

**UNIVERSIDADE FEDERAL DE SANTA CATARINA
BIBLIOTECA UNIVERSITÁRIA**

Carlos Bonetti

**SILQ 2 - SISTEMA DE INTEGRAÇÃO LATTES-QUALIS:
REESTRUTURAÇÃO E MELHORIAS**

Florianópolis

2016

Carlos Bonetti

**SILQ 2 - SISTEMA DE INTEGRAÇÃO LATTES-QUALIS:
REESTRUTURAÇÃO E MELHORIAS**

Tese submetida ao Programa de Graduação
em Ciência da Computação para a obtenção
do Grau de Bacharel.
Orientador: Prof. Dr. Carina F. Dor-
neles

Florianópolis

2016

Ficha de identificação da obra elaborada pelo autor,
através do Programa de Geração Automática da Biblioteca Universitária da UFSC.

SobrenomeAutor, NomeAutor

Título do trabalho : Subtítulo do trabalho / NomeAutor

SobrenomeAutor ; orientador, NomeOrientador

SobrenomeOrientador ; coorientador, NomeCoorientador

SobrenomeCoorientador. - Florianópolis, SC, 2014.

100 p.

Tese (doutorado) - Universidade Federal de Santa
Catarina, Centro Tecnológico. Programa de Pós-Graduação em
Ciência da Computação.

Inclui referências

1. Ciência da Computação. 2. Exemplo de ficha
catalográfica. I. SobrenomeOrientador, NomeOrientador. II.
SobrenomeCoorientador, NomeCoorientador. III. Universidade
Federal de Santa Catarina. Programa de Pós-Graduação em
Ciência da Computação. IV. Título.

Carlos Bonetti

**SILQ 2 - SISTEMA DE INTEGRAÇÃO
LATTES-QUALIS: REESTRUTURAÇÃO E MELHORIAS**

Esta Tese foi julgada aprovada para a obtenção do Título de “Bacharel”, e aprovada em sua forma final pelo Programa de Graduação em Ciência da Computação.

Florianópolis, 15 de maio 2016.

Prof. Dr. Renato Cislaghi
Coordenador de Projetos

Banca Examinadora:

Prof. Dr. Carina F. Dorneles
Orientador

Prof. Dr. Raul Sidnei Wazlawick

Prof. Ronaldo dos Santos Mello

RESUMO

O texto do resumo deve ser digitado, em um único bloco, sem espaço de parágrafo. O resumo deve ser significativo, composto de uma sequência de frases concisas, afirmativas e não de uma enumeração de tópicos. Não deve conter citações. Deve usar o verbo na voz passiva. Abaixo do resumo, deve-se informar as palavras-chave (palavras ou expressões significativas retiradas do texto) ou, termos retirados de thesaurus da área.

Palavras-chave: Palavra-chave 1. Palavra-chave 2. Palavra-chave 3.

ABSTRACT

Resumo traduzido para outros idiomas, neste caso, inglês. Segue o formato do resumo feito na língua vernácula. As palavras-chave traduzidas, versão em língua estrangeira, são colocadas abaixo do texto precedidas pela expressão “Keywords”, separadas por ponto.

Keywords: Keyword 1. Keyword 2. Keyword 3.

LISTA DE FIGURAS

| | | |
|----------|---|----|
| Figura 1 | Exemplo de conjunto de dados do Qualis 2014. | 23 |
|----------|---|----|

LISTA DE TABELAS

| | | |
|----------|---|----|
| Tabela 1 | Número de periódicos extraídos dos dados Qualis | 30 |
|----------|---|----|

