

Evento: XXV SEMINÁRIO DE INICIAÇÃO CIENTÍFICA

**INTEGRAÇÃO DE APLICAÇÕES APLICADA À EXTRAÇÃO E QUALIFICAÇÃO
AUTOMÁTICA DE PUBLICAÇÕES DE PESQUISADORES: UM CASO
BASEADO NO CURRÍCULO LATTES¹**

**APPLICATION INTEGRATION APPLIED TO AUTOMATIC EXTRACTION
AND QUALIFICATION OF RESEARCHER'S PUBLICATIONS: A CASE
BASED ON THE LATTES CURRICULUM**

Matheus H. Rehbein², Rafael Z. Frantz³, Fabricia Roos-Frantz⁴

¹ Pesquisa de Iniciação Científica desenvolvida no Grupo de Pesquisa em Computação Aplicada (GCA) da UNIJUI.

² Bolsista de Iniciação Científica PIBIC/FAPERGS (mrehbein45@gmail.com)

³ Professor Colaborador (rzfrantz@unijui.edu.br)

⁴ Professora Orientadora (frfrantz@unijui.edu.br)

1 Introdução

As empresas necessitam de aplicações para auxiliar e até automatizar seus processos de negócio. Normalmente, essas aplicações são desenvolvidas com propósitos específicos, por diferentes empresas e com diferentes tecnologias. O resultado dessas aplicações é um ecossistema de software heterogêneo composto de aplicações que não foram criadas com o intuito de trabalhar em conjunto. A integração de aplicações empresariais disponibiliza metodologias e ferramentas para modelar e desenvolver integração entre as aplicações heterogêneas, sem modifica-las.

Existem diversas plataformas que dão suporte a integração de aplicações. Entre elas destacam-se: Camel (IBSEN; ANSTEY, 2010), Spring (FISHER et al., 2012), Mule (DOSSOT; D'EMIC; ROMERO, 2014) e Guaraná (FRANTZ; CORCHUELO; ROOS-FRANTZ, 2016). Essas ferramentas utilizam o paradigma de mensagens para realizar integração, o qual consiste em troca de mensagens entre as aplicações. Cada ferramenta tem sua própria linguagem de domínio específico (DSL)(FOWLER, 2010), além de oferecer suporte a diversos padrões de integração, documentados por Hohpe e Woolf (HOHPE; WOOLF, 2004), facilitando o desenvolvimento de integrações.

Um problema presente no dia-a-dia de pesquisadores é a integração entre currículos Lattes, sistemas de qualificações e sites de grupos de pesquisa. É possível resolver este problema utilizando integração de aplicações. Uma solução de integração pode ser projetada para tomar como entrada os currículos Lattes, os quais tornar-se-ão mensagens, aplicar rotas, modificar e transformar essas mensagens para extrair e qualificar as publicações. Por fim, gerar um HTML compatível com o template de grupos de pesquisa.

Este trabalho tem com objetivo modelar e implementar uma solução de integração entre Currículo Lattes, qualificações (Qualis e H-Index) e Sites de grupos de pesquisa utilizando a ferramenta de integração Apache Camel.

O restante deste artigo se divide nas seguintes seções: a Seção 2 descreve a metodologia

Evento: XXV SEMINÁRIO DE INICIAÇÃO CIENTÍFICA

utilizada; a Seção 3 introduz a ferramenta de integração Apache Camel; a Seção 4 apresenta a modelagem e a implementação da solução; e, por último, a Seção 5 relata as principais conclusões.

2 Metodologia

Após realizar revisão bibliográfica, passou-se a estudar as principais ferramentas, verificou-se a facilidade de modelar e entender a solução de integração utilizando Guaraná DSL, portanto, optou-se pela utilização da mesma. Posteriormente, tornou-se necessário realizar a implementação da solução. Observou-se que o Camel possui suporte para as tarefas utilizadas no modelo conceitual e, além disso, a implementação poderia ser feita utilizando a linguagem de programação Java, na qual possibilita o desenvolvimento de interface gráfica para facilitar a usabilidade.

Devido a utilização de arquivos no formato XML tornou-se necessário realizar traduções dos arquivos, graças a este requisito foi necessário o estudo de uma linguagem que faria o tratamento e modificações nos arquivos, chamada de XSLT. Camel oferece suporte para realizar traduções com XSLT, portanto, tornou-se viável a utilização da mesma.

