute os resultados e os trabalhos futuros.

2 CONCEITOS BÁSICOS

Este capítulo apresenta os conceitos básicos para o entendimento do trabalho. A seção 2.1 introduz os principais conceitos da bibliometria. A seção 2.2 descreve o conteúdo do currículo Lattes e sua utilidade como fonte para a criação de indicadores. Finalmente, a seção 2.3 apresenta o funcionamento da ferramenta scriptLattes, um componente importante da solução proposta.

2.1 Bibliometria

A bibliometria é o processo de extrair e analisar dados de produções para entender o que elas representam (BELTER, 2015). O termo foi sugerido por Pritchard (1969) como uma melhor forma de se referir a estatísticas de bibliografia. A bibliometria é usada na construção de uma variedade de indicadores, mas os mais simples são recomendados porque dependem de poucos dados bibliográficos e seu cálculo é transparente (WILDGAARD; SCHNEIDER; LARSEN, 2014). Entre alguns desses indicadores estão os indicadores de produção e colaboração (GUTIÉRREZ-SALCEDO et al., 2017; THOMSON REUTERS, 2008).

Indicadores de produção são a medida bibliométrica mais básica e estão associados ao número total de publicações (THOMSON REUTERS, 2008). Características da publicação, como ano e tipo, são consideradas para definir o indicador. Por exemplo, o total de publicações em cada ano representa a evolução das publicações em relação ao tempo.

Indicadores de colaboração mostram a relação entre autores e instituições. De acordo com Gutiérrez-Salcedo et al. (2017), a rede de colaboração mais comum é a rede de coautoria e pode ser usada para identificar, por exemplo, grupos de colaboradores, comunidades isoladas e autores influentes. A rede de coautoria é uma rede não direcionada, onde o nó representa os autores e a ligação entre dois nós mostra que esses autores produziram juntos pelo menos uma vez (DING, 2011).

Indicadores podem ajudar pesquisadores a escolher em qual periódico submeter o seu trabalho e podem auxiliar organizações na tomada de decisões em relação a nomeações, promoções e financiamento (DURIEUX; GEVENOIS, 2010). A bibliometria também é usada para avaliar a performance de pesquisa por formuladores de políticas, diretores e administradores de pesquisa, especialistas de informação e bibliotecários (THOMSON REUTERS, 2008).

2.2 O Currículo Lattes

O currículo Lattes é um currículo eletrônico disponível publicamente na Internet. É usado por pesquisadores, estudantes, professores e outros profissionais. As informações

são criadas e mantidas pelo usuário. No final de 2016, já haviam mais de 3,5 milhões¹ de currículos cadastrados. Sua popularidade pode estar associada ao fato de que é usada para avaliar candidatos na obtenção de bolsas e auxílios (AMORIN, 2003).

As informações do currículo estão divididas entre dados gerais e dados de produção (AMORIN, 2003). Os dados gerais incluem identificação, endereço, titulação, idiomas, entre outras informações básicas. Os dados de produção incluem a lista de produções bibliográficas, técnicas, artísticas e culturais, além das orientações concluídas e outras informações. Finalmente, cada produção é identificada por um título, autores, ano e demais informações que dependem do tipo da produção.

Bassoli (2017) avaliou o currículo Lattes para a geração de indicadores bibliométricos, usando a Universidade Federal de São Carlos (UFSCar) como caso de uso. A principal conclusão da pesquisa é que os metadados dos currículos são suficientemente completos e de qualidade para essa finalidade. Consequentemente, isto significa que as informações do currículo podem ser usadas para avaliar instituições de ensino e pesquisa. Além disso, o seu conteúdo é mais completo do que a Web of Science, outra possível fonte de dados bibliométricos.

Outros estudos usaram a PL como fonte de informação (AMARAL et al., 2016; MENA-CHALCO; DIGIAMPIETRI; OLIVEIRA, 2012). A PL permite criar um panorama da produção científica e dos programas de pós-graduação. As informações da plataforma podem também ser utilizadas para avaliar um programa de pós-graduação (MACIEL et al., 2017). Porém, a PL tem a desvantagem de que seus dados são preenchidos manualmente pelos seus usuários, então pode haver erros de digitação (SILVA; SMIT, 2009).

2.3 scriptLattes

O scriptLattes (MENA-CHALCO; CESAR-JR, 2013) é um software livre² desenvolvido para extrair dados do Lattes. Na sua versão original, os dados são extraídos diretamente da página HyperText Markup Language (HTML) do currículo. Entretanto, esta opção tornou-se difícil com a adição de um *captcha* no acesso aos currículos disponibilizados *online*. A outra opção é utilizar o formato Extensible Markup Language (XML).

A Figura 1 mostra uma versão simplificada do formato do currículo em XML. Cada currículo inclui um identificador e uma data de atualização. Informações básicas do usuário estão dentro do elemento *DADOS-GERAIS*. Produções como trabalho em eventos, artigos publicados, livros e capítulos são listados no elemento *PRODUCAO-BIBLIOGRAFICA*. Produções como *softwares*, patentes, desenhos industriais estão no

¹ Estatísticas da base de currículos Lattes, como o número total de currículos cadastrados, estão disponíveis no painel Lattes: <<http://estatico.cnpq.br/painelLattes/>>.

² O código-fonte está disponível em <<https://bitbucket.org/scriptlattes/scriptlattes>>.

elemento *PRODUCAO-TECNICA*. Produções artísticas e culturais são listadas no elemento *OUTRA-PRODUCAO*. Por fim, dados como formação complementar, participações em bancas, eventos e congressos são listados em *DADOS-COMPLEMENTARES*.

Figura 1 – Estrutura simplificada do currículo em XML.

```
<CURRICULO-VITAE NUMERO-IDENTIFICADOR DATA-ATUALIZACAO >
  <DADOS-GERAIS NOME-COMPLETO NOME-EM-CITACOES-BIBLIOGRAFICAS PAIS-DE-NACIONALIDADE >
    <RESUMO-CV />
    <OUTRAS-INFORMACOES-RELEVANTES />
    <ENDERECO></ENDERECO>
    <FORMACAO-ACADEMICA-TITULACAO></FORMACAO-ACADEMICA-TITULACAO>
    <ATUACOES-PROFISSIONAIS></ATUACOES-PROFISSIONAIS>>
    <AREAS-DE-ATUACAO></AREAS-DE-ATUACAO>
    <IDIOMAS></IDIOMAS>
    <PREMIOS-TITULOS></PREMIOS-TITULOS>
  </DADOS-GERAIS>
  <PRODUCAO-BIBLIOGRAFICA></PRODUCAO-BIBLIOGRAFICA>
  <PRODUCAO-TECNICA></PRODUCAO-TECNICA>
  <OUTRA-PRODUCAO></OUTRA-PRODUCAO>
  <DADOS-COMPLEMENTARES></DADOS-COMPLEMENTARES>
</CURRICULO-VITAE>
```

Fonte: elaborado pelo autor.

O scriptLattes extrai as informações básicas e a lista de produções de cada currículo. A Tabela 1 apresenta os tipos de produções extraídos do currículo em XML. As produções estão distribuídas em diversos tipos que podem ser uma (A) produção bibliográfica ou (B) orientação concluída. No final da execução, o scriptLattes armazena uma lista com todas as produções do grupo de currículos processados.

Tabela 1 – Tipos de produção acadêmica extraídos do currículo em XML pelo scriptLattes.

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. Supervisões e orientações concluídas
Supervisão de pós-doutorado
Tese de doutorado
Dissertação de mestrado

Fonte: Adaptado de MENA-CHALCO e CESAR-JR (2013)

Outra função do *script* é remover produções duplicadas (MENA-CHALCO; CESAR-JR, 2013). Caso dois currículos possuírem a mesma publicação por motivos de coautoria, essa publicação é extraída duas vezes. Por isso, o *script* compara todas as produções com títulos aproximados ou idênticos e somente uma será compilada na lista final. Os títulos são comparados utilizando a distância Levenshtein e precisam ser pelo menos 80% similares (MENA-CHALCO; CESAR-JR, 2013; NAVARRO, 2001). A lista de coautores é identificada junto com a publicação.

A remoção de duplicações é essencial para a criação de indicadores de produção porque nesse tipo de indicador é feita a contagem total de produções. Se a lista de produções possuir duplicações, a contagem estará comprometida.

Para o *scriptLattes*, produções duplicadas também identificam colaborações. Por isso, esse processo é aproveitado para criar um grafo de coautoria. A partir da lista de coautores presente em cada publicação, um método chamado *calcularCombinacoes* identifica todas as coautorias em pares. Por exemplo, para os coautores A_1 , A_2 e A_3 são retornados 3 pares: A_1A_2 , A_1A_3 e A_2A_3 . Esse método é usado para gerar a matriz de adjacência que representa o grafo de coautoria. No grafo, o nó identifica um autor e a aresta identifica uma produção em comum.

Para executar o *script*, é necessário indicar o identificador (ID) Lattes dos currículos que serão extraídos em um arquivo de entrada (FRANCISCO, 2014). Opcionalmente, é possível indicar o nome do pesquisador, o período e o rótulo. O período permite definir o ano das publicações que serão extraídas e o rótulo é usado para classificar o grafo de coautoria.

A última etapa do *scriptLattes* é gerar um conjunto de páginas HTML (FERRAZ; QUONIAM, 2013) com relatórios sobre o grupo processado (mais detalhes na seção 3.2 dos Trabalhos Relacionados). O trabalho proposto aprimora isto com a adição dos dados extraídos em um BD e uma nova aplicação *web* para apresentar os indicadores.

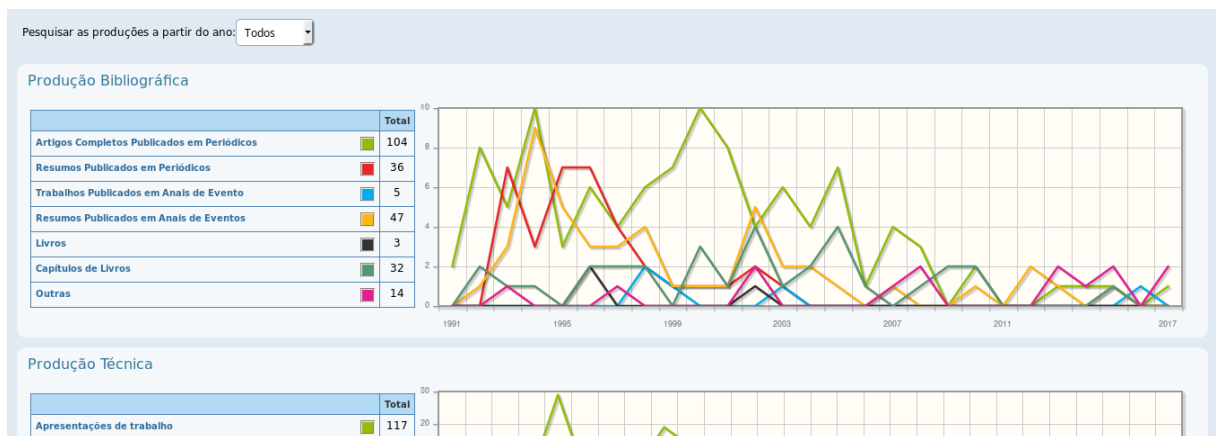
3 TRABALHOS RELACIONADOS

Este capítulo apresenta cinco soluções que disponibilizam dados bibliométricos do currículo Lattes na *web*. O capítulo é finalizado com uma comparação entre eles.

3.1 Indicadores do Currículo Lattes

A própria PL (CNPQ, 2018) disponibiliza a opção para visualizar indicadores de um currículo individual, como ilustrado na Figura 2. Entre os indicadores, encontram-se gráficos quantitativos de produções bibliográficas e técnicas, separados pelo seu tipo. Há, também, gráficos para periódicos mais publicados e autores com quem o pesquisador mais publicou. O usuário pode escolher indicadores de um ano específico.

Figura 2 – Indicadores de um currículo Lattes.



Fonte: CNPq (2018)

3.2 scriptLattes

O scriptLattes, além de ser uma ferramenta de extração de dados da PL, gera indicadores bibliométricos (MENA-CHALCO; CESAR-JR, 2013). A ferramenta gera um grupo de páginas em HTML com indicadores de produções bibliográficas, produções técnicas, produções artísticas, orientações, projetos, prêmios e eventos, além de um grafo de colaboração (FERRAZ; QUONIAM, 2013).

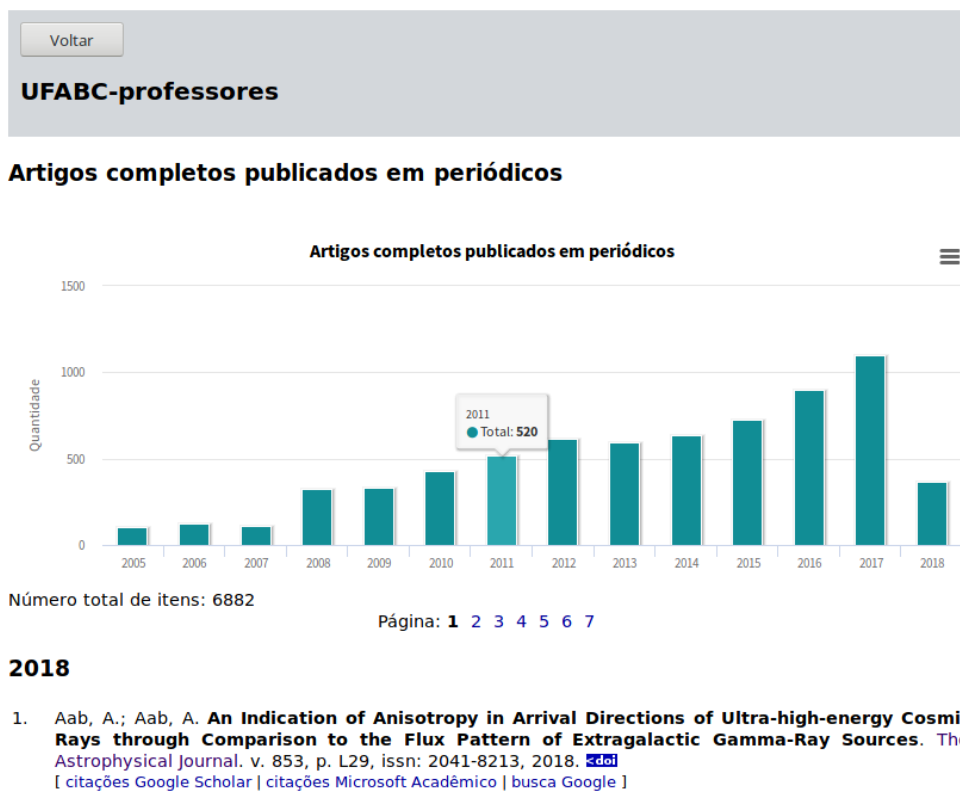
A Figura 3 apresenta um exemplo de saída dos relatórios gerados e disponibilizados online (UFABC, 2019). Para cada tipo de produção, uma página é gerada com um gráfico em barras com o total de produções em cada ano. Abaixo do gráfico, a bibliografia contabilizada pelo indicador é listada.