LISTA DE ABREVIATURAS E SIGLAS

SUMÁRIO

| | | |
|----------|---|----|
| 1 | INTRODUÇÃO | 17 |
| 1.1 | OBJETIVOS | 18 |
| 1.2 | PROCEDIMENTOS METODOLÓGICOS | 19 |
| 2 | CONCEITOS E FUNDAMENTAÇÃO TEÓRICA | 21 |
| 2.1 | PRODUÇÃO E AVALIAÇÃO CIENTÍFICAS NO BRASIL | 21 |
| 2.1.1 | Qualis | 22 |
| 2.1.2 | A Plataforma Lattes | 23 |
| 2.2 | TODO: CONCEITOS COMPUTACIONAIS | 24 |
| 3 | O SISTEMA DE INTEGRAÇÃO LATTES-QUALIS (SILQ) | 25 |
| 3.1 | HISTÓRIO E VISÃO GERAL DO SILQ-1 | 25 |
| 3.1.1 | “Trabalhos futuros” citados pelo SILQ-1 | 26 |
| 3.1.1.1 | Controle de persistência e transações | 26 |
| 3.1.1.2 | Automatização da atualização dos dados Qualis | 26 |
| 3.1.1.3 | <i>Fine-tuning</i> da função de similaridade | 27 |
| 3.1.1.4 | Considerar outras informações do pesquisador | 27 |
| 3.1.1.5 | Relação entre orientações e produções associadas | 27 |
| 3.1.1.6 | Busca por informações de eventos | 28 |
| 3.1.1.7 | Feedback do usuário | 28 |
| 3.1.1.8 | Período de participação em grupo de pesquisa | 28 |
| 3.2 | SILQ 2 | 29 |
| 3.2.1 | Extração e inserção dos novos dados Qualis | 29 |
| 3.2.2 | Visão geral da nova arquitetura | 30 |
| 4 | CONCLUSÃO | 31 |
| | REFERÊNCIAS | 33 |

1 INTRODUÇÃO

Introdução copiada do relatório de Proposta da disciplina passada. **SERÁ ALTERADA!**

A Plataforma Lattes, criada e mantida pela CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico), é um sistema de informação responsável pela integração da base de dados de currículos, grupos de pesquisa e instituições. O Currículo Lattes se tornou o padrão nacional no registro da vida científica de estudantes e professores e é hoje adotada por institutos e universidades de todo o país. (CNPQ, 2015b)

No Currículo Lattes é possível a inserção de dados gerais do pesquisador, produção bibliográfica, orientações, citações, entre outros dados. No módulo Produção Bibliográfica, por exemplo, é possível a inserção de artigos publicados ou aceitos para publicação em periódicos indexados pelo ISSN. (CNPQ, 2015a)

A classificação da qualidade da produção intelectual é realizada pela CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior) através de um conjunto de procedimentos denominado Qualis. O Qualis afere a qualidade de artigos e outras produções a partir da análise da qualidade dos veículos de divulgação, ou seja, periódicos científicos. Esta classificação é realizada pelas áreas de avaliação em um processo anual de atualização, sendo os veículos enquadrados em estratos indicativos de qualidade (A1 - o mais elevado; A2; B1; B2; B3; B4; B5; C - com peso zero. (CAPES, 2015a)

Apesar da Plataforma Lattes possuir um módulo de inclusão de publicações e permitir a definição do veículo onde este foi publicado, não há qualquer tipo de conexão entre os sistemas Lattes e Qualis, ou seja, o processo de avaliação de uma publicação (que é feita através da avaliação do veículo onde este foi publicado) era realizado de forma manual.

O Sistema de Integração Lattes Qualis (SILQ) surgiu no ano de 2015, desenvolvido como Trabalho de Conclusão de Curso dos alunos Felipe Nedel Mendes de Aguiar e Maria Eloísa Costa do curso de Ciência da Computação da Universidade Federal de Santa Catarina (UFSC). Orientados pela Prof. Carina F. Dorneles, da mesma instituição, o objetivo do sistema é a classificação automática de produções científicas do currículo Lattes do pesquisador, no quesito artigos e participações de conferências, através de busca de similaridade de dados extraídos do WebQualis, a partir de uma interface web amigável.

(AGUIAR; COSTA, 2015)

A primeira versão do sistema foi finalizada no primeiro semestre de 2015 e encontra-se disponível de forma pública e gratuita através do endereço <http://silq.inf.ufsc.br/>.

A primeira versão do SILQ, no entanto, possui em seu banco de dados de eventos e periódicos extraídos do Qualis apenas dados referentes ao triênio 2010-2012, já que, até metade de 2015, o Qualis realizava classificações de forma trienal. No momento de escrita desta proposta, no entanto, o Qualis alterou seu modus operandi para classificações anuais, já disponibilizando dados referentes aos anos 2013 e 2014, que ainda não estão na base de dados SILQ.

Outra questão relacionada à primeira versão do SILQ é a falta de alguns recursos que poderiam facilitar o processo de avaliação e acompanhamento de currículos, como gráficos de classificações dentro de grupos de pesquisa. Outra característica importante seria a capacidade de integração com outros sistemas, facilitando o reuso do serviço disponibilizado pelo SILQ em trabalhos futuros.