3 Apache Camel

A organização Apache é responsável por diversos projetos Open Source, entre eles está o Camel (IBSEN; ANSTEY, 2010). Camel é um framework de integração entre aplicações empresariais que implementa padrões de integração impostos por Hohpe e Woolf (HOHPE; WOOLF, 2004). Para este trabalho optou-se pela utilização da ferramenta Apache Camel. Ela foi desenvolvida na linguagem java e possui seu código open source. Um diferencial entre o Camel e as demais tecnologias é que ele permite que a integração seja feita com Fluent API Java, XML ou Scala DSL (IBSEN; ANSTEY, 2010).

Camel implementa o paradigma de integração por mensagens, sendo possível aplicar rotas, modificá-las e transformá-las de acordo com a necessidade da integração. Além dos padrões documentados por Hohpe e Woolf (HOHPE; WOOLF, 2004), Camel prove de uma vasta quantidade de Componentes de integração, como por exemplo o componente que é utilizado nos Translators, chamado de XSLT.

4 Caso de Estudo

4.1 Ecossistema de Software

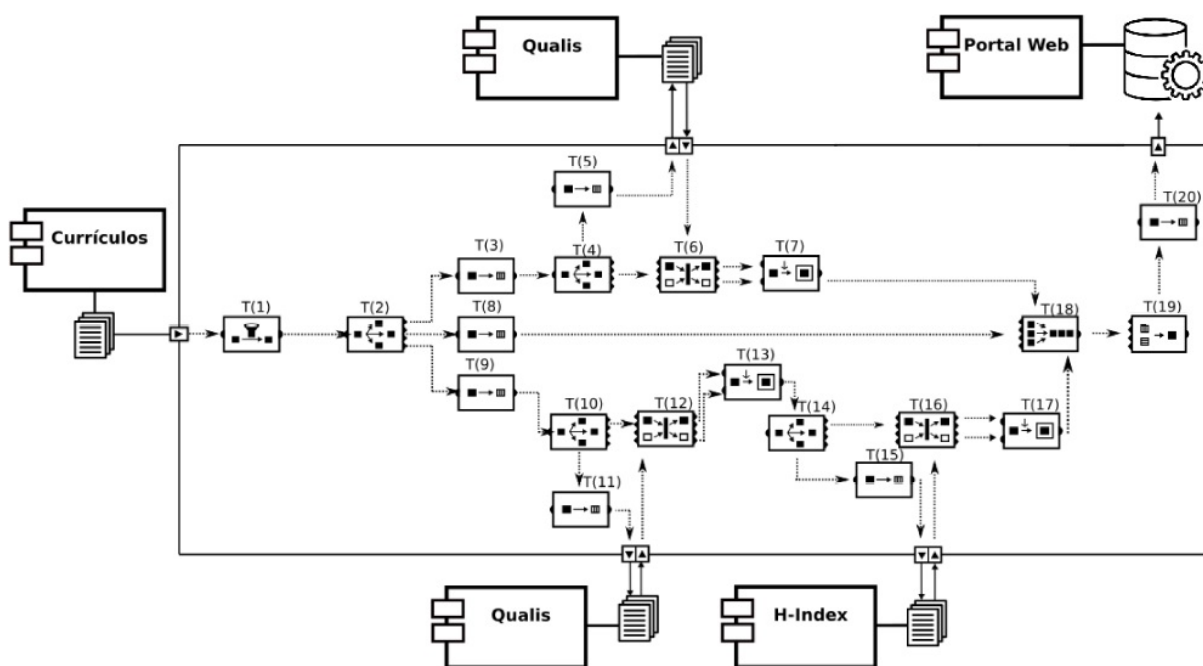
O ecossistema de software é constituído por quatro aplicações: Currículo Lattes, Qualis, H-Index e a saída, que será uma página web. Currículos Lattes é uma fonte de diversas informações de pesquisadores, como por exemplo as publicações. Qualis é uma base de dados que contém informações referente a qualidade de um evento ou periódico. Já o h-index, também uma base de dados, é uma forma de representa a quantidade de citações que o evento ou periódico possui. Por fim, a página web que será gerada como saída da solução.

Evento: XXV SEMINÁRIO DE INICIAÇÃO CIENTÍFICA

4.2 Modelo Conceitual

Para projetar uma solução adequada, é necessário desenvolver a solução por meio de um modelo conceitual. Este modelo descreve os componentes da solução e o fluxo de execução que define as relações entre esses componentes.

Figura 1: Modelo Conceitual da Solução de Integração



Fonte: do autor, 2017.