A página com a rede de colaborações apresenta um grafo não ponderado, onde os autores estão ligados caso exista uma produção em comum entre eles, e um grafo

ponderado, onde o valor da aresta está relacionado a quantidade de colaborações entre os autores. A Figura 4 apresenta uma parte do grafo ponderado que é representado totalmente em uma única imagem e, como há uma grande quantidade de nós, a qualidade da imagem não permite ampliar o suficiente para visualizar o nome do nó e o valor da aresta.

Algumas funcionalidades não estão presentes no scriptLattes (MENA-CHALCO; CESAR-JR, 2013). Não é possível adicionar novos currículos aos indicadores já processados. Para isso, é necessário processar todo o grupo novamente. Além disso, o *script* não extrai informações como formação acadêmica e áreas de atuação. Essa última é importante para a criação de indicadores com base na titulação e área do pesquisador.

Figura 3 – Uma das páginas com indicadores da UFABC gerados pelo scriptLattes.

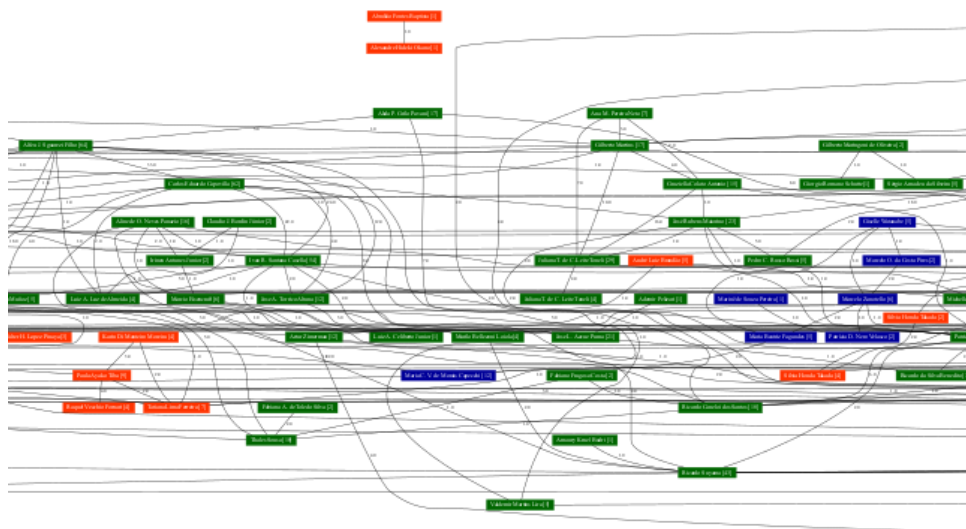


Fonte: UFABC (2019)

3.3 Sistema Somos

O Somos é um sistema que apresenta informações sobre especialidades e indicadores de produção de centros acadêmicos e científicos (FUNDEP, 2018). Exceto os dados sobre laboratórios e propriedade intelectual, todas as informações são extraídas da PL. Esse sistema é comercializado e pode ser adquirido por outras instituições.

Figura 4 – Uma parte do grafo ponderado da UFABC gerado pelo scriptLattes.



Fonte: UFABC (2019)

Como exemplo, temos o Somos aplicado aos pesquisadores associados à Universidade Federal de Minas Gerais (UFMG) (UFMG, 2019). A Figura 5 apresenta uma página que inclui indicadores gerais da instituição, ou seja, dados agrupados de todos os pesquisadores que foram processados pela ferramenta. Pode-se especificar, através de uma busca, indicadores de um professor, uma unidade acadêmica ou um departamento. Grafos de colaboração não são apresentados com essa ferramenta.

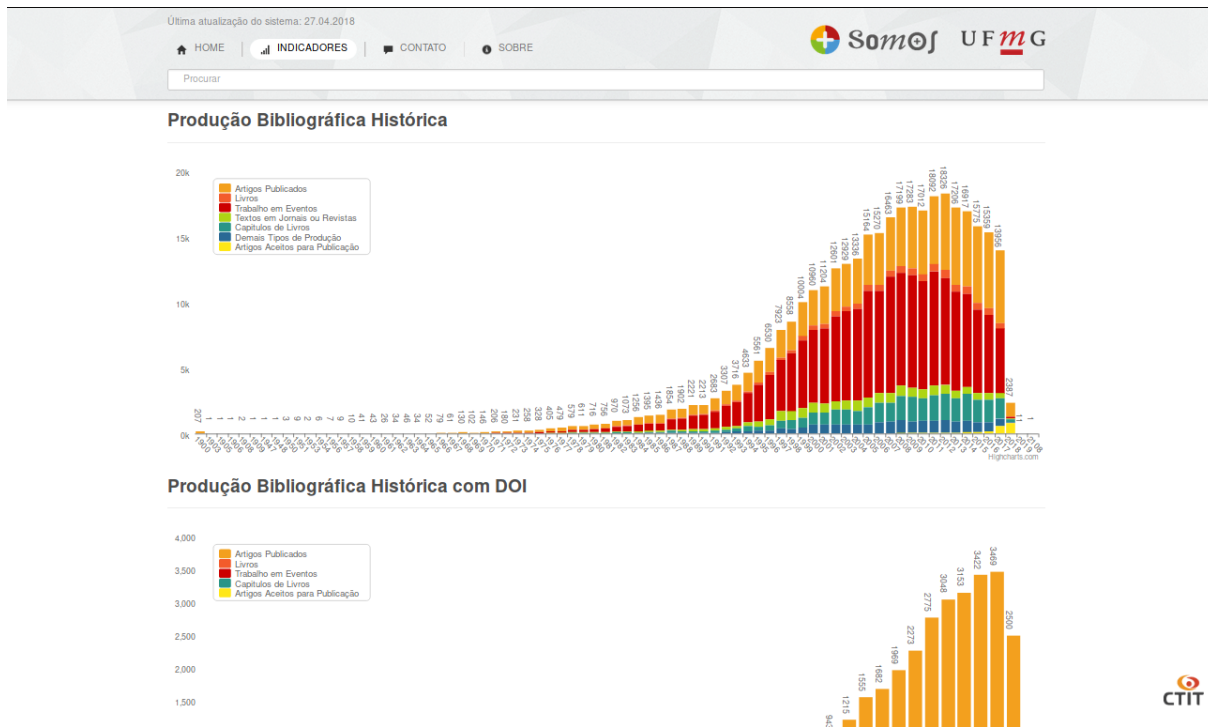
Na página de indicadores, encontram-se gráficos separados para mostrar a evolução das produções, ou seja, total de produções em cada ano. Gráficos em barra mostram o total de produções bibliográficas e orientações concluídas. Gráficos em pizza mostram o total de artigos e patentes por unidade acadêmica. Nos gráficos em barra, ainda é possível filtrar os tipos que serão contabilizados.

3.4 Stela Experta

A Stela Experta é um sistema *web* para visualizar informações de uma Instituição de Ensino Superior (IES) com o objetivo de auxiliar nas suas atividades de gestão (STELATEK, 2018). Indicadores bibliométricos são apresentados para dados de produções e projetos. Algumas informações são retiradas do Lattes dos seus docentes e discentes, enquanto outras são informadas pela instituição. A ferramenta é comercializada e tem acesso restrito, portanto, uma versão de demonstração foi usada com um conjunto de currículos de exemplo.

Há três formas distintas para visualizar os indicadores: perfil (indicadores de produção), temáticas (mapa de palavras-chave) e comparativos (indicadores de produção

Figura 5 – Página de indicadores da UFMG gerados pelo Somos.



Fonte: UFMG (2019)

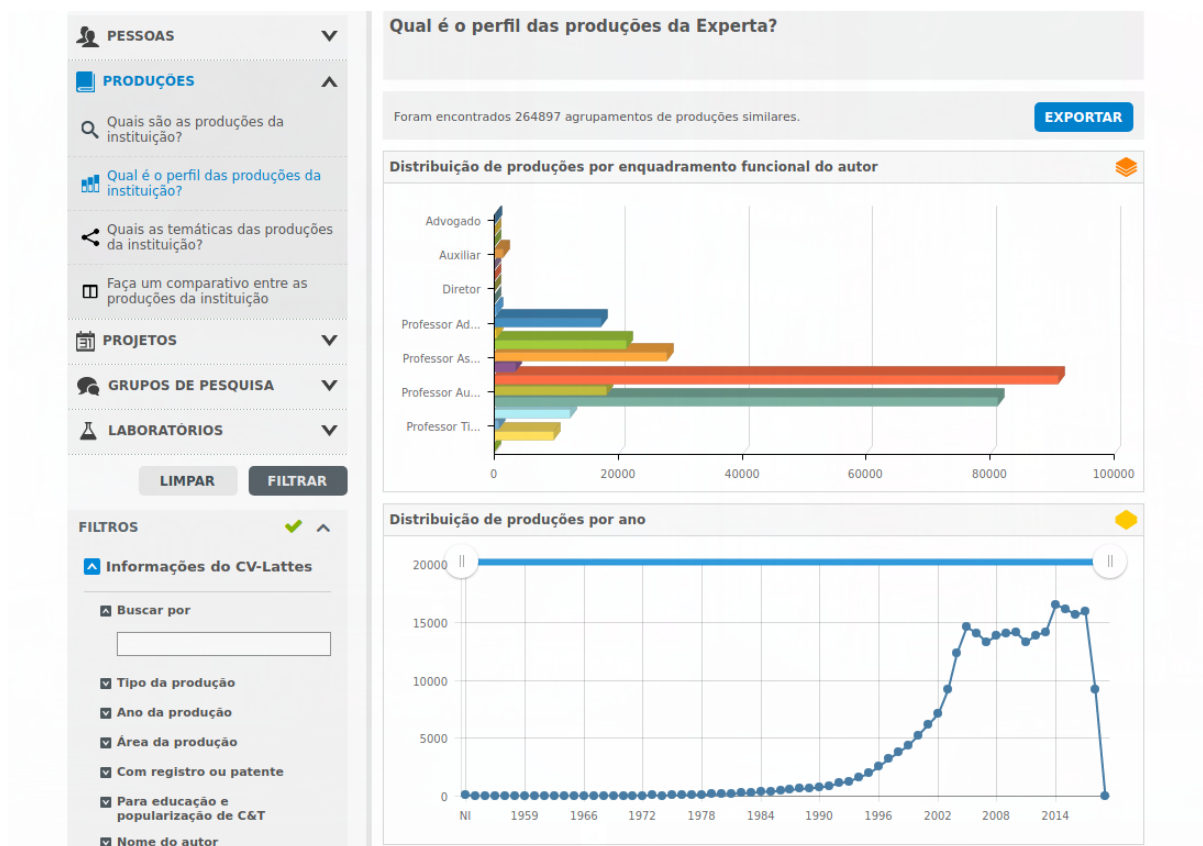
em gráficos em pilha). Cada uma é apresentada em uma página com opções de filtros disponibilizados com base na visualização e tipo de dado escolhido.

A Figura 6 apresenta a interface ao selecionar o perfil das produções que são classificadas em vários tipos: por enquadramento funcional do autor (professor, diretor, professor adjunto, professor assistente, etc), por ano, por grande área do conhecimento, por tipo de produção (bibliográfica, técnica ou artística), entre outros. O perfil dos projetos tem uma classificação semelhante: por enquadramento funcional do participante (auxiliar, professor adjunto, professor titular, etc), por natureza, por ano de início, entre outros. Cada um recebe um gráfico em barra, em linha ou em pizza, apresentados um abaixo do outro.

Filtros são apresentados conforme o tipo de dado e visualização selecionados. Para as produções, é possível filtrar por tipo de produção, ano de produção, área de produção, com registro ou patente, nome do autor, produção do grupo de pesquisa, etc. Já para os projetos, é possível filtrar por: instituição de execução, país da instituição de execução, instituição financiadora, país da instituição financiadora, ano de início, ano de fim, natureza do projeto, etc. Além disso, ao clicar em um gráfico, o usuário é redirecionado para uma lista dos dados aplicando o filtro conforme o item do gráfico clicado.

Na visualização de comparativos são apresentados os mesmos gráficos, porém em pilha. É necessário escolher pelo menos um tipo de dado para ser comparado na pilha

Figura 6 – Indicadores das produções da Stela Experta.



Fonte: StelaTek (2018)

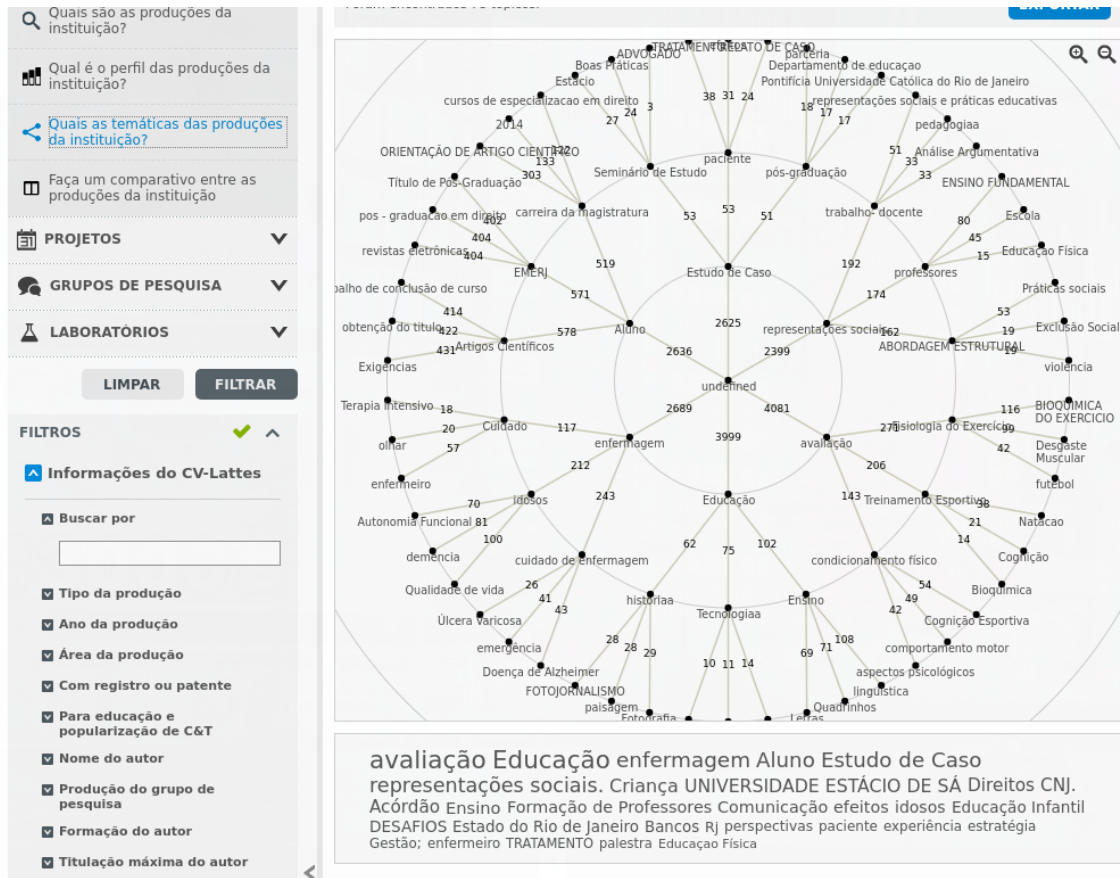
do gráfico. Entre as produções estão disponíveis: nome do autor, curso de atuação do autor, lotação institucional do autor, titulação máxima do autor, etc. Já entre os projetos estão disponíveis: nome de quem informou, curso de atuação do participante, lotação institucional do participante, titulação máxima do participante, etc.