A proposta deste trabalho, portanto, é a criação de uma segunda versão do SILQ, desenvolvida a partir do código existente da primeira, a fim de realizar as mudanças sugeridas acima, junto com a atualização da base de dados e arquitetura do sistema para a inclusão das novas qualificações Qualis em um ritmo anual.

1.1 OBJETIVOS

O objetivo geral deste trabalho é a reestruturação arquitetural do Sistema de Integração Lattes-Qualis (SILQ) de forma a permitir a inclusão de melhorias de usabilidade, API de integração com outros sistemas e atualização da base de dados conforme novas classificações disponibilizadas pelo Qualis, bem como um estudo de impacto do desempenho do sistema após a inclusão destes novos registros.

Como objetivos específicos, é possível citar:

- Reestruturação da arquitetura e banco de dados do SILQ a fim de suportar classificações de eventos e periódicos disponibilizados em um ritmo anual;
- Atualização do banco de dados do sistema com as últimas classificações disponibilizadas pelo Qualis (anos 2013 e 2014);
- Estudo de impacto de desempenho do sistema após a reestruturação do banco de dados e inclusão das novas classificações

Qualis;

- Criação de uma API pública de disponibilização dos serviços do SILQ, via camada de aplicação REST para integração com outros sistemas;
- Alterações na interface do sistema incluindo migração de framework de interface, novos gráficos de acompanhamento de grupos de pesquisa e pesquisadores;
- Melhorias gerais: geração de tabela excel a partir de qualificações, recuperação de senhas para usuários cadastrados, melhorar a precisão da comparação por similaridade, levantamento de produção por veículo de publicação.

1.2 PROCEDIMENTOS METODOLÓGICOS

A primeira etapa do trabalho envolverá a migração tecnológica da camada de interface e desacoplamento entre server e client side. Para isto, serão utilizadas práticas ágeis de desenvolvimento de software incluindo desenvolvimento orientado a testes e eventuais modelagens lógicas de arquitetura de sistema usando práticas da Engenharia de Software.

A segunda etapa, a de migração de base de dados e estudo de impacto terá um caráter mais exploratório e quantitativo: pretende-se analisar o desempenho do sistema sob diferentes modelos lógicos de banco de dados e coletar dados pertinentes que levem à escolha de um modelo em detrimento do outro.

Os softwares que serão utilizados em todas as etapas do trabalho são do tipo software livre, incluindo IDE de desenvolvimento, banco de dados e servidor de aplicação. Para a publicação da versão final do sistema continuará a ser utilizado os servidores e domínio da UFSC, onde a primeira versão do SILQ já encontra-se publicado.

Os recursos físicos e ambientes de trabalho utilizados no processo de desenvolvimento deste trabalho serão todos de posse pessoal do autor.

2 CONCEITOS E FUNDAMENTAÇÃO TEÓRICA

Neste capítulo serão feitas uma breve descrição e conceituação básica do processo de produção e avaliação científica no Brasil, com o objetivo de familiarizar o leitor com a problemática envolvida neste processo e o que o SILQ propôs para automatizá-lo. Também serão introduzidos os conceitos tecnológicos e computacionais que de alguma forma foram explorados pelo sistema para chegar a este objetivo.

Esta revisão de literatura, porém, não será exaustiva ao ponto de revisitar os mesmos assuntos abordados no trabalho original de (AGUIAR; COSTA, 2015) e focará somente nos pontos onde este novo trabalho divergiu. Alguns assuntos fundamentais, entretanto, foram expostos novamente com o intuito de facilitar o processo de entendimento deste artigo como um trabalho independente. É o caso, por exemplo, das seções relacionadas aos Qualis, Lattes e das tecnologias base utilizadas no sistema.

2.1 PRODUÇÃO E AVALIAÇÃO CIENTÍFICAS NO BRASIL

A produção científica é uma das atividades universitárias que cumprem uma função básica dentro das instituições e que merece notável destaque. É através dela que o conhecimento produzido na Universidade é difundido à sociedade, externalizando descobertas para outros pesquisadores e para a própria comunidade. (CRISPIM, 2014)

No Brasil, o Ministério de Ciência, Tecnologia e Inovação (MCTI) é o órgão de administração direta que possui, entre sua lista de competências, a política nacional de pesquisa científica, tecnológica e inovação; e o planejamento, coordenação, supervisão e controle das atividades da ciência e tecnologia. (MCTI, 2016?)

