



CRICTE 2017

XXVIII Congresso Regional de Iniciação Científica e Tecnológica em Engenharia



MALS: UMA FERRAMENTA PARA EXTRAÇÃO E CLASSIFICAÇÃO AUTOMÁTICA DE PUBLICAÇÕES DO CURRÍCULO LATTES

Matheus H. Rehbein

Acadêmico do Curso Ciência da Computação
Universidade Regional do Noroeste do Estado do RS (UNIJUI)
mrehbein45@gmail.com

Rafael Z. Frantz

Professor do Programa de Pós-Graduação em Modelagem Matemática
Universidade Regional do Noroeste do Estado do RS (UNIJUI)
rzfrantz@unijui.edu.br

Fabricia Roos-Frantz

Professor do Programa de Pós-Graduação em Modelagem Matemática
Universidade Regional do Noroeste do Estado do RS (UNIJUI)
frfrantz@unijui.edu.br

Resumo. Pesquisadores costumam divulgar suas publicações científicas no site de seu grupo de pesquisa. Isso requer um grande esforço não só para criar e manter a lista de publicações atualizadas, mas também por ser de interesse acrescentar informações extras que qualificam a publicação. Tendo em vista que o Currículo Lattes é uma fonte rica e atualizada de informações sobre as publicações, é possível utilizá-lo para realizar a extração das publicações de forma automática, reduzindo o esforço de atualizar a lista de publicações a cada nova publicação. Além disso, é possível obter a qualificação das publicações que estão contidas em outras bases de dados, assim como: Qualis e h-index, essas qualificações irão representar a qualidade da publicação. Este artigo tem como objetivo apresentar uma ferramenta de software que permite, a partir de uma solução de integração de aplicações, extrair de forma automática publicações dos currículos Lattes e enriquece-las com seus respectivos Qualis e h-index, e gerar uma página web com esses dados.

Palavras-chave: Desenvolvimento de Software. Integração de Aplicações. Integração de Sistemas.

1. INTRODUÇÃO

Os Currículos Lattes são uma fonte rica e confiável das atividades acadêmicas e de pesquisa de professores e pesquisadores brasileiros. Os currículos estão disponíveis na Plataforma Lattes, e podem ser exportados à arquivos no formato XML.

Uma das principais informações dos Currículos Lattes são as publicações científicas. Elas são divididas em três principais grupos: publicações em eventos, publicações em periódicos, e livros. Publicações em periódicos e em eventos possuem diversas qualificações que representam sua qualidade, entre elas estão o Qualis e o h-index, que podem ser obtidos a partir do WebQualis e de uma base de dados específica ao h-index.

Geralmente, a atualização de listas de publicações em sites contendo as publicações e qualificações é feita de forma manual. Dessa forma, manter as informações

das publicações atualizadas requer muito esforço. Para realizar este processo automaticamente, é possível utilizar uma solução de integração de aplicações (Zimmermann *et al.* [1]). A integração poderá ser configurada para ter como entrada os currículos. Aplicando rotas, modificadores e transformadores será possível ter como saída uma página web contendo a descrição das publicações, junto com as qualificações.

Este trabalho tem como objetivo desenvolver uma integração entre a Plataforma Lattes, WebQualis e a base de dados que contém o *h-index*, para obtenção e qualificação das publicações de pesquisadores. O produto final será um software com interface gráfica para a utilização do usuário.

O restante deste trabalho se divide nas seguintes seções: A Seção 2 apresenta a metodologia utilizada; já a Seção 3 apresenta a sequência de instruções realizada até a geração de uma página web; a Seção 4 introduz a aplicação com suas principais funcionalidades; por fim, a Seção 5 relata as principais conclusões e as futuras funcionalidades adicionadas a aplicação.

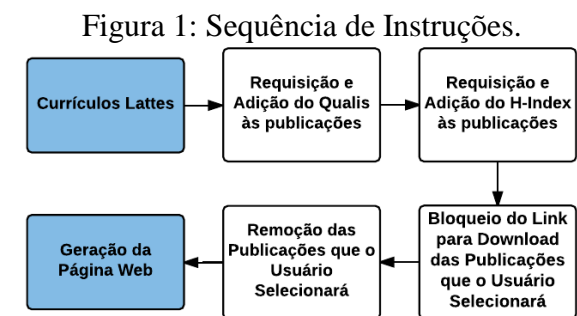
2. METODOLOGIA

A metodologia de trabalho adotada constituiu-se em, primeiramente identificar e documentar o problema que seria abordado neste trabalho. Posteriormente, tornou-se a selecionar as plataformas a serem utilizadas. A solução de integração foi modelada usando a linguagem Guaraná DSL (Frantz *et al* [2]), por ser uma linguagem de alta expressividade e de fácil aprendizado. Já a implementação da solução foi realizada a partir da plataforma Apache Camel (Ibsen *et al* [3]), por apresentar um fácil entendimento durante o desenvolvimento, além disso, pôde-se utilizar a Fluent API (Fowler [4]) para implementar a aplicação. Por fim, o desenvolvimento da interface gráfica foi feito utilizando a linguagem de programação

Java 8 e o ambiente de desenvolvimento Eclipse neon.3.