A Figura 7 apresenta a página para o mapa de palavras-chave que podem ser extraídas das palavras-chave da produção, do título da produção, das informações adicionais da produção, do título do projeto ou descrição do projeto. O mapa é apresentado em uma rede onde os nós representam palavras-chave e as arestas conectam duas palavras-chave de uma mesma produção ou projeto. Ao clicar no nó, a palavra-chave é transferida para o centro do grafo. Ao clicar no peso da aresta, a página com a lista de produções que contém as palavras-chave é apresentada.

3.5 e-Lattes

O e-Lattes é uma evolução do scriptLattes que extrai e apresenta indicadores da produção acadêmica brasileira a partir dos currículos Lattes (E-LATTES, 2019). Uma ferramenta de extração, codificada na linguagem de programação R, obtém e processa

Figura 7 – Mapa de palavras-chave da Stela Experta.



Fonte: StelaTek (2018)

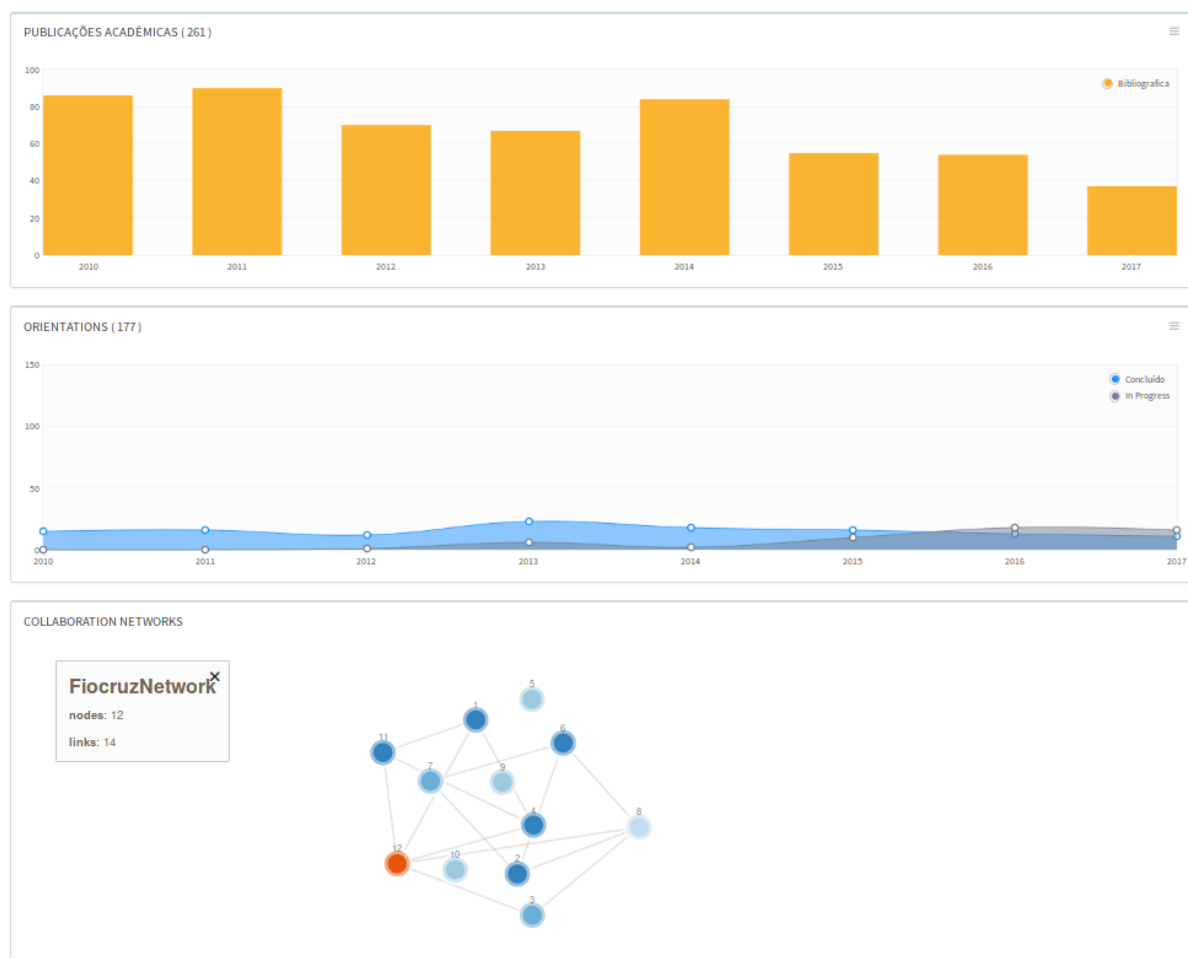
as informações de um grupo de currículos em XML (SAMPAIO; BATISTA JUNIOR; MENA-CHALCO, 2018). Os resultados dos grupos processados são concentrados em uma única aplicação *web* e seus indicadores são apresentados de forma fácil e aberta¹.

A Figura 8 apresenta os indicadores bibliométricos de produção e coautoria de um dos grupos analisados pelo e-Lattes: um gráfico para as produções bibliográficas, outro para as orientações e um grafo para a rede de coautoria. Orientações concluídas são comparadas com orientações em andamento. Na rede de coautoria, ao clicar no nó e na aresta, é possível visualizar o nome do autor, o total de ligações com outros nós e o peso da ligação.

Outros 179 grupos são apresentados, porém, cada um isolado um do outro e associado a um conjunto próprio de currículos. Características da produção, como período, tipo de publicação ou tipo de orientação, podem ter sido selecionadas para definir quais produções serão incluídas na contagem dos indicadores.

¹ Foi necessário fazer apenas um cadastro para acessar os indicadores.

Figura 8 – Indicadores bibliométricos do eLattes.



Fonte: E-LATTES (2019)

3.6 Análise Comparativa

A Tabela 2 apresenta as diferenças entre os sistemas apresentados em relação a solução proposta, comparando as seguintes funcionalidades: (1) indicadores que agrupam dados bibliográficos de um grupo inteiro (por exemplo: uma instituição), (2) indicadores individuais de cada pesquisador do grupo analisado, (3) indicadores de um grupo customizado de pesquisadores do grupo analisado, (4) indicadores que apresentam a relação entre a quantidade total de produções, (5) indicadores visuais que permitem analisar a relação entre os pesquisadores, (6) dados do gráfico de produção podem ser alterados usando a sua interface, (7) dados do grafo de coautoria podem ser alterados usando a sua interface e (8) permite o acesso público aos indicadores,

O indicador bibliométrico de produção está presente em todos os trabalhos relacionados, sendo que o Somos e o Lattes apresentam apenas esse tipo de indicador. O scriptLattes e o e-Lattes incluem adicionalmente uma rede de colaboração. Os indicado-

res são geralmente gerados para um grupo, como uma instituição, com exceção do Lattes, que apresenta indicadores apenas de um currículo. Além disso, a maioria desses sistemas disponibilizam os indicadores publicamente ou através de um *login*.

A solução proposta contempla todas as funcionalidades apresentadas na tabela. Indicadores de produção e redes de coautorias são apresentados e podem ser gerados para um grupo específico de membros. A Stela Experta é a única que inclui essa funcionalidade, mas somente para indicadores de produção. Além disso, somente a solução proposta apresenta redes de coautorias interativas.

Tabela 2 – Análise comparativa dos sistemas bibliométricos.

#	Funcionalidade	scriptLattes	Somos	Stela	e-Lattes	Lattes	Proposta
1	Indicadores de todo o grupo analisado	✓	✓	✓	✓		✓
2	Indicadores do membro do grupo analisado		✓	✓			✓
3	Indicadores de grupo customizado de membros			✓			✓
4	Gráficos quantitativos de produção	✓	✓	✓	✓	✓	✓
5	Grafos de coautoria	✓			✓		✓
6	Gráficos de produção interativos		✓	✓		✓	✓
7	Grafos de coautoria interativos						✓
8	Livre acesso	✓	✓		✓	✓	✓

Fonte: elaborado pelo autor.

4 SOLUÇÃO PROPOSTA

Este capítulo descreve as características e o funcionamento do sistema proposto. A seção 4.1 apresenta as funcionalidades, a seção 4.2 explica como a arquitetura está organizada, a seção 4.3 explica o funcionamento da extração de dados e as seções 4.4 e 4.5 apresentam detalhes sobre a aplicação *web* e o BD.

4.1 O Sistema Proposto

A ideia central do sistema é apresentar indicadores bibliométricos de produção e colaboração, de forma simples e intuitiva, gerados a partir dos currículos Lattes. Para extração dos dados, uma adaptação do scriptLattes é responsável por obter os dados de um grupo específico de pesquisadores, geralmente membros de uma instituição. Uma aplicação *web* serve de interface para o usuário final, que poderá visualizar graficamente e interativamente os dados de instituições e pesquisadores.

O foco da aplicação *web* é apresentar diferentes indicadores de produção e coautoria conforme a interação do usuário, com base nas seguintes características:

- Apresentar gráficos para os indicadores de produção. O total de produções pode estar associado a uma propriedade da produção como ano, tipo ou característica do seu autor.
- Apresentar grafos para os indicadores de coautoria. Arestas podem ser ponderadas para indicar a quantidade de produções em comum.
- Disponibilizar opções de filtros e navegação que podem ser combinados para que o usuário possa especificar indicadores de um pesquisador, um grupo de pesquisadores ou um tipo de produção, por exemplo.

4.2 A Arquitetura do Sistema

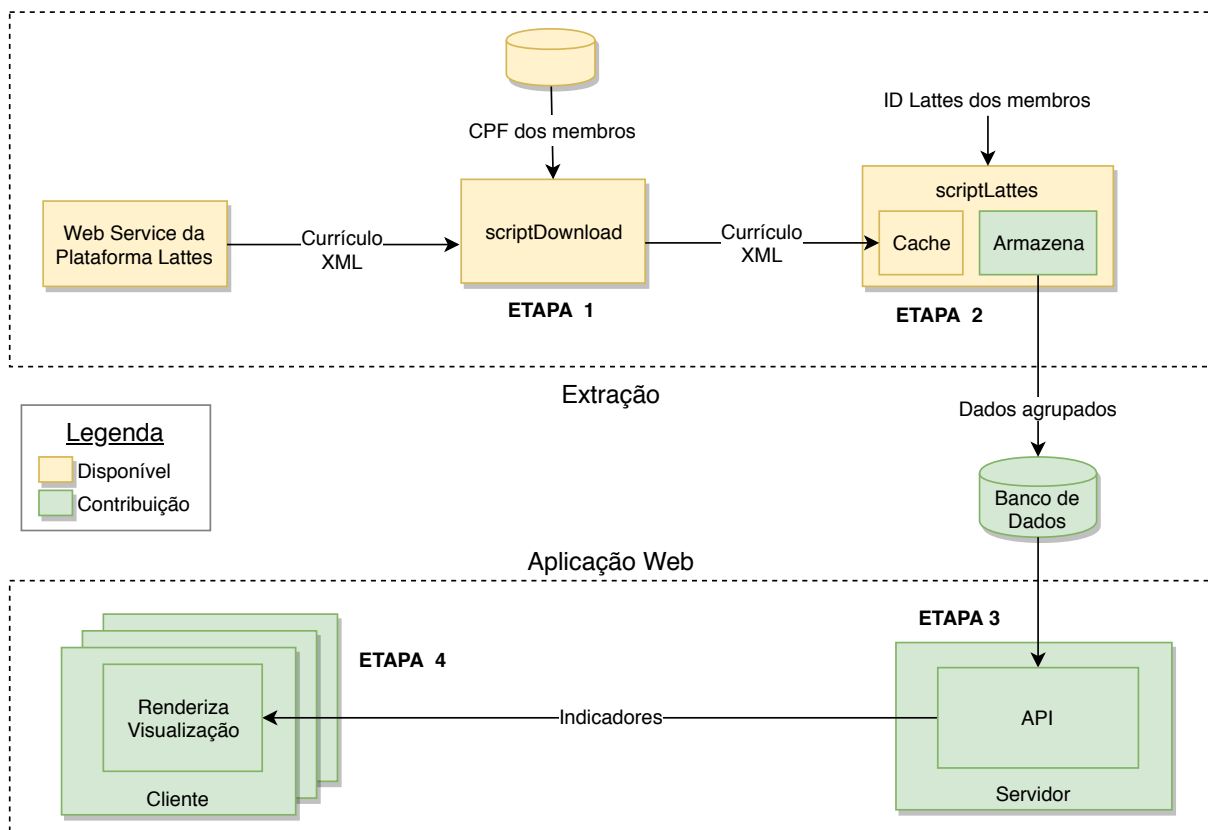
A Figura 9 ilustra o fluxo de dados e a arquitetura do sistema, que está dividido em duas partes: a extração dos dados e a aplicação *web*. Cada uma dessas partes são compostas de programas. Alguns programas estão disponíveis para uso (em amarelo), enquanto outros são contribuições do trabalho (em verde).

Um conjunto de currículos são processados por esses programas em uma sequência de etapas:

1. O scriptDownload é responsável por obter os Cadastros de Pessoas Físicas (CPF's) dos pesquisadores da instituição, descobrir o ID Lattes a partir do CPF e fazer o *download* do currículo em XML a partir do ID Lattes.
2. O scriptLattes recebe os currículos e extrai seus dados. Uma nova funcionalidade “Armazena” envia esses dados para um BD.

3. O servidor acessa os dados do BD e gera indicadores. Uma Application Programming Interface (API) oferece meios de acesso aos indicadores para qualquer cliente acessar.
4. O cliente seleciona os indicadores e gera uma requisição para o servidor. O servidor retorna os indicadores solicitados e o cliente renderiza visualmente para o usuário.