Uma das agências subordinadas ao MCTI é o CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico), órgão que tem como principais atribuições “fomentar a pesquisa científica e tecnológica e incentivar a formação de pesquisadores brasileiros” (CNPQ, 2016).

Analizando não só a produção científica, cabe ao Ministério da Educação (MEC), através de sua fundação subordinada, a CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior) a avaliação dos programas de pós-graduação das universidades Brasileiras (CAPES, 2015b). Um dos principais critérios de avaliação se dá pela quantidade e qualidade das produções científicas, através de um conjunto de

procedimentos denominado Qualis (CAPES, 2015a).

2.1.1 Qualis

A classificação da qualidade da produção intelectual é realizada pela CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior) através de um conjunto de procedimentos denominado Qualis. O Qualis afere a qualidade de artigos e outras produções a partir da análise da qualidade dos veículos de divulgação, ou seja, periódicos científicos. Esta classificação é realizada pelas áreas de avaliação em um processo anual de atualização, sendo os veículos enquadrados em estratos indicativos de qualidade (A1 - o mais elevado; A2; B1; B2; B3; B4; B5; C - com peso zero. (CAPES, 2015a)

O resultado desta estratificação é uma lista com a classificação dos veículos de publicação utilizados pelos programas de pós-graduação (*journals*, revistas, periódicos etc), separados por área e ano de avaliação (um mesmo veículo tem uma avaliação diferente para cada ano de avaliação e área de conhecimento). Desta forma, a avaliação da qualidade de uma produção de um pesquisador é realizado de forma indireta, atribuindo-se a esta produção o estrato dado àquele veículo onde ela foi publicada.

Note-se que o mesmo periódico, ao ser classificado em duas ou mais áreas distintas, pode receber diferentes avaliações. Isto não constitui inconsistência, mas expressa o valor atribuído, em cada área, à pertinência do conteúdo veiculado. Por isso, não se pretende com esta classificação que é específica para o processo de avaliação de cada área, definir qualidade de periódicos de forma absoluta. (CAPES, 2015a)

Até o final do ano de 2015, como cita o trabalho de (AGUIAR; COSTA, 2015, p. 31), os dados os resultados das avaliações Qualis encontravam-se estruturados em documentos no formato PDF e XLS. Também encontravam-se disponível apenas dados referentes ao triênio de 2010-2012. No início do ano de 2016, porém, a Qualis alterou seu modo de operação e passou a realizar avaliações anuais, já disponibilizando os novos dados de 2013 e 2014, junto com os antigos, em formato CSV¹ através do WebQualis² (Hoje plataforma Sucupira).

¹Comma-separated values

²<http://qualis.capes.gov.br/webqualis/>

Figura 1 – Exemplo de conjunto de dados do Qualis 2014.

