

Integração de sistemas como solução para implantação de um repositório institucional de artigos científicos

► Abstract

▼ Manuscript

Journal

Academic Publisher

Integração de sistemas como solução para implantação de um repositório institucional de artigos científicos

São João da Boa Vista - SP, 0

Centro Universitário das Faculdades
Associadas de Ensino, UNIFAE

,

São João da Boa Vista - SP

,

Brasil

2018

O crescimento das produções científicas

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 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, essas ações visam a ampla divulgação na Web dos resultados de suas produções científicas e também permitir a posterior manipulação desses dados. 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.

1. Introdução

A tarefa de disseminar conhecimento tem como auxiliares as ferramentas e sistemas que possibilitam armazenar e disponibilizar artefatos que descrevem trabalhos e pesquisas produzidas no meio acadêmico.

Muitas instituições de ensino possuem suas próprias bases de publicações, denominadas como Repositórios Institucionais (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).

1.1. Justificativa

Publicar os resultados alcançados nos trabalhos de conclusão de curso apenas internamente, no âmbito da Instituição de Ensino Superior (IES) onde foram realizados, pode não proporcionar a divulgação e alcance desejados aos projetos apresentados pelos alunos que se encontram em processo de finalização de um curso de graduação. Portanto, a proposta de elaborar um portal de publicações de trabalhos surgiu da necessidade de tornar o conteúdo de tais projetos publicável e de facilitar e ampliar a sua divulgação na rede mundial de computadores, uma vez que se nota o crescimento da mobilidade e da facilidade no acesso às informações.

1.2. Objetivos

Diante da necessidade apresentada, objetiva-se neste trabalho desenvolver uma aplicação web para a publicação de trabalhos de conclusão de curso, desenvolvidos por alunos em ambiente acadêmico. Prevê-se neste sistema que os alunos possam se cadastrar, submeter suas publicações para divulgação na internet, por meio de um portal específico, e através do mesmo efetuar consultas sobre as publicações já disponíveis.

Para viabilizar a construção do sistema web apresentado, é proposto fazer uso de técnicas de integração de aplicações, principalmente para viabilizar a obtenção do arquivo correspondente que será utilizado como base para a transformação do documento originalmente publicado em um formato compatível com as normas estabelecidas pelo órgão regulamentador da internet (W3C - World Wide Web Consortium), para que finalmente possa Centro Universitário das Faculdades Associadas de Ensino, UNIFAE, São João da Boa Vista – SP, Brasil. 2 ser publicado em uma página web onde o trabalho

ficará disponível para visualização e download.

Com essa aplicação, almeja-se colaborar para tornar público os trabalhos realizados por alunos de uma IES, fazendo com que o conhecimento adquirido possa ser cada vez mais difundido.

1.3. Objetivos Específicos

Para que o desenvolvimento da aplicação seja realizado, objetiva-se também:

- Analisar as ferramentas necessárias que poderão ser úteis na conversão dos textos originais para o formato a ser empregado na base de publicações. □
- Implementar as integrações entre os softwares necessários.

1.4. Problematização

Tomando como premissa o conceito de que um trabalho de conclusão de curso é rigorosamente analisado por profissionais da área técnica e científica por diversas vezes antes de ser dado como finalizado, e que também segue à risca os padrões estabelecidos pela ABNT (Associação Brasileira de Normas Técnicas), ou pela própria instituição de ensino, torna-se viável a ideia da construção de um software que reconheça padrões estruturais de texto e trabalhe em função deles para derivar em uma saída com conteúdo manipulável, propício para publicação em diversos formatos.

O cenário atual evidencia que a UNIFAE ainda não dispõe de meios que proporcionem à comunidade em geral, que deseja conhecer mais sobre o potencial científico da instituição, um ambiente virtual de pesquisa que armazene e também possibilite a consulta de trabalhos executados por seus alunos. Logo, presume-se, que a criação de um portal que se encarregue de executar tais papéis, poderia transformar a visibilidade externa da faculdade e também dos trabalhos de seus graduandos.

Portanto o desenvolvimento de um software que faça o intermédio entre o processo de reconhecimento, tratamento, reestruturação de um documento científico para publicação online, e a criação de um portal que armazene este conteúdo, vem de encontro com as necessidades da universidade, que visa, através do conhecimento transmitido na relação Centro Universitário das Faculdades Associadas de Ensino, UNIFAE, São João da Boa Vista – SP, Brasil. 3 aluno-professor, formar profissionais cada vez mais aptos e dispostos a buscar soluções em favor das carências da sociedade.