Figura 9 – Arquitetura e fluxo de dados do sistema.



Fonte: elaborado pelo autor.

4.3 Extração de Dados

Os currículos são obtidos do *web service* da PL no formato XML. Para adquirir o currículo de um determinado membro, é necessário seu ID Lattes, o qual pode ser obtido através do CPF com o programa `scriptDownload`.

O `scriptLattes` é utilizado para fazer um *parsing* dos currículos em formato XML, agrupar as produções, remover duplicações e identificar coautorias. A função seguinte do `scriptLattes` é gerar indicadores com os dados extraídos em páginas HTML. Portanto, foi proposto uma nova funcionalidade para mapear os dados para um BD, possibilitando o desenvolvimento de aplicações *web* mais dinâmicas.

Tomando como base os pesquisadores de uma instituição e considerando que há colaboração em pesquisa, há o problema de produções duplicadas. Por exemplo, quando dois pesquisadores são coautores de um artigo, o registro do artigo é encontrado nos seus respectivos arquivos XML. É importante que essa duplicação seja tratada apropriadamente antes dessa informação ser inserida no BD, caso contrário, ao solicitar um indicador que depende da contagem de muitas produções, todas elas precisariam ser comparadas a procura de duplicações. Portanto, esse custo adicional afetaria na interatividade da aplicação *web*.

4.4 Aplicação Web

A aplicação *web* está dividida entre o *frontend* (executado no cliente) e o *backend* (executado no servidor). O *backend* é composto de uma API, que disponibiliza indicadores gerados a partir dos dados presentes no BD. Já o *frontend*, recebe os indicadores da API através de requisições Hypertext Transfer Protocol (HTTP), como também renderiza os componentes da interface, incluindo os indicadores.

Para impedir que a interatividade com a interface do usuário seja prejudicada por recarregamento de páginas, a aplicação *web* segue o modelo Single-Page Application (SPA). Nesse modelo, a primeira requisição ao servidor retorna uma única página com a interface do usuário e a lógica da aplicação. Após isso, a página retornada faz requisições apenas à API que retorna indicadores ou outra informação do BD. Quando os dados são retornados, somente uma parte da página é atualizada, permitindo assim, a criação de uma interface dinâmica e com indicadores mais interativos.

4.5 Armazenamento de Dados

Com base na arquitetura proposta, o BD pode ser modelado em uma estrutura menos normalizada e otimizada para leituras. Em uma estrutura como essa, a informação é distribuída em menos tabelas físicas e redundâncias (dados repetidos em múltiplos registros) são permitidas. Como consequência, operações como atualizar ou remover precisariam ser repetidas onde o dado está duplicado, tornando o BD vulnerável a inconsistências caso não sejam executadas apropriadamente. Entretanto, o BD é criado na etapa de extração e a aplicação web faz somente leituras. Portanto, é possível criar uma estrutura mais focada em facilitar buscas, atingindo menos tabelas para gerar o indicador.

5 IMPLEMENTAÇÃO

Este capítulo apresenta a implementação do sistema proposto. A seção 5.1 apresenta linguagens de programação, bibliotecas e sistemas usados em todas as camadas da implementação. A seção 5.2 apresenta a estrutura do BD. A seção 5.3 apresenta a ferramenta de extração, adaptada do scriptLattes. Por fim, as seções 5.4, 5.5, 5.6 e 5.7 apresentam a implementação das principais funcionalidades do Openlattes, a aplicação *web* desenvolvida para a solução proposta.

5.1 Principais Tecnologias

Para auxiliar na implementação da arquitetura apresentada na seção 4.2, um conjunto de tecnologias foram usadas. O BD escolhido foi o MongoDB (MONGODB INC., 2019a), um Sistema Gerenciador de Banco de Dados (SGBD) baseado em documentos, *opensource*, flexível e com operações avançadas de busca (MONGODB INC., 2019c). A nova funcionalidade do scriptLattes foi implementada diretamente no seu código-fonte com a linguagem de programação Python junto com a biblioteca PyMongo (MONGODB INC., 2019b), a forma recomendada para interagir com o MongoDB usando Python. A aplicação *web* foi implementada em JavaScript (JS) com o auxílio de várias bibliotecas como o Apollo GraphQL (METEOR DEVELOPMENT GROUP, 2019), React (FACEBOOK OPEN SOURCE, 2018) e Semiotic (MEEKS, 2018).

O MongoDB (MONGODB INC., 2019a) permite simplificar a geração de indicadores e a implementação da arquitetura proposta. Os dados são armazenados em um formato semelhante ao JavaScript Object Notation (JSON), que por sua vez, é semelhante ao formato de objetos de um programa. Essa característica facilita na implementação do mapeamento dos dados nas duas interfaces com o BD que precisam ser implementadas (o *script* de extração e a aplicação *web*). O formato também permite uma estrutura flexível para a geração de indicadores. O MongoDB ainda inclui recursos de busca suficientes para gerar os indicadores necessários. Por exemplo, o método *aggregate* permite combinar uma série de operações, como filtrar, agrupar, ordenar, etc.

A aplicação *web* foi implementada com a linguagem de programação JS e as linguagens HTML e Cascading Style Sheets (CSS). Essas três linguagens são interpretadas nativamente por navegadores e, portanto, foram escolhidas para a implementação do *frontend*. JS também é usado no *backend* com a tecnologia Node.js (NODE.JS FOUNDATION, 2019). Dessa forma, as duas camadas da aplicação foram desenvolvidas com a mesma linguagem de programação.

O React (FACEBOOK OPEN SOURCE, 2018) foi a biblioteca usada na implementação do *frontend* em JS. Ela facilita a criação de aplicações *web* dinâmicas, interativas e que seguem o modelo SPA. Na arquitetura proposta, componentes da interface do usuário precisam ser atualizados conforme a interação do usuário, principalmente quando dados são retornados da API. Com React, os componentes são implementados de forma

declarativa e são automaticamente renderizados quando necessário.

A Semiotic (MEEKS, 2018) é uma biblioteca de visualização de dados compatível com React e com suporte para gráficos e grafos. Ela permite criar visualizações de vários tipos e de forma abstraída, ou seja, basta escolher a visualização e customizá-la. Essa característica agiliza a criação, não só dos indicadores propostos, mas também de novos indicadores, caso a aplicação seja expandida futuramente.

As bibliotecas Material-UI (CALL-EM-ALL, 2018), Apollo GraphQL (METEOR DEVELOPMENT GROUP, 2019) e Mongoose também foram utilizadas. Material-UI disponibiliza uma série de componentes de interface customizáveis e compatíveis com React. Apollo GraphQL foi utilizado na implementação da API, auxiliando na criação da API e na busca pelos dados. E, por fim, Mongoose foi utilizado para acessar o MongoDB pelo Node.js.

5.2 Estrutura do Banco de Dados

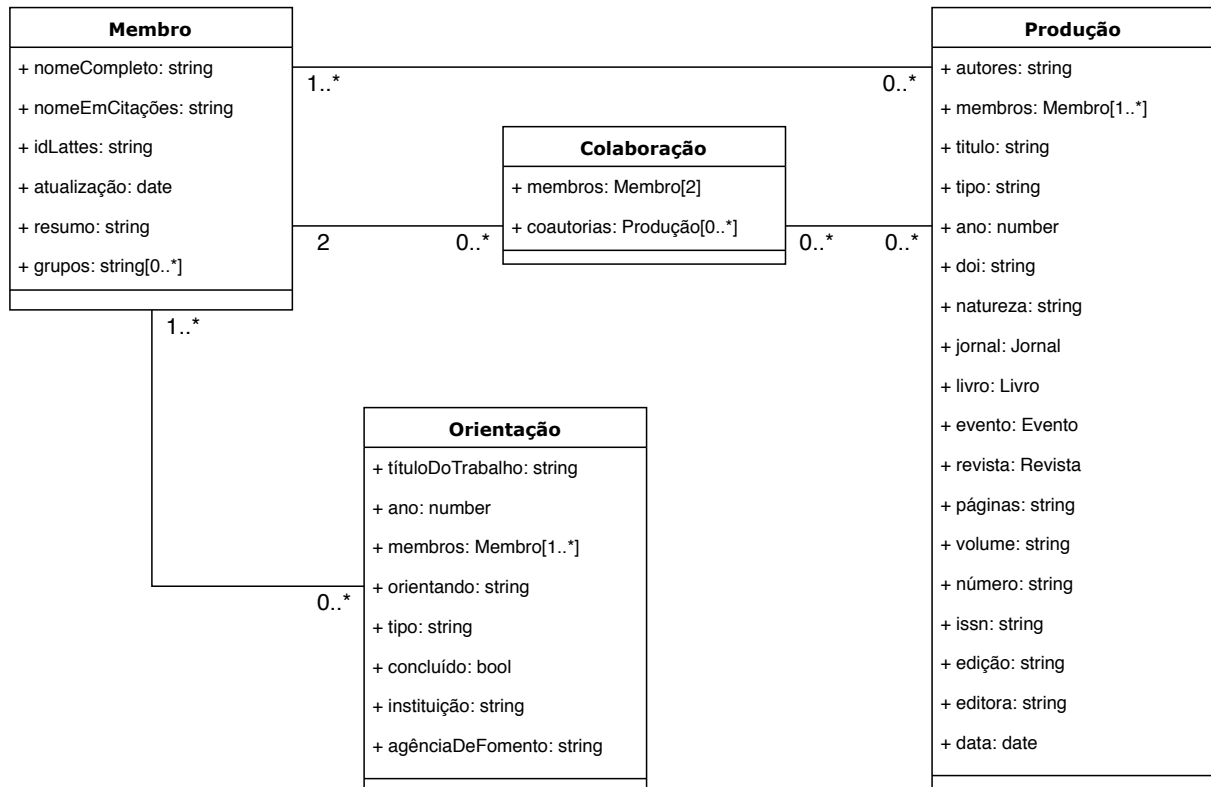
O esquema do BD está representado como um diagrama de classe conforme a Figura 10. Cada classe representa uma coleção. Uma coleção é um conjunto de documentos, ou seja, um conjunto de registros. Cada propriedade pública representa um campo opcional do documento. Dessa forma, o campo não está necessariamente presente em todos os documentos da coleção. Adicionalmente, cada coleção recebe automaticamente um campo ID que é usado para distinguir o registro e fazer as associações entre as coleções. As associações e a cardinalidade representam a relação entre duas coleções, onde uma armazena um ou mais IDs de outra coleção.

Quatro coleções armazenam todos os dados do BD. A coleção *Membro* armazena os dados de cada pesquisador. As coleções *Produção* e *Orientação* armazenam respectivamente os dados das produções bibliográficas e das orientações concluídas. Futuramente, a coleção *Produção* pode ser estendida para armazenar produções técnicas e artísticas inserindo um campo *categoria*, por exemplo. Da mesma forma, a coleção *Orientação* pode ser estendida para armazenar orientações em andamento. A coleção *Colaboração* armazena uma lista de arestas para a construção do grafo de coautoria. O campo *membros* armazena os dois nós e o campos *produções* armazena as produções bibliográficas em comum. A quantidade de produções define o peso da aresta.

As coleções *Produção* e *Orientação* estão relacionadas com a coleção *Membro* através do ID do membro. A coleção *Produção* contém o campo *membros* que armazena a lista de IDs dos autores que registram aquela produção em seus currículos. A coleção *Orientação* também contém o campo *membros*, mas representa o ID dos orientadores do trabalho. O autor do trabalho está no campo *orientando*. A lista vai incluir um ou muitos IDs que estão registrados na coleção *Membro*, enquanto que o registro do membro está presente em nenhuma, uma ou mais produções ou orientações.

A coleção *Colaboração* está associada às coleções *Membro* e *Produção*. Essa coleção

Figura 10 – Diagrama de classe representando o esquema do Banco de Dados.



Fonte: elaborado pelo autor.

armazena os IDs dos membros coautores referente a coleção *Membro* e os ID das produções bibliográficas referente a coleção *Produção*. São exatamente dois IDs dos membros e pelo menos um ID das produções bibliográficas. Um registro de uma produção bibliográfica ou de um membro está presente em nenhuma ou muitas colaborações.

5.3 Extração de Dados

Uma ferramenta foi criada a partir do scriptLattes com a nova funcionalidade para armazenar os dados extraídos para o MongoDB. Algumas funcionalidades foram mantidas como o *parser* de currículos e o processamento que agrupa produções e identifica coautorias.

Os dados de membros, produções e orientações estão armazenados em listas de objetos que são mapeados para o BD seguindo a estrutura apresentada na seção 5.2. Primeiro, os membros são inseridos, depois as produções e orientações. Entre as informações das produções e orientações, a lista de IDs dos membros coautores são inseridas no campo *membros*. Após armazenar cada produção, o campo *membros* é aproveitado para registrar as colaborações. O método *calcularCombinacoes* do scriptLattes transforma a lista de coautores, presente no campo *membros* em pares. O par de IDs dos membros e o ID da

produção são inseridos nos campos *membros* e *coautorias* de uma colaboração. Caso o par de membros já esteja registrado nas colaborações, o registro é atualizado incrementando o campo *coautorias* com ID da produção.

5.4 API

O *frontend* obtém os indicadores e demais dados da interface do usuário através de requisições à API. A Tabela 3 apresenta as 6 requisições que foram necessárias: *membros*, *produções*, *orientações*, *indicadorDeProdução*, *nós* e *arestas*. Cada uma recebe parâmetros de entrada que podem ser combinados para fazer uma busca mais específica. Ao receber uma requisição, os métodos *find* ou *aggregate* do MongoDB são chamados para obter os dados que são retornados em JSON.

Tabela 3 – Resquisições e opções da API.

Requisição	Parâmetro	Valor
produções	membros	Lista com os IDs dos membros
	tipos	Lista com os tipos de produção
	grupo	Nome de um grupo
	ano	Um ano específico
	nomeDoMembro	Nome de um membro
orientações	membros	Lista com os IDs dos membros
	tipos	Lista com os tipos de produção
	grupo	Nome de um grupo
	ano	Um ano específico
	nomeDoMembro	Nome de um membro
membros	membros	Lista com os IDs dos membros
	lattesIDs	Lista os Lattes IDs dos membros
indicadorDeProdução	coleção	“produções” ou “orientações”
	por	“ano” ou “membro”
	membros	Lista com os IDs dos membros
	grupo	Nome do grupo
	nós	Lista com os IDs dos membros
arestas	membros	Lista com os IDs dos membros

Fonte: elaborado pelo autor.

As requisições *produções*, *orientações* e *membros* retornam, respectivamente, os registros das produções, orientações e membros. O parâmetro *membros* permite especificar somente os membros de interesse por uma lista de IDs e, quando não indicado, todos os membros são retornados. O parâmetro *lattesIDs* também permite especificar uma lista de membros, mas usando os IDs Lattes. Nas requisições *produções* e *orientações*, também é possível indicar *tipos*, *grupo*, *ano* e *nomeDoMembro* para filtrar os resultados com essas informações.