| | | | |
|-----------|---|-----------------------|----|
| 1041-4347 | IEEE Transactions on Knowledge and Data Engineering (Print) | CIÊNCIA DA COMPUTAÇÃO | A1 |
| 0018-9464 | IEEE Transactions on Magnetics | CIÊNCIA DA COMPUTAÇÃO | B4 |
| 0278-0062 | IEEE Transactions on Medical Imaging (Print) | CIÊNCIA DA COMPUTAÇÃO | A1 |
| 1536-1233 | IEEE Transactions on Mobile Computing | CIÊNCIA DA COMPUTAÇÃO | A2 |
| 1520-9210 | IEEE Transactions on Multimedia | CIÊNCIA DA COMPUTAÇÃO | A2 |
| 2162-237X | IEEE Transactions on Neural Networks and Learning Systems | CIÊNCIA DA COMPUTAÇÃO | A1 |
| 0018-9499 | IEEE Transactions on Nuclear Science | CIÊNCIA DA COMPUTAÇÃO | B1 |
| 1045-9219 | IEEE Transactions on Parallel and Distributed Systems (Print) | CIÊNCIA DA COMPUTAÇÃO | A2 |
| 0885-8950 | IEEE Transactions on Power Systems | CIÊNCIA DA COMPUTAÇÃO | B2 |
| 0098-5589 | IEEE Transactions on Software Engineering | CIÊNCIA DA COMPUTAÇÃO | A1 |
| 1083-4427 | IEEE Transactions on Systems, Man and Cybernetics. Part A | CIÊNCIA DA COMPUTAÇÃO | A2 |
| 1094-6977 | IEEE Transactions on Systems, Man and Cybernetics. Part B | CIÊNCIA DA COMPUTAÇÃO | A2 |
| 0018-9545 | IEEE Transactions on Vehicular Technology | CIÊNCIA DA COMPUTAÇÃO | A1 |
| 1063-8210 | IEEE Transactions on Very Large Scale Integration (VLSI) Systems | CIÊNCIA DA COMPUTAÇÃO | A2 |
| 1077-2626 | IEEE Transactions on Visualization and Computer Graphics | CIÊNCIA DA COMPUTAÇÃO | A2 |
| 1536-1276 | IEEE Transactions on Wireless Communications | CIÊNCIA DA COMPUTAÇÃO | A1 |
| 1536-1284 | IEEE Wireless Communications | CIÊNCIA DA COMPUTAÇÃO | A1 |
| 2162-2337 | IEEE Wireless Communications Letters | CIÊNCIA DA COMPUTAÇÃO | B4 |
| 1932-4537 | IEEE eTransactions on Network and Service Management | CIÊNCIA DA COMPUTAÇÃO | B3 |
| 1932-8540 | IEEE-RITA | CIÊNCIA DA COMPUTAÇÃO | B5 |
| 1545-5963 | IEEE/ACM Transactions on Computational Biology and Bioinformatics | CIÊNCIA DA COMPUTAÇÃO | B1 |
| 0916-8532 | IEICE Transactions on Information and Systems | CIÊNCIA DA COMPUTAÇÃO | B1 |
| 1751-861X | IET Computers & Digital Techniques (Online) | CIÊNCIA DA COMPUTAÇÃO | B1 |
| 1751-8601 | IET Computers & Digital Techniques (Print) | CIÊNCIA DA COMPUTAÇÃO | B1 |
| 1751-8806 | IET Software (Print) | CIÊNCIA DA COMPUTAÇÃO | B1 |
| 1091-9856 | INFORMS Journal on Computing | CIÊNCIA DA COMPUTAÇÃO | B1 |
| 1526-5528 | INFORMS Journal on Computing (Online) | CIÊNCIA DA COMPUTAÇÃO | B1 |
| 2090-4355 | ISRN Communications and Networking | CIÊNCIA DA COMPUTAÇÃO | C |
| 2237-4523 | Iberoamerican Journal of Applied Computing | CIÊNCIA DA COMPUTAÇÃO | B5 |
| 0262-8856 | Image and Vision Computing | CIÊNCIA DA COMPUTAÇÃO | A1 |

2.1.2 A Plataforma Lattes

Desde a década de 80 notava-se a necessidade de um formulário padrão para registro dos currículos de pesquisadores brasileiros que possibilitasse a busca e seleção de especialistas bem como a geração de um mapa sobre a distribuição da pesquisa científica no Brasil. Em 1999, o CNPq lançou a Plataforma Lattes³, um sistema de informação para registro e consulta de currículos acadêmicos, e padronizou o Currículo Lattes como o formulário de currículo a ser utilizado no âmbito do Ministério da Ciência e Tecnologia e CNPq. (CNPQ, 2016, 2015b)

Entre os dados contidos em um currículo Lattes, estão os dados gerais do pesquisador, produção bibliográfica, orientações, citações, formação acadêmica, etc. No módulo Produção Bibliográfica, por exemplo, é possível a inserção de artigos publicados em periódicos indexados pelo ISSN⁴, livros e capítulos, textos em jornais ou revistas, trabalhos publicados em anais de eventos, entre outras formas de produções

³<http://lattes.cnpq.br/>

⁴International Standard Serial Number

(CNPQ, 2015a). Estes dados de publicações seriam usados pela primeira versão do SILQ (AGUIAR; COSTA, 2015) para a avaliação automática de currículos.

A Plataforma Lattes, além de apresentar seus currículos através de uma interface web usando HTML, também disponibiliza os dados em formato XML⁵, uma linguagem de marcação semi-estruturada legível por humanos e máquinas (AGUIAR; COSTA, 2015, p. 34). Foi este o formato utilizado pelo SILQ para a extração dos dados Lattes de um pesquisador, conforme relatado em (AGUIAR; COSTA, 2015).