2. Desenvolvimento

2.1. Referencial Teórico

2.1.1. Integração de aplicações

O conceito de integração de aplicações, definido por MARTINS (2005) como a “partilha de informações e dados entre aplicações em rede ou fontes de dados”, corrobora com a ideia de que com o aumento da quantidade de Sistemas Legados, Web Services, e softwares em geral, que inevitavelmente com o passar do tempo acabam por perder eficiência ou tornam-se obsoletos, muitas corporações tem adotado estratégias que viabilizam a integração de novos sistemas, ao invés da criação de novos softwares que abranjam de forma geral o escopo de um projeto.

Tais estratégias são denominadas EAI (Enterprise Application Integration), que consistem em um conjunto de boas práticas e padrões que auxiliam desenvolvedores e profissionais da área da tecnologia no processo de integração de aplicações.

De acordo com RUH, MAGINNIS & BROWN (2002 , p. 137) a adoção da abordagem EAI, torna-se mais viável para corporações pois a substituição total de um sistema em funcionamento pode ser arriscada, tanto do ponto de vista econômico quanto do ponto de vista técnico, visto que o processo de desenvolvimento exige grande investimento, profissionais capacitados, além de tempo até sua finalização.

Diante disto, as corporações tem optado pelo desenvolvimento de API's (Application Programming Interface), que são responsáveis por implementar novas tecnologias com o intuito de complementar os requisitos não contemplados pela aplicação anterior, além de proporcionar a comunicação entre sistemas distintos, independentes de linguagem de programação ou arquitetura de software utilizadas.

Dentre os diversos padrões e técnicas de integração de aplicações presentes no mercado, não é correto afirmar que existem técnicas melhores, que se sobressaíam ante as demais. Todas possuem suas vantagens e desvantagens, e cabe à equipe responsável pela Centro Universitário das Faculdades Associadas de Ensino, UNIFAE, São João da Boa Vista – SP, Brasil. 4 integração escolher qual delas é mais apropriada para a demanda (SORDI & MARINHO, 2007).

De acordo com as abordagens discutidas por SORDI & MARINHO (2007), os recursos de integração de sistema são divididos em três mecanismos: mensagem (messaging), chamadas (call interface) e transferência de arquivos (data access/file transfer).

Para o desenvolvimento do presente trabalho, utilizou-se o mecanismo de chamadas, aplicado por meio de um web service que através da execução de requisições HTTP, possibilitou o envio e recebimento de dados e arquivos,

bem como a comunicação com a plataforma OTS.

2.1.2. Protocolo HTTP e API's Restful

HRADIL & SKLENÁK (2017) ressalta que grande parte das aplicações desenvolvidas atualmente utilizam a WEB como meio de comunicação com o usuário, uma vez que a necessidade de acesso aos dados utilizando os mais variados tipos de dispositivos tem se tornado cada vez maior.

Dentre as diversas formas de implementação de web services, destaca-se o estilo arquitetural REST (Representational State Transfer) no desenvolvimento desta aplicação, pois proporciona um conjunto padronizado de requisições e respostas enviadas a um servidor, viabilizando assim, uma comunicação clara entre sistemas diferentes. A arquitetura REST se baseia no uso do protocolo HTTP (Hypertext Transfer Protocol), considerado como o protocolo básico de transferência de dados e arquivos de hipermídia.

O protocolo HTTP é dividido em métodos, que permitem a realização de diferentes tipos de ações via requisição. São utilizadas no desenvolvimento da aplicação proposta os seguintes métodos e finalidades: □ POST: gravação de dados; □ GET: consulta de dados; □ PUT: alteração de dados existentes; □ DELETE: remoção de dados do servidor.

2.1.3. Extração de informações de arquivos estruturados

Centro Universitário das Faculdades Associadas de Ensino, UNIFAE, São João da Boa Vista – SP, Brasil. 5

A abordagem da extração de informação de documentos de textos estruturados é um conceito bastante abordado na área da Tecnologia da Informação. Atrelado a este conceito, podemos observar que em grande parte dos casos, faz-se necessário o reconhecimento e extração de conteúdo para a obtenção de dados estruturados, que quando aplicados em determinado cenário, passam a ter maior significado semântico para posterior manipulação e obtenção dos chamados metadados, que por sua vez, resumem-se na ideia de “dado sobre o dado” ou na “catalogação do dado ou descrição do recurso eletrônico.”, de acordo com [\(SOUZA, VENDRUSCULO & MELO, 2000\)](#).