A requisição *indicadorDeProdução* retorna os indicadores de produção e os parâmetros são utilizados para definir o seu tipo. O parâmetro *coleção* define se a contagem

vai ser de produções bibliográficas ou orientações concluídas. O parâmetro *por* define se a contagem vai ser por ano ou por membro. O parâmetro *membros* gera o indicador para um determinado grupo de membros ou apenas um membro. Por fim, o parâmetro *grupo* gera o indicador apenas para os membros do grupo especificado.

A Figura 11 apresenta um resumo da resposta do *indicadorDeProdução*. Nesse exemplo, os valores do parâmetro *por* é “ano” e do parâmetro *coleção* é “produções”. Para possibilitar a criação do gráfico em pilha, a estrutura retorna o valor total de produções de cada ano separados por tipo de produção. Se o parâmetro *por* for “membro”, então o indicador será de apenas os 30 membros com o maior número de produções. O limite de 30 membros permite apresentar um número razoável de barras no gráfico sem deixá-lo muito extenso.

Figura 11 – Formato do indicador de produção retornado pela API.

```
{
  "indicadorDeProducao": [
    { "ano": 2015, "tipo": "Resumo em Congresso", "total": 30 },
    { "ano": 2014, "tipo": "Resumo em Congresso", "total": 28 },
    ...
    { "ano": 2015, "tipo": "Capitulo de Livro", "total": 16 },
    { "ano": 2014, "tipo": "Capitulo de Livro", "total": 11 },
    ...
  ]
}
```

Fonte: elaborado pelo autor.

As requisições *nós* e *arestas* são usadas em conjunto para gerar o grafo de coautoria. A requisição *nós* retorna informações dos membros que representam os nós. A requisição *arestas* retorna os dados da coleção *Colaboração*. As duas possuem o parâmetro *membros* que permite especificar uma lista de membros. Ao fazer isso, o grafo será gerado somente para os membros especificados.

A Figura 12 apresenta um exemplo da estrutura do grafo de coautoria. A lista de arestas contém dados suficientes para gerar o grafo: o ID dos dois membros coautores e o peso que é equivalente ao total de coautorias em comum. As informações do nó são opcionais e podem ser usadas para decorar a visualização como, por exemplo, colorir ou incluir o nome do membro no nó. Além do nome e grupos, outras informações do membro podem ser incluídas.

A Tabela 4 apresenta duas requisições para gerenciar os grupos dos pesquisadores. Os grupos estão armazenados no campo *grupos* da coleção *Membro* do BD. A requisição *adicionaGrupo* armazena um novo grupo através de um nome e uma lista de membros. O nome do grupo será armazenado no campo *grupos* de todos os membros especificados na lista de membros. A requisição *removeGrupo* remove um grupo existente através de seu

Figura 12 – Formato do grafo retornado pela API.

```

{
  "nos": [
    { "id": 1, "nome": "Marcos", "grupos": ["A"] },
    { "id": 2, "nome": "Maria", "grupos": ["A", "B"] },
    { "id": 3, "nome": "Bruno", "grupos": ["B"] }
    ...
  ],
  "arestas": [
    { "id1": 1, "id2": 3, "peso": 5 },
    { "id1": 2, "id2": 3, "peso": 8 }
    ...
  ]
}

```

Fonte: elaborado pelo autor.

nome. O grupo será removido do campo *grupos* de todos os membros que fizerem parte dele.

Tabela 4 – Requisições da API para alterar grupos.

Requisição	Parâmetro	Valor
adicionaGrupo	membros grupo	Lista com os IDs dos membros Nome de um grupo
removeGrupo	grupo	Nome de um grupo

Fonte: elaborado pelo autor.

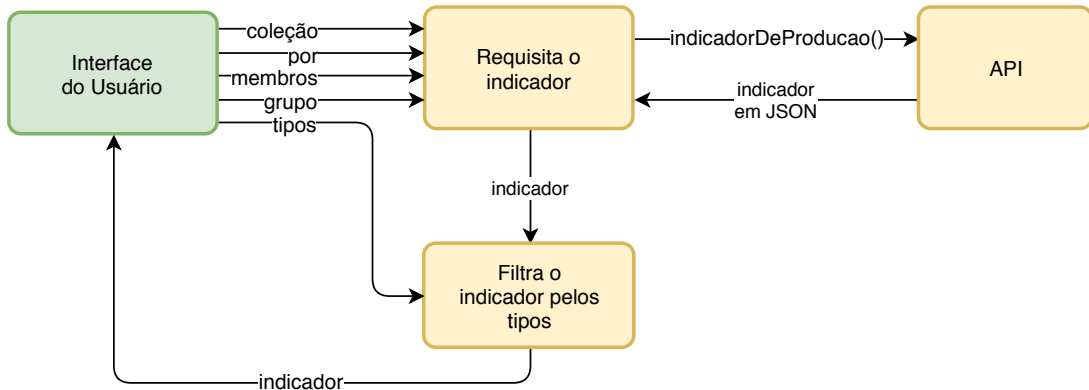
5.5 Indicador de Produção

A Figura 13 apresenta como o indicador de produção é gerado na interface do usuário. Os parâmetros *coleção*, *membros*, *grupo* e *por* são definidos pelo usuário interagindo com a interface. A requisição *indicadorDeProdução* é disparada com esses parâmetros e o indicador é retornado em JSON pela API. A opção *tipos* contém uma lista com os tipos de produção, também definida como entrada pela interface. O indicador é filtrado mantendo somente os tipos de produção contidos nessa lista.

5.6 Grafo de Coautoria

A Figura 14 apresenta como a rede de coautoria é gerada na interface do usuário. O parâmetro *membros* é definido como entrada na interface e usado para disparar as requisições *nós* e *arestas*.

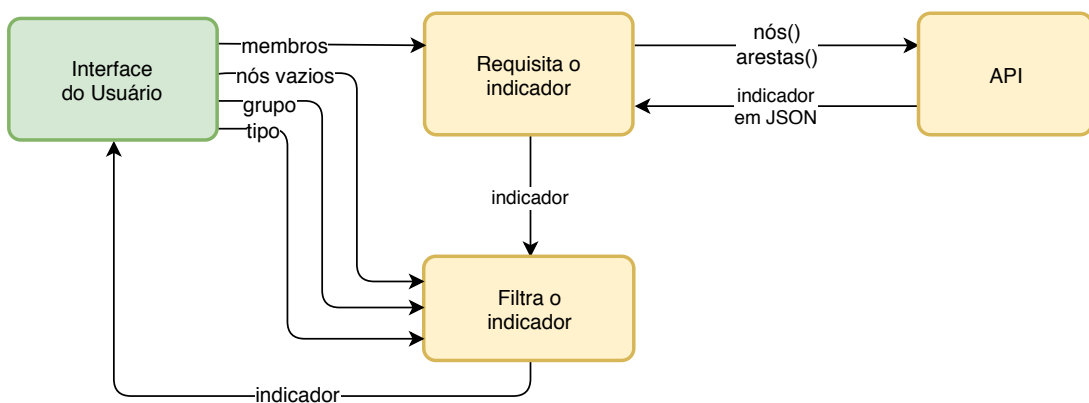
Figura 13 – Como o indicador de produção é gerado.



Fonte: elaborado pelo autor.

As opções *nós vazios*, *grupo* e *tipo*, também definidas como entrada pela interface do usuário, são usadas para filtrar o grafo após ele ser retornado da API. A opção *nós vazios* pode ter o valor “verdadeiro” ou “falso”. Se for “verdadeiro”, o grafo vai incluir os nós que não possuem ligações. A opção *grupo* é usada para filtrar o grafo para incluir somente os membros que pertencem ao grupo especificado. A opção *tipo* é usada para filtrar o grafo para incluir somente o tipo de produção especificado.

Figura 14 – Como o grafo de coautoria é gerado.



Fonte: elaborado pelo autor.

5.7 Armazenamento de Grupos no Navegador

A tecnologia IndexedDB foi utilizada para armazenar e gerenciar grupos locais. Essa tecnologia permite o armazenamento de dados estruturados no próprio navegador. A Figura 15 ilustra o formato de um grupo salvo no navegador. Cada grupo é composto

de um ID, um nome, e a lista de IDs Lattes dos membros desse grupo. Para gerar um indicador para o grupo armazenado, basta utilizar a lista do campo *membros* ao requisitar um indicador de produção ou uma rede de coautoria.

Figura 15 – Formato de um grupo armazenado no navegador.

```
{
  id: 1,
  nome: "Grupo A",
  membros: ["11111", "22222", "33333"]
}
```

Fonte: elaborado pelo autor.

6 RESULTADOS

Este capítulo apresenta os resultados obtidos utilizando os currículos dos docentes da UNIPAMPA. A seção 6.1 apresenta a obtenção e extração dos dados até a criação do BD. As seções 6.2, 6.3, 6.4, 6.5 e 6.6 apresentam o Openlattes utilizando o BD gerado na etapa de extração. Por fim, a seção 6.7 apresenta a avaliação da solução proposta através de um teste com o usuário.

6.1 Extração e Geração do Banco de Dados

O scriptDownload foi executado para obter os currículos em XML dos professores. Os CPFs de 743 docentes foram usados para adquirir os seus IDs Lattes. Os CPFs são adquiridos do diretório de informações institucionais da UNIPAMPA. O *script* acessa a *web service* da PL para obter a lista de IDs Lattes e, a partir dessa lista, obtém os currículos.

Na sequência, os currículos foram usados como entrada na execução do *script* de extração. A lista de IDs Lattes, produzida na etapa anterior, foi reutilizada no arquivo de entrada do scriptLattes. Nesse arquivo, foi indicado na coluna do rótulo, um campus arbitrário para cada professor para que possa ser usado como um grupo de classificação. Finalmente, o scriptLattes foi executado e 50464 registros foram armazenados no BD: 743 na coleção *Membro*, 48277 na coleção *Produção*, 922 na *Colaboração* e 522 na coleção *Orientação*.

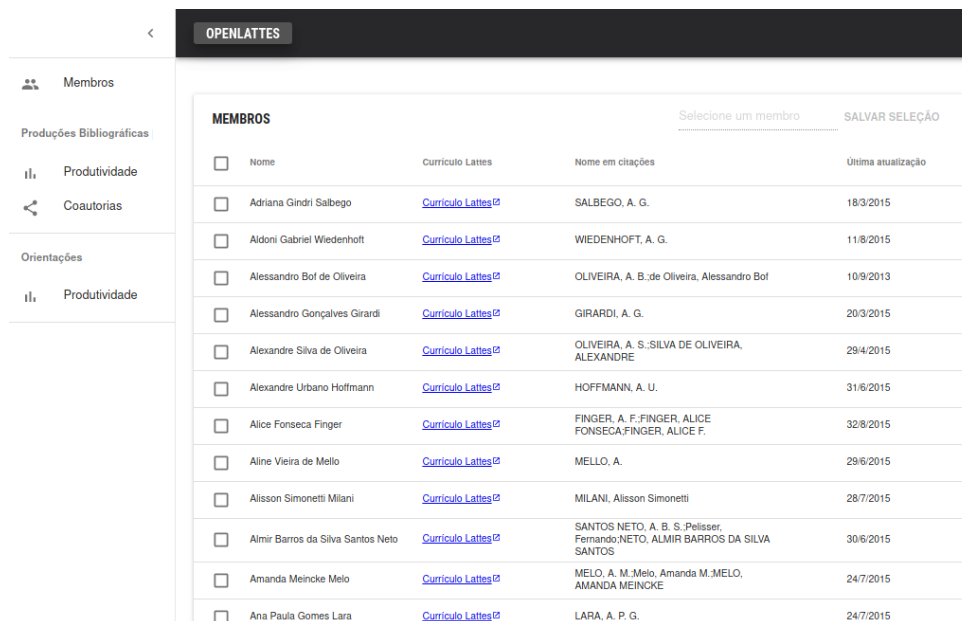
6.2 Página Inicial

A Figura 16 apresenta a interface do usuário ao acessar o Openlattes. A página com a lista dos docentes é apresentada junto com um menu principal de navegação, situado no lado esquerdo da página. O menu está separado em categorias que determinam a informação que o indicador irá contabilizar: produções bibliográficas ou orientações concluídas. Dentro de cada categoria, o menu *Produtividade* apresenta os indicadores de produção e o menu *Coautorias*, disponível apenas para as produções bibliográficas, apresenta a página com as redes de coautorias.

6.3 Lista e Seleção de Professores

A página principal apresenta a lista de professores que também pode ser acessada pelo menu *Membros*. Uma tabela apresenta todos os professores da UNIPAMPA com o nome completo, o *link* para o currículo Lattes, os nomes em citações e a data da última atualização do currículo. A tabela permite que um conjunto de professores sejam selecionados para serem usados como filtros nos indicadores. Quando há pelo menos um

Figura 16 – Interface do usuário da página inicial do Openlattes.

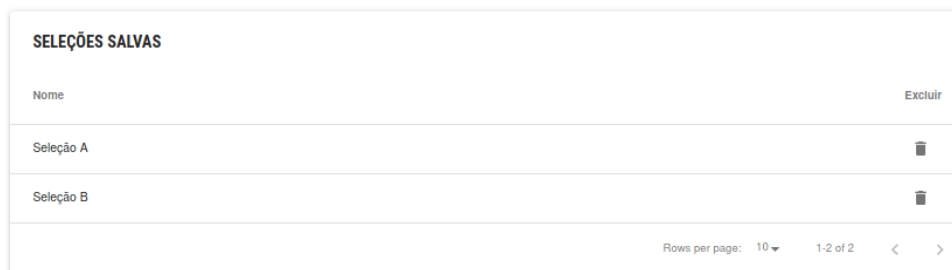


The screenshot shows the Openlattes user interface. On the left is a navigation menu with categories like 'Membros', 'Produções Bibliográficas', 'Produtividade', 'Coautorias', and 'Orientações'. The main area displays a table titled 'MEMBROS' with columns for selection, name, curriculum link, citation name, and last update date. The table lists 14 members with their respective details.