3. GERAÇÃO DA LISTA

A Figura 1 representa a sequência de processos que a aplicação seguirá até a geração da lista de publicações. Inicialmente a aplicação solicitará ao usuário a entrada dos Currículos Lattes. Em seguida será feita a busca na base de dados do *Qualis* e adicionará a informação ao corpo da mensagem, em um campo específico ao *Qualis*. O mesmo processo anterior será feito com o *h-index*, buscará e o adicionará ao corpo da mensagem. Posteriormente será solicitado ao usuário os trabalhos que ele não desejará liberar o link para download e os trabalhos que ele não desejará publicar na página HTML. Por fim, será gerada a página Web.



Fonte: do autor, 2017.

A requisição do *Qualis* das publicações em periódicos é feita a partir do ISSN, ou seja, será um valor único, diminuindo a possibilidades de erros. O mesmo ocorre à busca do *h-index*. Diferente das outras, a busca do *Qualis* das conferências é feita a partir do nome da conferência. Sendo assim, existe a possibilidade de o usuário colocar a edição do evento antes do nome, que ocasionaria em erros, porém, essa possibilidade foi tratada durante o desenvolvimento. Outro possível erro, é de os usuários colocar, após o nome, a sigla do evento, porém, está possibilidade também foi tratada.

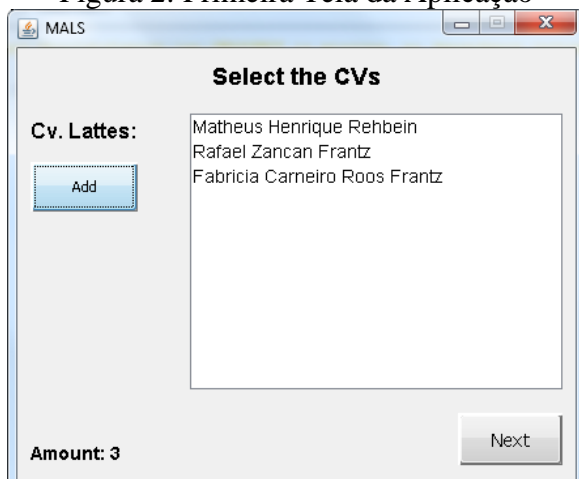
Para solucionar este problema de uma maneira correta, seria o Currículo Lattes ter um identificador de valor único, assim como ISSN, para as conferências. Utilizando esta política, seria possível diminuir gradativamente a possibilidade de erros com os nomes das conferências.

4. APLICAÇÃO PROPOSTA

Nesta seção será apresentado a aplicação desenvolvida para dar suporte a execução das atividades desenvolvidas do processo de geração automática das listas de publicações científicas de pesquisadores.

Ao executar a aplicação, a mesma inicia um processo passo-a-passo, correspondente a uma sequência de interfaces gráficas para auxiliar o usuário durante a geração das listas. A geração das publicações requer três passos do usuário. Sendo o primeiro apresentado pela Fig. 2, onde o usuário deverá selecionar os currículos dos quais deseja extrair as publicações. Nesta tela, uma vez selecionado o arquivo XML correspondente ao currículo Lattes, serão exibidos os nomes completos dos pesquisadores detentores destes currículos.

Figura 2: Primeira Tela da Aplicação



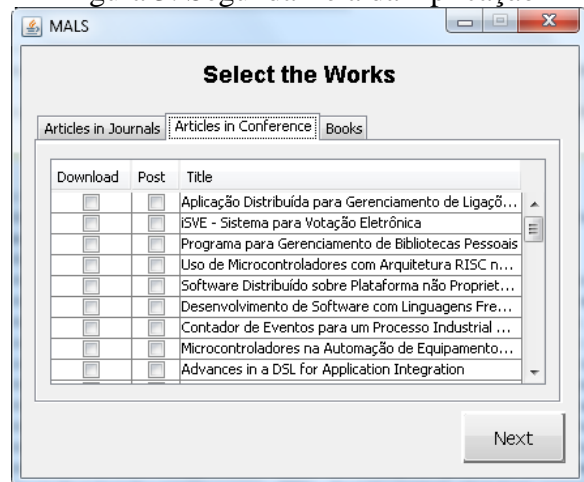
Fonte: do Autor, 2017.

Uma vez selecionado os currículos, a aplicação irá extrair todas as publicações de artigos em periódicos, em conferências e

livros e então apresentará ao usuário em uma próxima interface, conforme a Fig. 3. Nesta tela, o usuário deverá indicar quais publicações farão parte da página web, assim como quais terão link ativo para habilitar o download das mesmas.

Os links de download das publicações seguirão um padrão. Posteriormente, eles deverão ser criados no servidor do site, como o mesmo padrão, para ser possível realizar o download.