Após serem extraídos, tais metadados passam a ser tratados como estruturas sujeitas a processamento computacional e podem ser utilizados para a criação de conteúdo em diversos formatos, inclusive os voltados à publicação online. Os metadados, segundo [VANTI \(2010\)](#), também são utilizados para a produção de dados estatísticos voltados à medição de índices de produtividade de centros de pesquisa, identificando áreas e cursos que

necessitam de uma abordagem diferenciada, visando o aumento de seu potencial, além de também serem utilizados como ferramenta auxiliar de gestão para a alocação de recursos dentro de uma instituição de ensino.

Este processo de extração faz uso de diversos softwares de automação, que partindo dos documentos de entrada, resultam em arquivos de formatos específicos, bem estruturados e especificamente criados para facilitar o processo de organização e identificação do conteúdo. Dentre os principais formatos de arquivos com tais características, destacam-se o Text Encoding Initiative (TEI) e o Extensible Markup Language (XML).

Os documentos TEI são definidos como uma estrutura de arquivo de texto, utilizada para preservar o conteúdo através da criação de uma estrutura de identificação, que viabiliza a leitura por meio de softwares específicos, voltados ao reconhecimento e extração de textos. A codificação TEI de um texto permite criar um documento com marcações para uso em programas especificamente desenvolvidos para a análise textual, que reconhecem os objetos e seus significados através das tags TEI. (RANKIN, 2015 p. 5) .

Os documentos XML podem ser determinados como uma linguagem de marcação com a qual pode ser especificada a forma dos dados no documento e definições semânticas. “Um arquivo eletrônico XML pode conter, simultaneamente, dados e a descrição da estrutura do documento, através do DTD-Data Type Definitions (gramáticas que conferem estrutura ao documento XML).” (ALMEIDA, 2002, p. 7).

Centro Universitário das Faculdades Associadas de Ensino, UNIFAE, São João da Boa Vista – SP, Brasil. 6

2.1.4. Open Typesetting Stack (OTS)

A API OTS é um serviço WEB open-source, mantido pela iniciativa Public Knowledge Project (PKP)¹, que visa efetuar a conversão de documentos para o formato XML-JATS, de maneira mais intuitiva que outras ferramentas, com abordagem simples, e entrega de uma interface gráfica de fácil usabilidade, conforme Figura 1.

Figura 1. Demonstração da interface de submissões à plataforma Open Typesetting Stack.

Fonte: Open Typesetting Stack².

Segundo GARNETT et al. (2017), esta ferramenta utiliza diversas bibliotecas de acesso aberto (como por exemplo, o software meTypeset) para a realização do processo de conversão de um documento. No processamento realizado pelo software meTypeset, de acordo com GARNETT, ALPERIN & WILLINSKY (2015), são empregadas grandes 1 Iniciativa Open Source desenvolvedora de soluções para auxílio à profissionais no ramo de publicações científicas. Disponível em: <https://pkp.sfu.ca/>.

2 <http://pkp-xml-demo.lib.sfu.ca/>

Centro Universitário das Faculdades Associadas de Ensino, UNIFAE, São João da Boa Vista – SP, Brasil. 7 quantidades de estruturas de dados que utilizam princípios de heurística e aplicação de parseamento baseado em lógica fuzzy, partindo de um documento DOCX, passando pela transformação para o formato TEI e posterior obtenção de um arquivo XML-JATS.

Na Figura 2 é demonstrado o fluxo de processamento da ferramenta meTypeSet, na qual pode ser observado que a entrada pode ser constituída de documentos contendo textos fornecidos a partir de diferentes ferramentas como Microsoft Office Word, ou LibreOffice. A partir dela, conforme mencionado, pode ser obtido um arquivo equivalente no formato XML.

Figura 2. Fluxo de processamento da ferramenta meTypeset.

Fonte: Eve (2015).

A API OTS foi escolhida neste trabalho, pois seu índice de assertividade e compatibilidade dos documentos de saída chega a uma média de 80% em relação a padrões exigidos pela especificação XML-JATS.