<input type="checkbox"/>	Nome	Currículo Lattes	Nome em citações	Última atualização
<input type="checkbox"/>	Adriana Gindri Salbego	Currículo Lattes	SALBEGO, A. G.	18/3/2015
<input type="checkbox"/>	Aldoni Gabriel Wiedenhof	Currículo Lattes	WIEDENHOFT, A. G.	11/8/2015
<input type="checkbox"/>	Alessandro Bof de Oliveira	Currículo Lattes	OLIVEIRA, A. B. de Oliveira, Alessandro Bof	10/9/2013
<input type="checkbox"/>	Alessandro Gonçalves Girardi	Currículo Lattes	GIRARDI, A. G.	20/3/2015
<input type="checkbox"/>	Alexandre Silva de Oliveira	Currículo Lattes	OLIVEIRA, A. S. SILVA DE OLIVEIRA, ALEXANDRE	29/4/2015
<input type="checkbox"/>	Alexandre Urbano Hoffmann	Currículo Lattes	HOFFMANN, A. U.	31/6/2015
<input type="checkbox"/>	Alice Fonseca Finger	Currículo Lattes	FINGER, A. F.; FINGER, ALICE FONSECA; FINGER, ALICE F.	32/8/2015
<input type="checkbox"/>	Aline Vieira de Mello	Currículo Lattes	MELLO, A.	29/6/2015
<input type="checkbox"/>	Alisson Simonetti Milani	Currículo Lattes	MILANI, Alisson Simonetti	28/7/2015
<input type="checkbox"/>	Almir Barros da Silva Santos Neto	Currículo Lattes	SANTOS NETO, A. B. S. Pelasser, Fernando; NETO, ALMIR BARROS DA SILVA SANTOS	30/6/2015
<input type="checkbox"/>	Amanda Meincke Melo	Currículo Lattes	MELO, A. M. Melo, Amanda M.; MELO, AMANDA MEINCKE	24/7/2015
<input type="checkbox"/>	Ana Paula Gomes Lara	Currículo Lattes	LARA, A. P. G.	24/7/2015

Fonte: elaborado pelo autor.

Figura 17 – Tabela para visualizar as seleção salvas.



The screenshot shows a table titled 'SELEÇÕES SALVAS'. It has two columns: 'Nome' and 'Excluir'. The table lists two saved selections, 'Seleção A' and 'Seleção B', each with a trash icon in the 'Excluir' column. At the bottom, there is a pagination control showing 'Rows per page: 10' and '1-2 of 2'.

Nome	Excluir
Seleção A	
Seleção B	

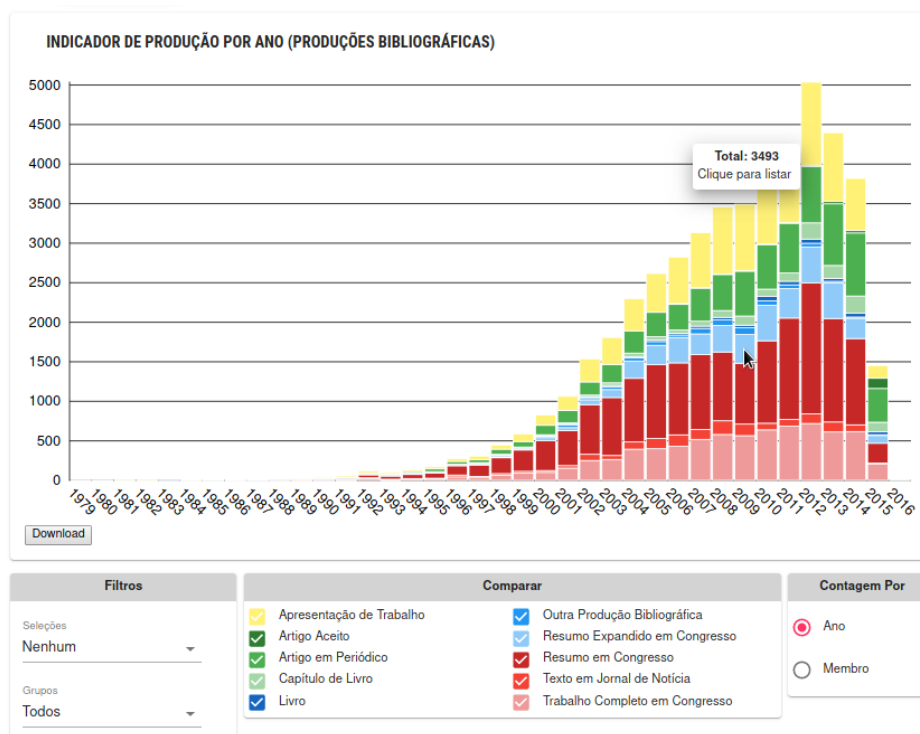
Fonte: elaborado pelo autor.

professor selecionado, o filtro pode ser aplicado ao navegar para um indicador pelo menu principal.

Múltiplas seleções podem ser salvas separadamente. Para isso, é necessário fazer uma seleção de professores e escolher um nome para identificar o grupo. Estão disponíveis no topo da tabela uma caixa de entrada de texto e um botão “Salvar Seleção” para essa função.

As seleções salvas ficam listadas em uma tabela, como na Figura 17, acima da tabela de professores. As seleções podem ser excluídas clicando no ícone da lixeira ao lado do nome da seleção criada. A tabela é atualizada conforme seleções são salvas e removidas. Ao fechar e acessar a aplicação novamente pelo mesmo navegador, as seleções

Figura 18 – Página dos indicadores das produções bibliográficas.



Fonte: elaborado pelo autor.

continuam salvas.

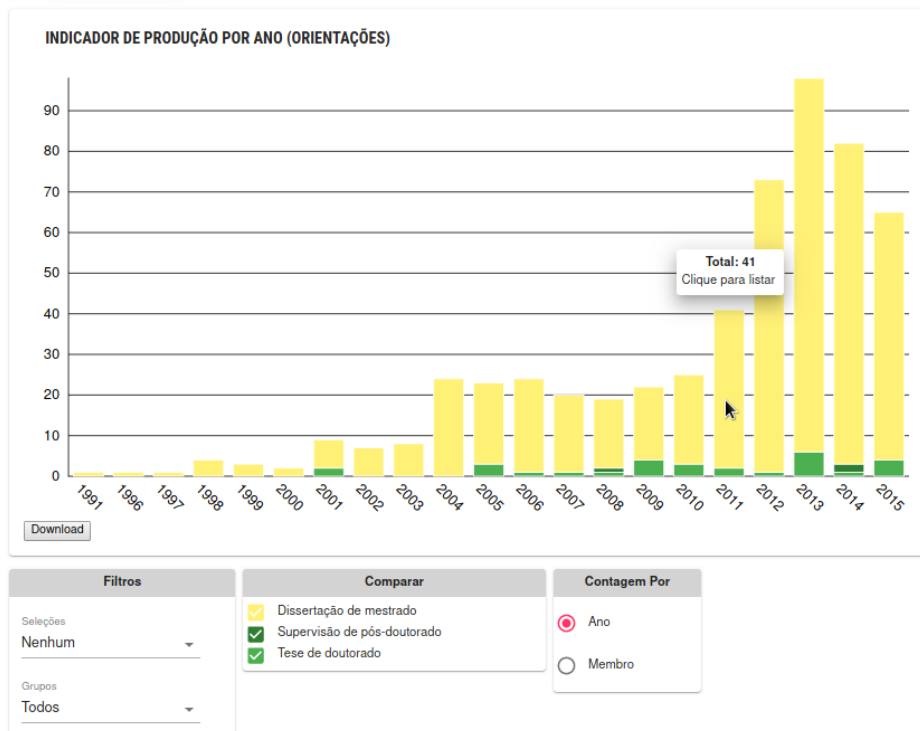
6.4 Indicadores de Produção

A Figura 18 apresenta a interface com os indicadores de produção ao acessar o menu *Produtividade* das produções bibliográficas. Ela é composta de um gráfico com um botão de *Download* e painéis com componentes de seleção. A estrutura é a mesma para os indicadores de produção das orientações, como mostra a Figura 19, mudando apenas os dados que estão sendo apresentados.

Outros detalhes são apresentados conforme a interação com a página. Ao mover o cursor em cima do gráfico, o total de publicações é mostrado. Ao clicar no botão *Download*, uma janela é apresentada para salvar os dados do estado atual do gráfico em um arquivo Comma Separated Values (CSV). Uma tabela como na Figura 20 é apresentada ao clicar no gráfico. São listados o título, autores e tipo de produção das produções contabilizadas pela coluna clicada.

Por padrão, o indicador de produção é classificado por ano e aplicado para todos os professores e tipos de produção, porém, é possível modificar o gráfico utilizando os seletores, como mostra a Figura 21. Os seletores estão separados em três painéis: *Filtros*, *Comparar* e *Contagem Por*.

Figura 19 – Página dois indicadores das orientações.



Fonte: elaborado pelo autor.

O painel *Filtros* permite visualizar a quantidade de produções para um grupo específico de professores, incluindo apenas um professor. Há dois seletores para essa função: *Seleções* e *Grupos*. *Seleções* filtra o gráfico com uma seleção de professores feita na página *Membros*. A seleção pode ser feita pelo nome da seleção salva ou pela última seleção escolhendo “Seleção Atual”. *Grupos* filtra o gráfico para os professores que pertencem a um determinado grupo pré-definido que, neste caso, são os campus definidos na etapa de extração.

O painel *Comparações* permite filtrar e comparar uma combinação de tipos de produções. Cada tipo representa uma cor no gráfico e apenas os tipos selecionados são apresentados. Dessa forma, é possível visualizar o total de produções dos tipos selecionados para cada coluna.

No painel *Contagem Por*, o gráfico pode ser alterado para calcular o total de produções por professor ao invés de ano, como mostra a Figura 22. Nesse caso, o gráfico é apresentado horizontalmente e apenas para os 30 professores com o número maior de produções. Ao atualizar um filtro como *Seleções* ou *Grupos*, os 30 professores com mais produções são recalculados. Por outro lado, ao atualizar os tipos de produções em *Comparações*, os 30 professores não são alterados.

Figura 20 – Tabela listando as produções da coluna 2011.

PRODUÇÕES DE 2011		
Título	Autores	Tipo
A Particle-Filtering Approach for Vehicular Tracking Adaptive to Occlusions	Scharcanski, Jacob; de Oliveira, Alessandro Bof; Cavalcanti, Pablo G.; Yari, Yessenia	Artigo em Periódico
Estimativa temporal de enchentes de rios em tempo linear	FINGER, A. F.; LORETO, A.; BESKOW, S.	Artigo em Periódico
Análises de Complexidade e Numérica do Modelo Hidrológico para Previsões de Inundações em Rios	FINGER, A. F.; LORETO, A. B.	Artigo em Periódico
On the Time Evolution of the Turbulent Kinetic Energy Spectrum for Decaying Turbulence in the Convective Boundary Layer	Goulart, A. G.; BODMANN, B. E. J.; VILHENA, M. T. M. B.; Soares, P. M. M.; MOREIRA, D. M.	Artigo em Periódico
Theoretical study of the decaying convective turbulence in a shear-buoyancy PBL	KIPPER, C. J.; Goulart, A.; DEGRAZIA, G.; Vilhena, M.T.; Soares, P.M.	Artigo em Periódico
On the new parameterisation of the eddy diffusivity for radioactive pollutant dispersion	MOREIRA, Davidson Martins; Soares, Pedro M.M.; Goulart, Antonio Gledson; Vilhena, Marco Tulio	Artigo em Periódico
Modelo de Estrutura de Rede com Reaproveitamento de Computadores	SCHAURICH, A. L.; SCHEPKE, C.	Artigo em Periódico
Agile methods and organizational culture: reflections about cultural levels	Toffo, Cristiano; Wazlawick, Raul Sidnei; Ferreira, Marcelo Giltrana Gomes; Forcellini, Fernando Antonio	Artigo em Periódico
Estudo técnico e econômico de um sistema móvel de extração de óleo essencial de eucalipto	VIVAN, G. A.; BARBOZA, F. S.; DA LUZ, M. L. G.; LUZ, C. A. S. da; RAMIREZ, O. P.; GOMES, M. C.; SOARES, F. C.	Artigo em Periódico
Generalised pulse width modulation approach for DC capacitor voltage balancing in diode-clamped multilevel converters	Grigoletto, F.B.; Pinheiro, H.	Artigo em Periódico

Rows per page: 10 1-10 of 4189 < >

Fonte: elaborado pelo autor.

6.5 Redes de Coautorias

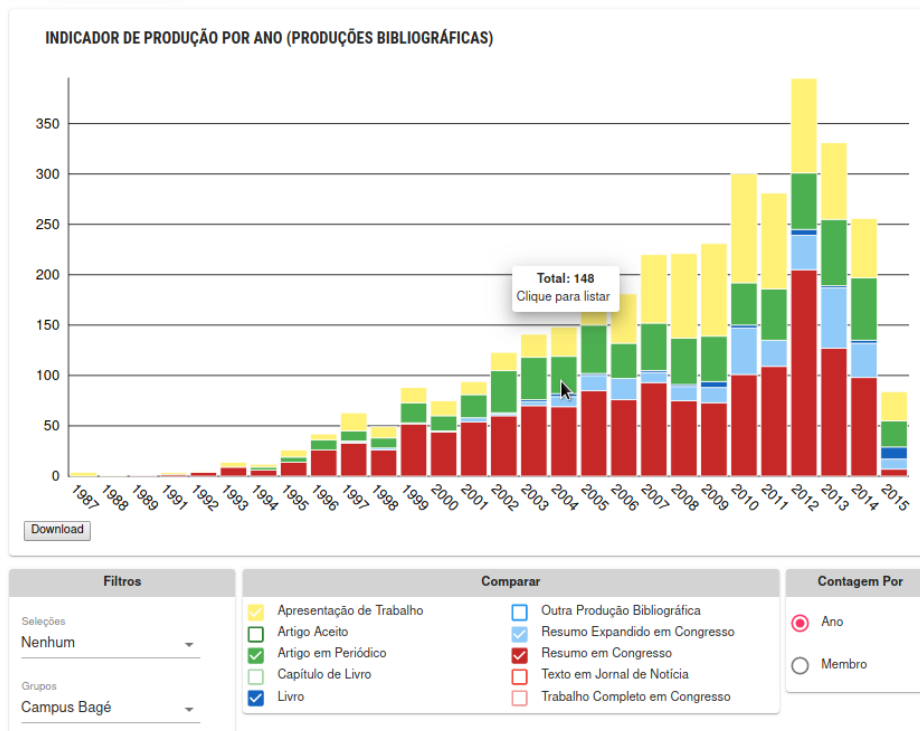
A Figura 23 apresenta a página das redes de coautorias, acessível pelo menu *Coautorias* das produções bibliográficas. Os nós do grafo representam o professor e as arestas representam os trabalhos em comum. O tamanho do nó representa a quantidade de coautorias e a largura da aresta representa a quantidade de trabalhos em comum. O autor pode ser identificado ao passar o cursor em cima do nó: o nome do autor e o total de coautorias é mostrado.

O painel é composto de quatro seletores e a Figura 24 apresenta a interface após eles serem selecionados. Os seletores *Seleções* e *Grupos* funcionam da mesma forma que no indicador de produção. O seletor *Tipo de Produção* limita a rede para apenas um dos tipos de produções. O seletor *Membros sem coautorias* permite visualizar os professores que não possuem coautorias.

