

## INTEGRAÇÃO DE SISTEMAS COMO SOLUÇÃO PARA IMPLANTAÇÃO DE UM REPOSITÓRIO INSTITUCIONAL DE ARTIGOS CIENTÍFICOS

CARLOS HENRIQUE FRANCISCO<sup>1</sup>, CRISTIANE YAE MI IMAMURA<sup>2</sup>

<sup>1</sup> Graduando em Engenharia de Software, Centro Universitário das Faculdades de Ensino - UNIFAE, [carloshsjbv@gmail.com](mailto:carloshsjbv@gmail.com).

<sup>2</sup> Mestre em Ciências pela USP-São Carlos. Professora dos curso de Engenharia de Software e Engenharia de Computação da UNIFAE.

Área de conhecimento (Tabela CNPq): 1.03.03.04-9 Sistemas de Informação

Apresentado no

9º Congresso de Inovação, Ciência e Tecnologia do IFSP

11 a 13 de dezembro de 2018 - Boituva-SP, Brasil

**RESUMO:** O crescimento das produções científicas das instituições de ensino e a progressiva adoção de políticas de Acesso Aberto (AA) por diversas bases indexadoras ao redor do mundo, principalmente na América Latina, tornam cada vez mais evidente a importância da adoção de um sistema de preservação e recuperação de dados estruturados provenientes de artigos científicos, nas instituições de ensino superior, visando a posterior manipulação dos dados obtidos e a sua disponibilização na WEB. Nesse sentido, este trabalho propõe o uso de técnicas de integração de sistemas no desenvolvimento de uma aplicação Web para o cadastro de artigos científicos e sua posterior disponibilização online. São demonstradas as soluções propostas para os desafios encontrados na integração de soluções novas com outras existentes neste cenário. Uma das principais integrações neste sistema leva em conta o uso da plataforma Open Typeset Stack, para obter documentos nos formatos HTML, XML e PDF, a partir de documentos científicos estruturados no formato DOCX.

**PALAVRAS-CHAVE:** sistema de publicações científicas; open typeset stack; recuperação de dados.

## SYSTEMS INTEGRATION AS A SOLUTION FOR THE IMPLEMENTATION OF AN INSTITUTIONAL REPOSITORY OF SCIENTIFIC ARTICLES

**ABSTRACT:** The growth of scientific production of educational institutions and the progressive adoption of Open Access (OA) policies by several scientific indexing bases around the world, mainly in Latin America, show the increasingly evident importance of the adoption of a system for preservation and recovery of structured data of scientific articles from higher educational institutions to latter manipulation and make them available on the WEB. In this way, this study proposes the use of systems integration techniques for a development of a scientific articles registration web application and its latter online availability. The solutions to the challenges found in the integration of new and exists solutions in this scenario are presented here. One of the main integrations at this system consider the use of Open Typeset Stack platform to get documents at the formats HTML, XML e PDF, from structured scientific documents at DOCX format.

**KEYWORDS:** scientific publication systems; open typeset stack; data retrieving.

## INTRODUÇÃO

Muitas instituições de ensino possuem suas próprias bases de publicações, denominadas como Repositórios Insitucionais (RIs), onde pesquisadores são encorajados a depositarem seus *preprints* e

*postprints* com o intuito de aumentar a visibilidade dos resultados de suas pesquisas (FURNIVAL & HUBBARD, 2011). Entretanto, para que os artefatos armazenados nessas bases possam ser efetivamente acessados, é necessário que eles sejam estruturados de alguma forma, seja em termos de formatações específicas, ou seja, do tipo de documento.

Objetiva-se através deste trabalho desenvolver uma aplicação web para a publicação de trabalhos de conclusão de curso, de uma instituição de ensino superior. Para isso, são usados aspectos da integração de sistemas que pode ser definida como a “partilha de informações e dados entre aplicações em rede ou fontes de dados numa organização” (MARTINS, 2005), no uso de uma ferramenta auxiliar no processo de conversão de documentos não estruturados disponíveis em formato Microsoft Word para XML (Extensible Markup Language).

## MATERIAL E MÉTODOS

O software que está sendo construído utiliza Web Services e baseia-se na arquitetura REST (Representational State Transfer), que utiliza princípios que fazem a internet escalável (FIELDING & TAYLOR, 2002).

A ferramenta destinada ao recebimento de requisições é denominada *Open Typeset Stack* (OTS) e, segundo GARNETT et al. (2017), é um serviço web open source mantido pela iniciativa *Public Knolegment Project* (PKP). A OTS é voltada à transformação de documentos no formato Microsoft Word para XML-JATS, com regras baseadas em *Data Type Definitions* (DTD) que são definidas como gramáticas que conferem estrutura ao documento XML (ALMEIDA, 2002).

O conceito de integração aplicado a este cenário utiliza requisições Hipertext Transfer Protocol (HTTP), tais requisições são utilizadas para definir como um cliente faz solicitações ao servidor, e como o servidor responde a este mesmo cliente (FARQUHAR et al., 1995).

## RESULTADOS E DISCUSSÃO

Para o desenvolvimento do software, foi estabelecida uma arquitetura para fazer com que o software possua o fluxo de ações, ilustrado na Figura 1, no qual: (i) O usuário submete o arquivo DOCX ao sistema; (ii) O arquivo é enviado via requisição HTTP ao serviço *Open Typeset Stack*, que se encarrega de converter o documento para outros formatos. (iii) Após o processamento, o sistema faz requisições à plataforma para verificar o status do processo de conversão; (iv) Caso os documentos já tenham sido processados, o sistema solicita via requisição HTTP o retorno dos arquivos da plataforma, que em múltiplos formatos são enviados como resposta; (v) Os arquivos vindos da Open Typeset Stack são armazenados no Sistema de Gerenciamento de Banco de Dados (SGBD) e, através de uma rotina interna proporcionada pelo sistema, o artigo científico é finalmente publicado online.