2.2. Materiais e Métodos 2.2.1. Front-end

O software proposto foi construído segundo a arquitetura Cliente-Servidor e para a apresentação dos recursos utilizados durante o desenvolvimento, podemos agrupá-los em categorias, para melhor entendimento:

Centro Universitário das Faculdades Associadas de Ensino, UNIFAE, São João da Boa Vista – SP, Brasil. 8

O front-end é parte relacionada à exibição do software em uma interface web, e para o desenvolvimento desta etapa foram utilizados conceitos básicos da área de desenvolvimento web como HTML (HyperText Markup Language) e CSS (Cascading Style Sheets) e Javascript somados à utilização do framework Angular em sua versão 6.1.7. A escolha do framework Angular deu-se devido sua popularidade, performance e especialmente à sua grande comunidade de desenvolvedores ao redor do mundo que destinam parte do seu tempo na criação de soluções e plugins que facilitam a vida dos desenvolvedores que a utilizam. Tais fatores fazem do Angular um dos melhores frameworks

Javascript da atualidade [SESHADRI, S. \(2018\)](#)).

2.2.2. Back-end

Nesta categoria, do lado servidor, foi utilizada a plataforma OTS para conversão dos documentos WORD em XML, HTML ou PDF. Para isso, foi necessária a implementação de web services que permitissem a integração desta plataforma com o software em desenvolvimento. Para a implementação dos web services, foi escolhida a arquitetura REST. Para a construção desta integração e das regras de negócio presentes no back-end foi utilizada a linguagem JAVA versão 8 em conjunto com o framework Spring Boot em sua versão 2.0.4, devido sua facilidade em acelerar o processo de configuração de uma aplicação Spring (ANTONOV, 2018).

2.2.3. Ferramentas estruturais e de apoio

Nesta categoria, serão listadas as ferramentas que tornaram mais fáceis a interação e o gerenciamento do projeto:

- Controle de versões: utilizou-se GIT somado à plataforma de hospedagem de códigos-fonte Bitbucket3.
- Organização de tarefas do projeto: utilizou-se o Trello4, que aliado à metodologia ágil Kanban, pode proporcionar uma excelente visão das tarefas a serem executadas.

3 <https://bitbucket.org>

4 <https://trello.com/>

Centro Universitário das Faculdades Associadas de Ensino, UNIFAE, São João da Boa Vista – SP, Brasil. 9 □ □

Armazenamento de dados: utilizou-se o SGBD Oracle Database Express Edition 11g Release 2, que proporciona ao desenvolvedor de forma gratuita facilidade na administração, manutenção5.

Armazenamento dos documentos: utilizou-se o sistema de arquivos do ambiente em que o sistema está hospedado. A estratégia utilizada visa a economia no armazenamento de dados, visto que a versão de SGBD é uma versão gratuita e possui um limite de armazenamento, portanto, foram utilizados campos VARCHAR(2) destinados ao armazenamento de referências ao diretório em que o arquivo foi armazenado.

3. Resultados e Discussões

Como resultado deste trabalho, foi construído um software que atende aos seguintes requisitos funcionais: □ Cadastro de alunos; □ Cadastro de cursos e turmas; □ Cadastro de submissões de documentos referentes a artigos ou trabalhos acadêmicos; □ Listagem de submissões de documentos por curso e turma; □ Listagem individual de um documento submetido, em uma página HTML, contendo as informações básicas de um artigo (título, autores, resumo e palavras-chave), bem como links para download do documento na versão PDF.

A principal integração implementada neste sistema foi realizada com o desenvolvimento de um web service para permitir que o usuário realize a submissão de documentos científicos estruturados no formato Word e que o software web obtenha como retorno, desta requisição, arquivos nos formatos HTML, XML e PDF que são produzidos pela plataforma OTS. Tais documentos são posteriormente disponibilizados online pelo software web, conforme ilustrado na Figura 3. Centro Universitário das Faculdades Associadas de Ensino, UNIFAE, São João da Boa Vista – SP, Brasil. 10

Figura 3. Diagrama de fluxo de arquivos no processo de submissão.

O processo de submissão dá-se da seguinte forma: i. O usuário submete o arquivo Word ao sistema. ii. A submissão é armazenada em uma fila temporária no sistema, que ao fim do dia através de um scheduler de tarefas, compõe o lote a ser enviado a plataforma OTS. Os arquivos são enviados um a um via requisição HTTP (método POST) à plataforma OTS, que se encarrega de converter o documento para outros formatos. iii. Devido a uma convenção adotada no desenvolvimento deste projeto, três horas após a submissão, o sistema faz requisições à plataforma (também através da utilização de um scheduler de tarefas), para verificar o status do processo de conversão. Tais status estão disponíveis para consulta na documentação do OTS no Github⁶. iv. Caso os documentos já tenham sido processados, o sistema solicita via requisição HTTP o retorno dos arquivos da plataforma, que os envia como resposta.