6.6 Página de Administração

A Figura 25 apresenta a página de administração. A tabela inferior lista todos os membros e a tabela superior lista todos os grupos desses membros. O funcionamento é semelhante ao da página *Membros*, porém, permite gerenciar grupos ao invés de seleções salvas. Essa interface está separada do resto da aplicação e funciona como uma página de configurações para ser usada por administradores.

Figura 21 – Página dos indicadores das produções bibliográficas com filtros aplicados.



Fonte: elaborado pelo autor.

6.7 Avaliação da Solução Proposta

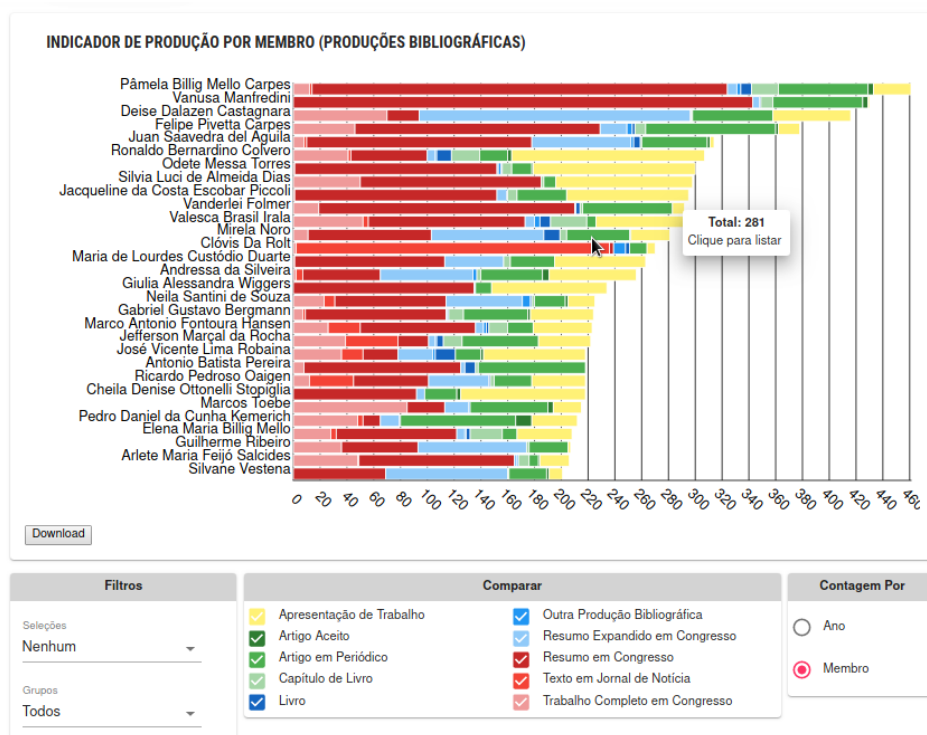
Com o objetivo de avaliar a solução proposta, a aplicação desenvolvida foi testada diretamente com o usuário. Foi utilizado a versão curta do User Experience Questionnaire (UEQ)¹ em combinação com um teste de usabilidade (LAUGWITZ; HELD; SCHREPP, 2008; SCHREPP; HINDERKS; THOMASCHEWSKI, 2017b). Dessa forma, o questionário ajuda a medir a qualidade da aplicação a partir da experiência do usuário e o teste de usabilidade ajuda a detectar problemas específicos e concretos (SCHREPP; HINDERKS; THOMASCHEWSKI, 2014).

A versão curta do UEQ avalia a experiência do usuário em duas escalas: qualidade pragmática e qualidade hedônica. A qualidade pragmática está relacionada aos objetivos que o usuário quer alcançar ao usar o produto, enquanto que a qualidade hedônica está relacionada ao prazer e a diversão (SCHREPP; HINDERKS; THOMASCHEWSKI, 2017b). Como mostra a tabela Tabela 5, o questionário é composto de 8 itens de termos opostos, os quatro primeiros itens avaliam a qualidade pragmática e os quatro últimos avaliam a qualidade hedônica.

O procedimento utilizado foi baseado em um estudo de validação usado para avaliar

¹ O questionário e a planilha para análise dos resultados estão disponíveis em <www.ueq-online.org>.

Figura 22 – Indicadores de produção classificado por professor.



Fonte: elaborado pelo autor.

Tabela 5 – Relação entre itens do questionário e qualidade avaliada.

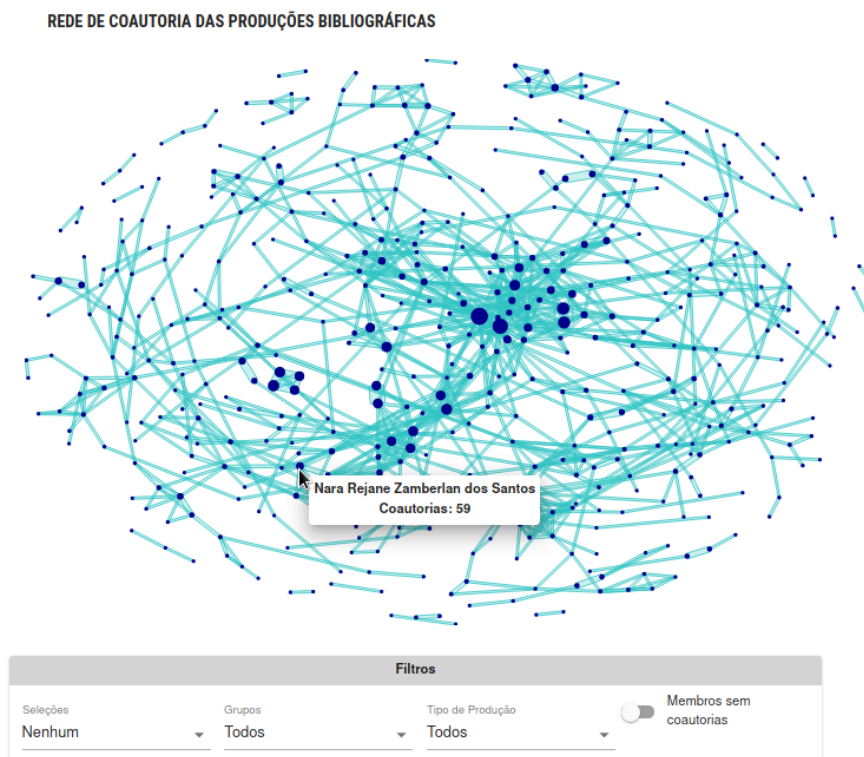
Itens	Escala
Obstrutivo — Condutor	Qualidade Pragmática
Complicado — Fácil	Qualidade Pragmática
Ineficiente — Eficiente	Qualidade Pragmática
Confuso — Evidente	Qualidade Pragmática
Aborrecido — Excitante	Qualidade Hedônica
Desinteressante — Interessante	Qualidade Hedônica
Convencional — Original	Qualidade Hedônica
Comum — Vanguardista	Qualidade Hedônica

Fonte: elaborado pelo autor.

o próprio questionário (LAUGWITZ; HELD; SCHREPP, 2008). Foram 9 participantes, todos estudantes de graduação da UNIPAMPA com experiência no uso de computadores. Para que o usuário explorasse todas as páginas e principais funcionalidades da aplicação logo antes de responder o formulário, foram solicitadas tarefas como: visualizar indicadores, visualizar informações contidas nos indicadores, aplicar filtros nos indicadores, listar informações sobre produções, fazer *download* dos dados do gráfico, etc.

Todos os participantes avaliaram a aplicação no mesmo ambiente: computador de mesa, processador AMD A8-6500B APU 3.5 GHz, 3.1 GB de memória RAM, monitor de

Figura 23 – Rede de coautoria de todos os professores e tipos de produções bibliográficas.



Fonte: elaborado pelo autor.

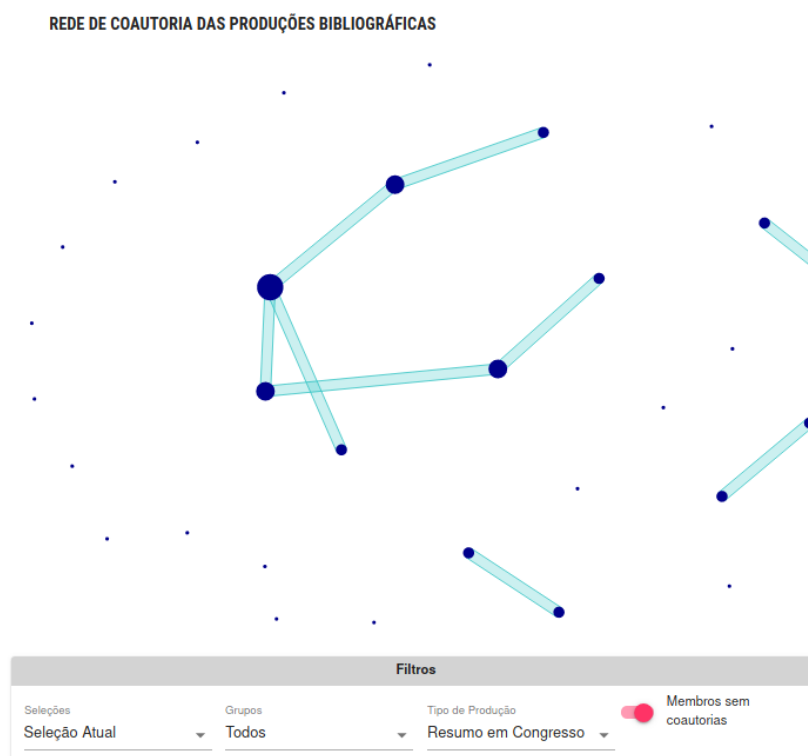
20 polegadas, sistema operacional Ubuntu 16.04 LTS e navegador Google Chrome versão 71.0.3578.98. A aplicação *web* estava disponível pela rede local.

O procedimento se repete para cada participante com os passos a seguir:

1. Explicar o objetivo da avaliação.
2. Explicar brevemente o Openlattes e os tipos de indicadores que apresenta.
3. Solicitar a execução das tarefas, disponíveis no Apêndice A, e observar a sua execução. Encorajar o participante a verbalizar durante a execução das tarefas.
4. Ao terminar todas as tarefas, encerrar a aplicação e solicitar para que o UEQ, disponível no Apêndice B, seja preenchido. .
5. Iniciar a aplicação novamente e permitir que o participante comente sobre problemas de usabilidade encontrados.
6. Discutir com o participante sobre problemas de usabilidade observados durante a execução das tarefas.

A maioria dos problemas de usabilidade detectados estão relacionados a intuitividade e falta de explicação sobre os dados apresentados. Por exemplo, a aplicação não

Figura 24 – Rede de coautoria com filtros aplicados.



Fonte: elaborado pelo autor.

informa que os dados saíram dos currículos Lattes dos docentes da UNIPAMPA. A rede de coautoria também tem esse problema, porque não informa o quê está representando e isso torna o seu entendimento mais demorado. Não há uma explicação de que os nós representam autores e arestas representam as produções em coautoria. Inclusive, não há explicação sobre o quê o tamanho do nó e a largura da aresta representam.

Os problemas detectados nos indicadores de produção são relativamente menos problemáticos. O título do gráfico apresenta o tipo de dado que está sendo contabilizado (produções bibliográficas ou orientações concluídas) de forma muito discreta. Além disso, quando o indicador é contabilizado por professor, não há nenhuma informação de que somente os 30 professores com mais produções são apresentados. A funcionalidade de listar as informações das produções contabilizadas também está discreta e pouco intuitiva.

Outros problemas dos indicadores de produção estão relacionados a apresentação do gráfico. Em alguns momentos, o eixo do total de produções apresenta valores decimais. Em outros momentos, quando o total de produções é muito baixo, a barra do gráfico fica invisível e o valor total pode ser interpretado como zero.

Os filtros também apresentaram alguns problemas de usabilidade. Não está claro que, ao alterar algum filtro, outro filtro pode ser alterado para seu valor padrão. Por isso, o usuário precisa repetir a seleção pois não sabe qual a ordem exata que deve escolher

Figura 25 – Página para gerenciar grupos.

GRUPOS			
Nome	Excluir		
Campus Alegrete	🗑️		
Campus Bagé	🗑️		
Campus Caçapava do Sul	🗑️		
Campus Dom Pedrito	🗑️		
Rows per page: 10 ▾ 1-4 of 4 < >			

MEMBROS				Selecione um membro	SALVAR GRUPO
<input type="checkbox"/>	Nome	Curriculo Lattes	Nome em citações	Última atualização	
<input type="checkbox"/>	Adriana Gindri Salbego	Curriculo Lattes	SALBEGO, A. G.	18/3/2015	
<input type="checkbox"/>	Aldoni Gabriel Wiedenhoft	Curriculo Lattes	WIEDENHOFT, A. G.	11/8/2015	
<input type="checkbox"/>	Alessandro Bof de Oliveira	Curriculo Lattes	OLIVEIRA, A. B.;de Oliveira, Alessandro Bof	10/9/2013	

Fonte: elaborado pelo autor.

os filtros. Também não está claro que o filtro *Seleções* está relacionado as seleções de membros feitas na página *Membros*. Essa confusão também é feita na perspectiva da página *Membros*: ao fazer uma seleção de professores, não está claro que é possível filtrar os indicadores pelo menu principal.

Entretanto, os problemas de usabilidade não afetaram gravemente na experiência do usuário. A Tabela 6 apresenta os resultados do questionário com valores na escala entre -3 e 3. Os valores são obtidos calculando a média de cada participante para cada qualidade e, por fim, a média de todos os participantes. Os resultados demonstram uma avaliação positiva, tanto para a qualidade pragmática quanto para a hedônica.

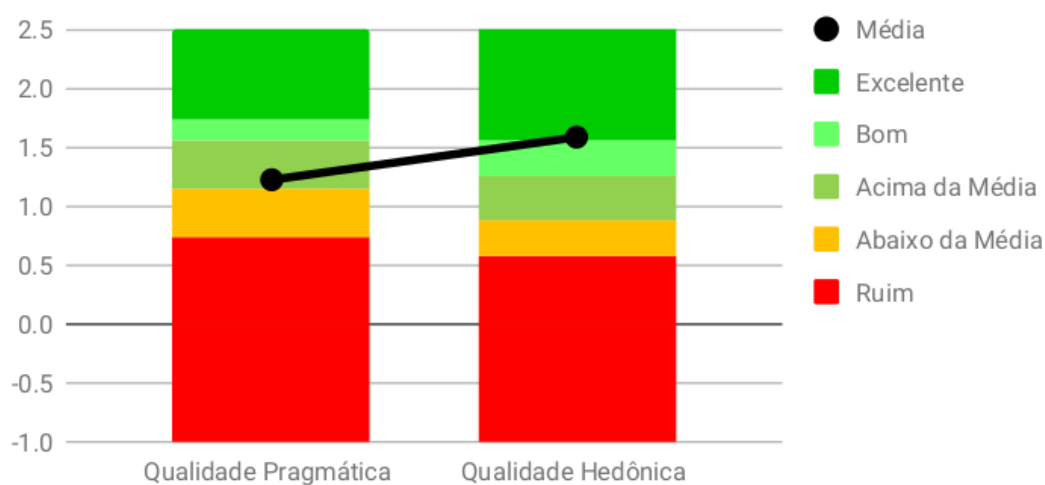
Tabela 6 – Resultado da versão curta do UEQ.