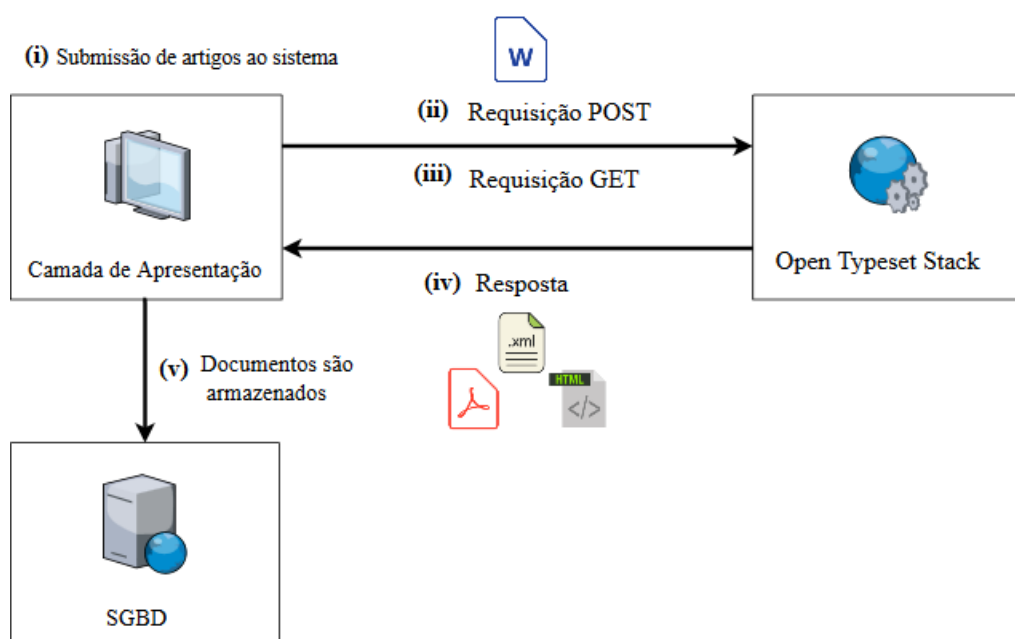


FIGURA 1. Fluxo de ações a ser desempenhado no software.

Embora o software esteja sendo construído em Java, em função de ter sido definida uma arquitetura baseada em REST e que faz uso de Web Services, já se nota desde a fase inicial a transparência no tratamento das requisições e a obtenção das respostas da plataforma OTS, independente da linguagem que ela esteja disponível. Essa transparência tem contribuído muito na promoção da integração desejada.

## CONCLUSÕES

A utilização de conceitos e técnicas de integração de sistemas é importante para a adaptação das organizações a novos padrões, ou para reaproveitamento de trabalhos existentes que, de certa forma, se relaciona com ideias elaboradas em um novo projeto de software, como neste trabalho em andamento.

Inicialmente foi possível vislumbrar a integração de um novo software com a plataforma já existente, chamada Open Typeset Stack (OTS), para desempenhar a importante função de facilitar a divulgação das pesquisas realizadas na instituição. No entanto, o processo de identificação e classificação de elementos textuais em dados bem estruturados, ou metadados, em XML permitem que aplicações futuras possam ser construídas para fornecer dados estatísticos sobre as pesquisas publicadas.

## REFERÊNCIAS

ALMEIDA, M. B. Uma introdução ao XML, sua utilização na Internet e alguns conceitos complementares. *Ci. Inf.*, Brasília, v. 31, n. 2, p. 5-13, Ago. 2002. Disponível em <[http://www.scielo.br/scielo.php?script=sci\\_arttext&pid=S0100-19652002000200001&lng=en&nrm=iso](http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0100-19652002000200001&lng=en&nrm=iso)>. Acesso em 07 mai. 2018.

FARQUHAR, A.; FIKES, R.; PRATT, W.; RICE, J. Collaborative ontology construction for information integration. Palo Alto, Califórnia: Universidade de Stanford. 1995.

FIELDING, R. T; TAYLOR, R.N. Principled design of the modern web architecture. *ACM Transactions on Internet Technology*, v.2, n. 2, p.115–150, 2002.

FURNIVAL, A.; HUBBARD, B. Acesso Aberto às Publicações Científicas: vantagens, políticas e advocacy. *InCID: Revista de Ciência da Informação e Documentação*, v. 2, n. 2, p. 160-177, 12 Dez. 2011.

GARNETT, A.; AUFREITER, M.; BUCHTALA, O.; Alperin, J. P. Introducing Texture: An Open Source WYSIWYG Javascript Editor for JATS. In: *Journal Article Tag Suite Conference (JATS-Con) Proceedings* [Internet]. Bethesda (MD): National Center for Biotechnology Information (US); 2017. Disponível em: < <https://www.ncbi.nlm.nih.gov/books/NBK425544/>>. Acesso em 07 mai. 2018

MARTINS, V. M. M. Integração de Sistemas de Informação: Perspectivas, normas e abordagens. 2005. 218f. Dissertação (Mestrado) – Universidade do Minho, Guimarães, Portugal, 2005.