2.2 TODO: CONCEITOS COMPUTACIONAIS

REST, API de Integração
XML - XPATH - Extração

Trabalhos correlatos:

<http://qualis.renesp.com.br/index.php>

⁵Extensible Markup Language

3 O SISTEMA DE INTEGRAÇÃO LATTES-QUALIS (SILQ)

O presente trabalho é uma continuação dos esforços iniciados por (AGUIAR; COSTA, 2015). Neste capítulo, portanto, será apresentado o histórico do trabalho original, os motivos que levaram ao desenvolvimento de um novo trabalho para sua continuação, comparações do que foi de fato alterado no sistema e por que estas alterações foram necessárias.

Para mais fácil compreensão e contextualização, durante o texto serão utilizados os termos “trabalho original”, “artigo original”, “SILQ-1” ou ainda “primeira versão do sistema” ou sinônimos para referir-se ao trabalho de (AGUIAR; COSTA, 2015). Em contrapartida, os termos, “SILQ-2”, “nova versão do sistema” e derivados referem-se a este trabalho.

3.1 HISTÓRIO E VISÃO GERAL DO SILQ-1

Este trabalho é uma continuação de (AGUIAR; COSTA, 2015), um Trabalho de Conclusão de Curso de alunos do curso de Ciência da Computação da UFSC, orientados pela Professora Carina F. Dorneles. O objetivo deste trabalho de 2015 era a criação de um sistema que deveria ser capaz de “qualificar produções científicas, no quesito artigos e participações em conferências, por busca por similaridade de dados com os dados extraídos do WebQualis” (AGUIAR; COSTA, 2015, p. 26-27). Este objetivo foi alcançado com a criação da primeira versão do Sistema de Integração Lattes-Qualis (SILQ), lançado no segundo semestre de 2015 e disponível no sítio <http://silq.inf.ufsc.br/>.

Apesar de estável e com sua função principal sendo desempenhada de forma satisfatória, o SILQ-1 deixou algumas lacunas e melhorias a serem desenvolvidas por trabalhos futuros. Segundo os próprios autores, “[...] o SILQ foi concebido para ser uma ferramenta de domínio público e vários projetos devem nascer a partir dele. A continuidade do projeto só tem a acrescentar ao mundo acadêmico [...]” (AGUIAR; COSTA, 2015, p. 79), o que motivou a criação deste trabalho para a continuação do projeto original. O ponto de partida deste trabalho, portanto, foi a seção “Trabalhos futuros” do artigo original.

3.1.1 “Trabalhos futuros” citados pelo SILQ-1

Na seção de “Conclusões e Tabalhos Futuros” do artigo original, (AGUIAR; COSTA, 2015) citam 8 itens de melhoria que poderiam ser revisitados por trabalhos futuros. Cada um destes itens está exposto nas próximas subseções junto uma explicação de quais foram as atitudes tomadas por este trabalho para resolvê-las.

3.1.1.1 Controle de persistência e transações

definir um novo container de execução para automatizar o uso de uma API de persistência. No momento, as transações são controladas de forma manual; (AGUIAR; COSTA, 2015, p. 79)

Esta melhoria trata-se de um detalhe de implementação do SILQ-1 que, devido à sua arquitetura, forçava a criação e gerenciamento manual de transações ao realizar acessos ao banco de dados. A automatização desta tarefa foi rapidamente alcançada no SILQ-2 com a utilização do *Spring Framework*, conforme será relatado na seção 3.2.2.

3.1.1.2 Automatização da atualização dos dados Qualis

as publicações periódicas dos documentos de avaliação do WebQualis levam à necessidade de atualizar o banco de dados periodicamente. A construção de uma funcionalidade automatizada é pertinente; (AGUIAR; COSTA, 2015, p. 79)

Os dados Qualis utilizados pelo sistema para realizar a avaliação de currículos são atualmente inseridos de forma manual. Na segunda etapa deste trabalho será incluído um módulo no sistema onde administradores poderão fazer *upload* de novos dados Qualis, conforme eles forem sendo divulgados e assim não necessitando inserí-los de forma manual no banco de dados.

Os dados Qualis 2014 e 2015, divulgados somente no ano de 2016, entretanto, já foram inseridos de forma manual no banco de dados do SILQ-2, passando por um processo de limpeza de dados descrito na seção 3.2.1.