Na Figura 1 apresenta-se o modelo conceitual da solução. É importante ressaltar que cada currículo representa uma mensagem, portanto, se a aplicação tiver como entrada 5 currículos terá 5 mensagens. A tarefa 1 é conhecida como "Filter", ele é responsável pela filtragem dos currículos, ou seja, se o arquivo de entrada não for um currículo Lattes, esse arquivo não será utilizado. Já a tarefa 2, denominada "Replicator" tem como objetivo encaminhar uma cópia de cada currículo para outros diretórios. Em seguida, as tarefas 3, 8 e 9, "Translator", como o próprio nome já diz, são utilizadas para traduzir um determinado conteúdo. A tarefa 3 será responsável pelos Artigos publicados em Eventos, a tarefa 8 pelos Livros e Capítulos de Livros publicados e a tarefa 9 pelos Artigos publicados em periódicos. Em seguida está a tarefa 4, novamente um Replicator, ela tem como objetivo enviar uma cópia para a tarefa 5 que irá traduzir para uma mensagem que o Qualis irá identificar, e também para a tarefa 6 que é uma "Correlator", a qual correlaciona mensagens. A tarefa 6 irá correlacionar a mensagem enviada pelo Qualis e a mensagem replicada pela tarefa 4, e pôr fim a tarefa 7, conhecida como "Context-based Content Enricher", irá adicionar os conteúdos obtidos na tarefa 6 ao corpo da mensagem. Posteriormente está a tarefa 18, conhecida como "Merger". Esta tarefa irá concatenar as mensagens obtidas das

Evento: XXV SEMINÁRIO DE INICIAÇÃO CIENTÍFICA

tarefas 7, 8 e 17 em uma única mensagem. Em seguida, terá a tarefa 19, chamada de "Assembler", ela irá construir uma nova mensagem a partir das mensagens obtidas na tarefa 18. Por fim estará a tarefa 20, que tem o objetivo de fazer com que seja gerado um HTML que o site irá reconhecer.

O processo executado entre as tarefas 3 e 7 é repetido novamente entre as tarefas 9 e 17, porém irá obter o Qualis e H-Index do artigos publicados em periódicos.

4.3 Implementação

Após realizar o modelo conceitual, foi desenvolvido a implementação da solução utilizando a Fluent API que o Camel disponibiliza. Desenvolvendo uma aplicação Java para implementar os métodos e, então utilizar o kit de desenvolvimento que o Camel disponibiliza.

5 Conclusão

Com o Guaraná tornou-se viável a modelagem conceitual da solução. Após a modelagem foi possível implementá-la com o Camel. O problema de integração é real e está presente no cotidiano de diversos pesquisadores. A solução apresentada para automatizar o processo necessita como entra somente o currículo Lattes do pesquisador e produz como saída uma página HTML. Esta saída pode ser editada para tornar-se compatível com o site em que será disponibilizado. Se posteriormente for desejável adicionar novas funcionalidades à aplicação, será possível, devido a baixa manutenção que a integração proporciona.

Palavras-Chave: Modelagem e Implementação de uma Solução de Integração; Apache Camel; Guaraná DSL.

Agradecimentos

Agradeço a Fapergs pela concessão da bolsa de Iniciação Científica que permitiu o desenvolvimento desta pesquisa. Também agradeço a todos os membros do grupo de pesquisa GCA por todo conhecimento compartilhado.

Referências

- DOSSOT, D.; D'EMIC, J.; ROMERO, V. Mule in action. [S.l.]: Manning, 2014. Citado na página 1.
- FISHER, M. et al. Spring integration in action. [S.l.]: Manning Publications Co., 2012. Citado na página 1.
- FOWLER, M. Domain-specific languages. [S.l.]: Pearson Education, 2010. Citado na página 1.
- FRANTZ, R. Z.; CORCHUELO, R.; ROOS-FRANTZ, F. On the design of a maintainable software development kit to implement integration solutions. Journal of Systems and Software, Elsevier, v. 111, p. 89-104, 2016. Citado na página 1.
- HOPPE, G.; WOLFF, B. Enterprise integration patterns: Designing, building, and deploying

Evento: XXV SEMINÁRIO DE INICIAÇÃO CIENTÍFICA

messaging solutions. [S.l.]: Addison-Wesley Professional, 2004. Citado 2 vezes nas páginas 1 e 2.

IBSEN, C.; ANSTEY, J. Camel in action. [S.l.]: Manning Publications Co., 2010. Citado 2 vezes nas páginas 1 e 2.