Figura 3: Segunda Tela da Aplicação



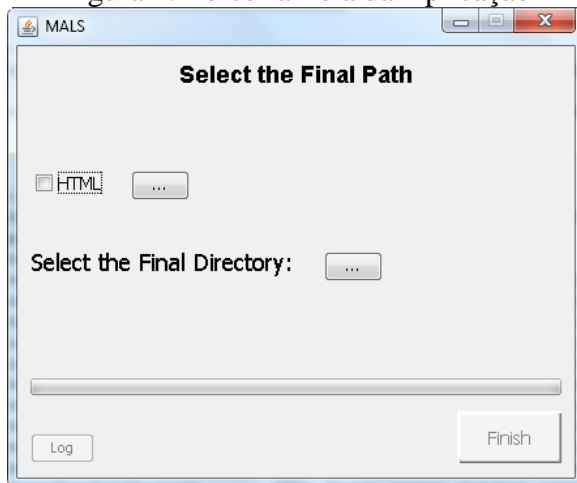
Fonte: do Autor, 2017.

Por fim, a Figura 4 apresenta a última tela da aplicação, nesta tela será iniciado o processo de geração da lista. Além da página web, também será gerado um arquivo no formato XML contendo todas as informações, que facilitará a verificação das publicações. Para iniciar o processo de geração das saídas é necessário selecionar um diretório para salvar os arquivos finais (XML e página Web).

Devido há grande ambiguidade de layouts de site, torna-se necessário uma maneira de modificar a página web gerada de acordo com o qual será postado. Para atender este requisito, deve-se selecionar o *CheckBox* e então clicar no botão ao lado. Assim, será possível editar o template que será gerado. Caso o usuário não edite o arquivo, será gerado em um template HTML padrão, onde não haverá formatações, apenas

uma página em branco com as publicações e qualificações.

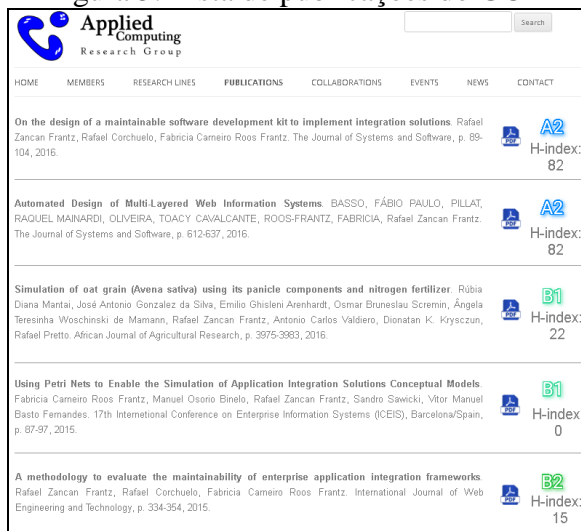
Figura 4: Terceira Tela da Aplicação



Fonte: do Autor, 2017.

A Figura 5 apresenta um exemplo de saída utilizado no Grupo de Pesquisa GCA. Nela é possível identificar os ícones de download, *Qualis* e *h-index* das publicações.

Figura 5: Lista de publicações do GCA



Fonte: do Autor, 2017.

5. CONCLUSÕES

Com o desenvolvimento desta aplicação, verificou-se que a integração entre aplicações se comportou de maneira esperada e auxiliou o desenvolvimento. Abstraiu-se muitos processos que seriam realizados manualmente, como por exemplo,

modificações e transformações do formato do currículo para um formato Web. Além disso, possui um baixo nível de dificuldade para adicionar uma nova funcionalidade à aplicação, como por exemplo, novos tipos de qualificação das publicações.

Como trabalhos futuros, será implantado um novo componente ao Apache Camel, o WordPress. Esse componente permitirá que seja postado automaticamente ao site.

Além do novo componente implantado, será adicionado uma nova funcionalidade ao software, onde será possível identificar diversas informações geradas das publicações. Sendo assim, poderá ser feita uma análise das publicações que o pesquisador realizou, como por exemplo, quantidade de artigos com *Qualis* A1. Para verificar estas informações poderá ser clicado no botão “log” da Fig. 3, ao encerramento do processo de geração.

Agradecimentos

Esta pesquisa foi parcialmente financiada pela FAPERGS por meio de bolsa de iniciação científica.

REFERÊNCIAS

- [1] O. Zimmermann, C. Pautasso, G. Hohpe, B. Woolf, “A Decade of enterprise integration patterns: A conversation with the authors”, *IEEE Software* vol. 33, 2016, 13-19.
- [2] R. Z. Frantz, R. Corchuelo, F. Roos-Frantz, “On the design of a maintainable software development kit to implement integration solutions.”, *Journal of Systems and Software*, vol. 111, 2016, pp 89-104.
- [3] C. Ibsen and J. Anstey, “Camel in Action”, Manning Publications Co., 2010.
- [4] M. Fowler, “Domain-specific languages”, Pearson Education, 2010.