3.1.1.3 *Fine-tuning* da função de similaridade

desenvolvimento de uma nova função de similaridade ou novos procedimentos de cálculo de similaridades. Foram testadas funções já bem conhecidas pela comunidade científica, com testes concretos para levantar qual delas é a mais indicada no escopo da ferramenta. Porém, é possível que se alcance resultados melhores utilizando uma combinação de funções, ou até mesmo criando uma nova especificamente para o cálculo de similaridade de nomes de periódicos e conferências; (AGUIAR; COSTA, 2015, p. 79)

Será realizado (?) - Hipótese

3.1.1.4 Considerar outras informações do pesquisador

levantamento das outras informações contidas no currículo, como o desempenho do pesquisador antes de ser incluído em um grupo de pós-graduação ou a quantidade e a qualidade de outras atividades acadêmicas; (AGUIAR; COSTA, 2015, p. 79)

Este item envolve o processamento de outras informações contidas no currículo Lattes de um pesquisador e, em um primeiro momento, foge do escopo original do SILQ que é a integração entre o Lattes e o Qualis, utilizando as publicações e informações de onde foram publicadas/apresentadas como item de avaliação.

Por este motivo, este item não foi abordado neste trabalho. Todavia, outros trabalhos de pesquisa relacionados estão paralelamente em desenvolvimento: **pegar informações dos trabalhos em desenvolvimento aqui.**

3.1.1.5 Relação entre orientações e produções associadas

levantamento da relação entre as atividades de orientação, como bolsas de iniciação científica, e as produções acadêmicas associadas; (AGUIAR; COSTA, 2015, p. 80)

Não foi abordado neste trabalho, pelos mesmos motivos do item

anterior. Os trabalhos citados acima também relacionam-se com este item.

3.1.1.6 Busca por informações de eventos

expansão do módulo da busca por informações de eventos. A complexidade do tema impediu o desenvolvimento completo desse módulo na ferramenta SILQ. Procedimentos automáticos de busca utilizando-se web crawlers e o tratamento correto dessas informações fornecem muitos temas para desenvolvimento de módulos novos; (AGUIAR; COSTA, 2015, p. 80)

A inserção de dados do Qualis Eventos¹ no sistema se dá de forma manual, à medida em que novas informações são disponibilizadas pela CAPES. O objetivo deste item seria a criação de um *web crawler* para captura automática de dados de eventos científicos na web. Devido à complexidade deste tema e por se tratar de um assunto independente, não relacionado com o foco deste trabalho, este item não foi abordado.

3.1.1.7 Feedback do usuário

permitir que o usuário auxilie a ferramenta na qualificação, ou seja, no caso dele observar algum matching errado, ele próprio pode sugerir o matching correto; (AGUIAR; COSTA, 2015, p. 80)

Será realizado (?) - Hipótese

3.1.1.8 Período de participação em grupo de pesquisa

indicar o período em que o professor está vinculado a um programa de pós-graduação, permitindo que as avaliações dos currículos também possam ser realizadas considerando-se apenas os períodos nos quais o pesquisador estava vinculado à esse programa. (AGUIAR; COSTA, 2015, p. 80)

¹dados e conceitos relacionados a eventos científicos disponibilizados pela CAPES para algumas áreas específicas

Será realizado (?). Inclusão do campo período ao incluir pesquisador em um grupo.

3.2 SILQ 2

É possível aumentar a acurácia de um sistema (de recomendação/avaliação) baseado em algoritmo de similaridade utilizando feedback do usuário?

Qual o Threshold (nível de similaridade) ideal para o SILQ?

3.2.1 Extração e inserção dos novos dados Qualis

Os dados Qualis de eventos e periódicos foram extraídos pelo trabalho original a partir de documentos em PDF e planilhas. Como relatado em (AGUIAR; COSTA, 2015, p. 50-52), os dados foram extraídos dos PDFs utilizando bibliotecas para manipulação deste formato e passaram por um processo de limpeza. A partir deste processo, aproximadamente 107 mil tuplas foram criadas representando os periódicos do triênio 2010-2012.

No primeiro semestre do ano de 2016, porém, a CAPES disponibilizou novos dados Qualis referentes ao ano de 2013 e 2014, alterando sua periodicidade de avaliação de trimestral para anual. Estes novos dados foram divulgados em formato CSV, junto com os antigos dados do triênio 2010-2012, mas desta vez separados por ano. Estes dados foram obtidos pela plataforma Sucupira (antigo WebQualis): <http://qualis.capes.gov.br/webqualis/>.