Escala	Média
Qualidade Pragmática	1,222
Qualidade Hedônica	1,583

Fonte: elaborado pelo autor.

A Figura 26 apresenta os resultados através de um *benchmark* criado para interpretar os resultado em comparação com outros produtos (SCHREPP; HINDERKS; THOMASCHEWSKI, 2017a). Os resultados são comparados com outros 280 estudos que aplicaram o UEQ em *softwares* como sistemas corporativos, páginas *web*, lojas virtuais e redes sociais.

Figura 26 – Qualidade do Openlattes em relação a outros produtos.



Fonte: elaborado pelo autor.

Em relação a esses estudos, o Openlattes tem uma qualidade pragmática acima da média, porém, quase abaixo da média. O resultado atende a expectativa porque, com base nos problemas de usabilidade encontrados, a interface não está intuitiva e explicativa o suficiente. Isso definitivamente afetou no que diz respeito as tarefas e objetivos que o usuário pretende alcançar ao usar a aplicação.

A qualidade hedônica foi avaliada como excelente, ou seja, a aplicação garante uma experiência agradável. Esse resultado pode estar atribuído ao fato das visualizações serem dinâmicas e interativas, o que torna a aplicação mais interessante. Através dos filtros, o usuário tem o controle para definir o indicador que é automaticamente apresentado conforme a sua interação. Os filtros também podem ser combinados, permitindo muitas possibilidades de visualizações.

7 CONSIDERAÇÕES FINAIS

Este trabalho apresentou a arquitetura e implementação de um sistema para coletar os dados dos currículos Lattes e apresentar indicadores bibliométricos pela *web*. Para a etapa de extração, o *scriptLattes* foi alterado para armazenar os dados dos currículos no MongoDB. O *Openlattes* foi desenvolvido para apresentar os indicadores de produção e coautoria de forma interativa, dinâmica e visual.

O sistema foi demonstrado utilizando os currículos dos docentes da UNIPAMPA. A etapa de extração funcionou como esperada: o *scriptDownload* obteve os currículos e o *script* de extração armazenou os dados de pesquisadores, produções bibliográficas, orientações e coautorias no MongoDB. Esse armazenamento foi crucial pois permitiu o acesso flexível aos dados da instituição, permitindo calcular os indicadores e obter os dados da interface.

Após implementar as funcionalidades propostas, o *Openlattes* foi avaliado com o usuário. A avaliação revelou que o *Openlattes* tem uma qualidade acima da média para seu uso prático, mas poderia ser melhor caso a interface fosse mais intuitiva e autoexplicativa. No que diz respeito a qualidade hedônica, a aplicação obteve um resultado excelente.

Outras instituições podem usar o sistema para apresentar seus indicadores na *web*¹, caso tenham acesso aos currículos em XML dos seus pesquisadores. Informações sobre a produção e colaboração acadêmica estarão disponíveis para estudantes, professores, pesquisadores e o público em geral. Além disso, a própria instituição pode utilizar essas informações, como o total de artigos de um determinado período, em relatórios de gestão. Obter esses dados manualmente dos currículos seria difícil e trabalhoso.

7.1 Trabalhos Futuros

Até agora, a coleta dos currículos com o *scriptDownload* foi feita uma única vez. A próxima etapa é tornar esse processo periódico, automático e integrado com o *script* de extração. Dessa forma, o BD e os indicadores permanecem atualizados com base em um período pré-definido. Além disso, seria ideal que o BD fosse sobrescrito somente quando a extração dos dados é finalizada. No modelo atual, os dados são escritos durante a extração e caso algum usuário estiver visualizando algum indicador, visualizará o indicador com os dados incompletos até a extração ser finalizada.

Outra etapa é melhorar o *parser* do *script* de extração para obter mais dados do currículo como produções técnicas, produções artísticas e orientações em andamento. Com essas informações armazenadas no BD, seria trivial gerar o indicador de produção para esses dados. Outras informações como titulação e área do pesquisador também ajudariam na classificação dos indicadores, principalmente na rede de coautoria.

Para melhorar a qualidade da aplicação e torná-la mais intuitiva, deseja-se conser-

¹ O *Openlattes* está disponível em: <<https://github.com/openlattes/openlattes>>.

tar os problemas de usabilidade identificados neste trabalho. A página onde estão listados os membros deveria sinalizar que os membros selecionados podem ser usados como filtro nos indicadores pelo menu principal. Outra melhoria é informar na rede de coautoria o quê os nós e arestas representam.

REFERÊNCIAS

- AMARAL, R. M. et al. Panorama da inteligência competitiva no brasil: os pesquisadores e a produção científica na plataforma lattes. **Perspectivas em Ciência da Informação**, FapUNIFESP (SciELO), v. 21, n. 4, p. 97–120, dez. 2016. Citado na página 20.
- AMORIN, C. V. Organização do currículo–plataforma lattes. **Pesqui Odontol Bras**, SciELO Brasil, v. 17, n. Supl 1, p. 18–22, 2003. Citado 2 vezes nas páginas 17 e 20.
- BASSOLI, M. **Avaliação do currículo lattes como fonte de informação para construção de indicadores: o caso da ufscar**. Dissertação (Mestrado) — Universidade Federal de São Carlos, 2017. Citado na página 20.
- BELTER, C. W. Bibliometric indicators: opportunities and limits. **Journal of the Medical Library Association: JMLA**, University Library System, University of Pittsburgh, v. 103, n. 4, p. 219–221, out. 2015. Citado na página 19.
- CALL-EM-ALL. **Material-UI**. 2018. Disponível em: <<https://material-ui.com/>>. Acesso em: 27 de novembro de 2018. Citado na página 36.
- CNPQ. **Plataforma Lattes**. 2018. Disponível em: <<http://lattes.cnpq.br/>>. Acesso em: 13 de junho de 2018. Citado 2 vezes nas páginas 17 e 23.
- DING, Y. Scientific collaboration and endorsement: Network analysis of coauthorship and citation networks. **Journal of Informetrics**, Elsevier, v. 5, n. 1, p. 187–203, 2011. Citado na página 19.
- DURIEUX, V.; GEVENOIS, P. A. Bibliometric indicators: Quality measurements of scientific publication. **Radiology**, Radiological Society of North America (RSNA), v. 255, n. 2, p. 342–351, mai. 2010. Citado na página 19.
- E-LATTES. **e-Lattes**. 2019. Disponível em: <<http://unb.elattes.com.br/>>. Acesso em: 09 de abril de 2019. Citado 2 vezes nas páginas 27 e 29.
- FACEBOOK OPEN SOURCE. **React**. 2018. Disponível em: <<https://reactjs.org/>>. Acesso em: 26 de novembro de 2018. Citado na página 35.
- FERRAZ, R. R. N.; QUONIAM, L. M. A utilização da ferramenta computacional scriptlattes para avaliação das competências em pesquisa no brasil. **PRISMA.COM**, n. 21, p. 222–234, 2013. Citado 2 vezes nas páginas 22 e 23.
- FRANCISCO, L. G. **Tutorial para o uso do scriptLattes**. [S.l.], 2014. Disponível em: <<http://scriptlattes.sourceforge.net/tutorial-para-o-uso-do-scriptLattes.pdf>>. Citado na página 22.
- FUNDEP. **Sistema Somos**. 2018. Disponível em: <<http://www.fundep.ufmg.br/solucoes/sistema-somos/>>. Acesso em: 11 de junho de 2018. Citado 2 vezes nas páginas 17 e 24.
- GREGOLIN, J. A. et al. Análise da produção científica a partir de indicadores bibliométricos. In: **Indicadores de ciência, tecnologia e inovação em São Paulo 2004**. [S.l.]: FAPESP, 2005. v. 1, cap. 5, p. 44. Citado na página 17.

- GUTIÉRREZ-SALCEDO, M. et al. Some bibliometric procedures for analyzing and evaluating research fields. **Applied Intelligence**, Springer Nature, dez. 2017. Citado na página 19.
- LAUGWITZ, B.; HELD, T.; SCHREPP, M. Construction and evaluation of a user experience questionnaire. In: SPRINGER. **Symposium of the Austrian HCI and Usability Engineering Group**. [S.l.], 2008. p. 63–76. Citado 2 vezes nas páginas 48 e 49.
- MACIEL, R. S. et al. A plataforma lattes como recurso estratégico para a gestão dos programas de pós-graduação. In: **XVIII Encontro Nacional de Pesquisa em Ciência da Informação**. [S.l.: s.n.], 2017. Citado na página 20.
- MEEKS, E. **Semiotic**. 2018. Disponível em: <<https://emeeks.github.io/semiotic/>>. Acesso em: 18 de junho de 2018. Citado 2 vezes nas páginas 35 e 36.
- MENA-CHALCO, J. P.; CESAR-JR, R. M. Prospecção de dados acadêmicos de currículos lattes através de scriptlattes. In: **Bibliometria e Cientometria: reflexões teóricas e interfaces**. [S.l.]: São Carlos: Pedro & João Editores. Maria Cristina Piumbato Innocentini Hayashi e Jacqueline Leta (Orgs.), 2013. p. 109–128. Citado 6 vezes nas páginas 17, 20, 21, 22, 23 e 24.
- MENA-CHALCO, J. P.; DIGIAMPIETRI, L. A.; OLIVEIRA, L. B. Perfil de produção bibliográfica dos programas brasileiros de pós-graduação em ciência da computação. In: **Encontro Brasileiro de Bibliometria e Cientometria**. [S.l.: s.n.], 2012. Citado na página 20.
- METEOR DEVELOPMENT GROUP. **Apollo**. 2019. Disponível em: <<https://www.apollographql.com/>>. Acesso em: 28 de março de 2019. Citado 2 vezes nas páginas 35 e 36.
- MONGODB INC. **MongoDB**. 2019. Disponível em: <<https://www.mongodb.com/>>. Acesso em: 03 de fevereiro de 2019. Citado na página 35.
- MONGODB INC. **PyMongo 3.7.2 Documentation**. 2019. Disponível em: <<https://api.mongodb.com/python/3.7.2/>>. Acesso em: 11 de fevereiro de 2019. Citado na página 35.
- MONGODB INC. **What is MongoDB?** 2019. Disponível em: <<https://www.mongodb.com/what-is-mongodb>>. Acesso em: 03 de fevereiro de 2019. Citado na página 35.
- NAVARRO, G. A guided tour to approximate string matching. **ACM computing surveys (CSUR)**, ACM, v. 33, n. 1, p. 31–88, 2001. Citado na página 22.
- NODE.JS FOUNDATION. **About Node.js**. 2019. Disponível em: <<https://nodejs.org/en/about/>>. Acesso em: 12 de março de 2019. Citado na página 35.
- PRITCHARD, A. Statistical bibliography or bibliometrics. **Journal of Documentation**, New York, v. 25, n. 4, p. 348–349, 1969. Citado na página 19.
- SAMPAIO, R. B.; BATISTA JUNIOR, A. de A.; MENA-CHALCO, J. P. e-Lattes: Um novo arcabouço em linguagem R para análise do currículo Lattes. In: **6 Encontro Brasileiro de Bibliometria e Cientometria**. [S.l.: s.n.], 2018. Citado na página 28.

SCHREPP, M.; HINDERKS, A.; THOMASCHEWSKI, J. Applying the user experience questionnaire (UEQ) in different evaluation scenarios. In: SPRINGER. **International Conference of Design, User Experience, and Usability**. [S.l.], 2014. p. 383–392. Citado na página 48.

SCHREPP, M.; HINDERKS, A.; THOMASCHEWSKI, J. Construction of a benchmark for the user experience questionnaire (UEQ). **IJIMAI**, v. 4, n. 4, p. 40–44, 2017. Citado na página 52.

SCHREPP, M.; HINDERKS, A.; THOMASCHEWSKI, J. Design and evaluation of a short version of the user experience questionnaire (UEQ-S). **IJIMAI**, v. 4, n. 6, p. 103–108, 2017. Citado na página 48.

SILVA, F. M.; SMIT, J. W. Organização da informação em sistemas eletrônicos abertos de informação científica & tecnológica: análise da plataforma lattes. **Perspectivas em ciência da informação**, v. 14, n. 1, p. 77–98, 2009. Citado na página 20.

STELATEK. **Stela Experta**. 2018. Disponível em: <<http://site.stelaexperta.com.br/>>. Acesso em: 07 de dezembro de 2018. Citado 4 vezes nas páginas 17, 25, 27 e 28.

THOMSON REUTERS. **Whitepaper using bibliometrics: a guide to evaluating research performance with citation data**. [S.l.]: Thomson Reuters, 2008. Citado na página 19.

UFABC. **UFABC-professores**. 2019. Disponível em: <<http://pesquisa.ufabc.edu.br/cientometria/UFABC-professores/>>. Acesso em: 03 de julho de 2019. Citado 3 vezes nas páginas 23, 24 e 25.

UFMG. **Somos UFMG**. 2019. Disponível em: <<http://www.somos.ufmg.br/>>. Acesso em: 03 de julho de 2019. Citado 2 vezes nas páginas 25 e 26.

WILDGAARD, L.; SCHNEIDER, J. W.; LARSEN, B. A review of the characteristics of 108 author-level bibliometric indicators. **Scientometrics**, Springer Nature, v. 101, n. 1, p. 125–158, set. 2014. Citado na página 19.

Apêndices

APÊNDICE A – LISTA DE TAREFAS

Lista de tarefas para avaliação do Openlattes

1. Descubra a quantidade total de produções bibliográficas no ano de 2009.
2. Visualize a quantidade de produções bibliográficas por ano, contabilizando apenas as produções dos tipos: apresentação de trabalho, artigo aceito, artigo em periódico, resumo em congresso e resumo expandido em congresso.
3. Descubra o total de orientações concluídas do professor do Campus Bagé com a quinta maior quantidade de orientações.
4. Liste os títulos das Teses de Doutorado orientadas no ano de 2015.
5. Descubra o professor do Campus Dom Pedrito com a maior quantidade de coautorias.
6. Visualize os professores do Campus Alegrete que não possuem coautorias em Capítulos de Livro
7. Visualize a rede de coautoria de apenas um professor.
8. Escolha 5 professores, salve essa seleção com o nome “Teste” e visualize o indicador de produções bibliográficas por ano para a seleção Teste. Remova a seleção salva “Teste”.
9. Faça o download do arquivo CSV do indicador de orientações por ano.