Como o conjunto de dados do SILQ-1 contava apenas com dados trimestrais, as tuplas originais não possuíam informação de ano (só podia-se deduzir que os dados faziam parte do triênio, mas não sabia-se qual ano). Desta forma, todos as planilhas divulgadas passaram por um novo processo de extração, incluindo os dados já inseridos no SILQ-1, mas agora considerando a nova informação de ano.

As cinco novas planilhas passaram por um processo de *data cleaning* manual:

1. Substituição de entidades HTML especiais por seu caractere correspondente. Ex.: “&” por “&”;
2. Normalização do campo ISSN para o formato “9999-9999”: alguns registros não possuíam o dígito separador ou omitiam os zeros à esquerda;

3. Correção de ISSNs errôneos: alguns registros possuíam o campo ISSN com dígitos faltando. Ex.: 0034-167 (Revista Brasileira de Enfermagem), cujo número correto seria 0034-**7**167. Estes casos foram tratados um a um e os números corretos identificados através de pesquisas na web.

As planilhas CSV foram então dadas como *input* à base de dados resultando na criação de 339204 tuplas, referentes a cada um dos anos do período divulgado 2012-2015. A Tabela 1 mostra o número de periódicos extraídos para cada ano de avaliação Qualis.

Tabela 1 – Número de periódicos extraídos dos dados Qualis

| Ano | Nº de periódicos extraídos |
|--------------|----------------------------|
| 2010 | 75786 |
| 2011 | 66171 |
| 2012 | 108272 |
| 2013 | 44437 |
| 2014 | 44538 |
| Total | 339204 |

Após esta etapa, verificou-se que cada registro da base de dados antiga do SILQ-1 estava contida nesta nova base de dados. Desta forma, cada registro de periódico agora contém os dados do veículo, incluindo nome e ISSN, conceito atribuído pela CAPES, área de avaliação e ano de avaliação.

A inclusão dos novos registros aumentou o número de publicações passíveis de serem avaliadas pelo sistema. Os dados Qualis de 2015 e 2016 ainda não foram divulgados no momento de escrita deste artigo, mas podem ser incluídos facilmente na base de dados assim que forem disponibilizados pela CAPES.

3.2.2 Visão geral da nova arquitetura

O que mudou. Por que mudou. Figurinha da nova Arquitetura. Seção com descrição superficial das tecnologias e ferramentas.

Novo esquema do banco de dados.

4 CONCLUSÃO

As conclusões devem responder às questões da pesquisa, em relação aos objetivos e hipóteses. Devem ser breves podendo apresentar recomendações e sugestões para trabalhos futuros.

REFERÊNCIAS

AGUIAR, F. N. M. de; COSTA, M. E. **SILQ - Sistema de Integração Lattes Qualis**. Florianópolis: Universidade Federal de Santa Catarina, Biblioteca Universitária, 2015.

CAPES. **Classificação da produção intelectual**. 2015. Disponível em: <<http://www.capes.gov.br/avaliacao/instrumentos-de-apoio/classificacao-da-producao-intelectual>>. Acesso em: 17/11/2015.

CAPES. **História e missão**. 2015. Disponível em: <<http://www.capes.gov.br/historia-e-missao>>. Acesso em: 22/05/2016.

CNPQ. **Módulo Produção Bibliográfica**. 2015. Disponível em: <<http://ajuda.cnpq.br/index.php/Módulo.Produção.Bibliográfica>>. Acesso em: 17/11/2015.

CNPQ. **Sobre a Plataforma Lattes**. 2015. Disponível em: <<http://www.cnpq.br/web/portal-lattes/sobre-a-plataforma>>. Acesso em: 17/11/2015.

CNPQ. **Conselho Nacional de Desenvolvimento Científico e Tecnológico: Institucional**. 2016. Disponível em: <http://cnpq.br/apresentacao_institucional/>. Acesso em: 22/06/2016.

CNPQ. **Histórico**: História do surgimento da plataforma lattes. 2016. Disponível em: <<http://memoria.cnpq.br/web/portal-lattes/historico>>. Acesso em: 31/05/2016.

CRISPIM, B. do A. (Editorial) A importância da publicação de artigos científicos: Uma abordagem para área de ciência animal. **Journal of the Selva Andina Animal Science**, v. 1, n. 1, 2014.

MCTI. **Ministério da Ciência, Tecnologia e Inovação: Institucional**. 2016? Disponível em: <<http://www.mcti.gov.br/institucional>>. Acesso em: 22/06/2016.