6 <https://github.com/pkp/ots>

Centro Universitário das Faculdades Associadas de Ensino, UNIFAE, São João da Boa Vista – SP, Brasil. 11 v. Os arquivos vindos do OTS são armazenados no sistema de arquivos e Sistema de Gerenciamento de Banco de Dados (SGBD).

Para a execução do processo de envio dos arquivos via requisição HTTP, utilizou-se a implementação de uma fila de submissões, onde os arquivos submetidos durante o dia são enfileirados, e devido a uma convenção adotada na implementação do projeto, definiu-se que

seus itens seriam enviados em uma requisição HTTP POST em determinado horário, e que após três horas o sistema faria outra requisição HTTP, para obter os arquivos anteriormente enviados (GET). Tal implementação tornou-se possível através da utilização de um campo do tipo boolean pré definido na estrutura da tabela de submissões, que assumiu o papel de flag, possibilitando a indicação se o arquivo havia ou não sido enviado à plataforma.

As requisições HTTP GET utilizaram como parâmetro de identificação o campo OTS_ID presente na tabela de submissões. Este campo corresponde ao identificador recebido por cada publicação no momento em que a requisição é recebida pela plataforma OTS.

Ao término do desenvolvimento deste trabalho, foi possível obter a integração de um novo software com a plataforma OTS, inicialmente vislumbrada para desempenhar a importante função de facilitar a divulgação das pesquisas realizadas na instituição, e a obtenção dos seguintes resultados: □ Realização de requisições HTTP com o método POST à plataforma OTS, permitindo com que o usuário submeta arquivos no formato Word e PDF. □ Realização de requisições HTTP com o método GET à plataforma OTS, viabilizando a consulta ao status de processamento dos arquivos submetidos e também a obtenção dos arquivos processados. □ Implementação de uma interface voltada ao repositório de artigos de uma instituição, que permite o cadastro de alunos, cursos, turmas e submissões de artigos, e possibilita ao público o acesso ao conteúdo científico submetido.

Diante dos resultados obtidos, constatou-se que a plataforma OTS cumpre a proposta definida por seus criadores no que tange ao cenário de transformação de um arquivo Word ou PDF em formatos diversos. Entretanto, é importante ressaltar que a qualidade obtida nos documentos de saída, vindos da plataforma, depende diretamente do arquivo fonte, e da forma pelo qual é estruturado e, embora tenha uma assertividade de 80%, algumas vezes o OTS produz resultados que não são suficientemente satisfatórios e que precisam de uma solução supervisionada ao término de seu trabalho, conforme mencionado em GARNETT et al. (2017) Centro Universitário das Faculdades Associadas de Ensino, UNIFAE, São João da Boa Vista – SP, Brasil. 12

4. Considerações Finais

O desenvolvimento do presente estudo possibilitou a criação de uma aplicação voltada à preservação de artigos científicos desenvolvidos por uma instituição de ensino, através da utilização de princípios de integração de

software, proporcionando à comunidade científica e ao público em geral, acesso ao conteúdo produzido por alunos de graduação, com o intuito da difusão do conteúdo científico e aumento da visibilidade da instituição frente à sociedade.

O presente trabalho possui sua aplicação voltada ao âmbito acadêmico, no entanto, também possibilita sua utilização como ferramenta de apoio no desenvolvimento de soluções que utilizem como fonte a plataforma OTS.

Para futuras implementações, será considerada a inclusão das seguintes funcionalidades:

- Melhoria na estruturação do banco de dados e estudo da viabilidade de uma possível migração para a tecnologia NOSQL, que facilita aspectos relacionados à busca rápida de dados e é comumente utilizada para a recuperação de um grande fluxo de dados.
- Implementação de um dashboard com dados estatísticos à respeito da quantidade de artigos publicados por curso, quais cursos precisam aumentar seu potencial científico em relação aos demais, dentre outros aspectos que auxiliem os gestores de uma instituição de ensino superior a impulsionarem a produção de material científico.

- Elaborar sistema de busca de artigos por título, palavras-chave e autores.

5. 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, Aug. 2002. Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S010019652002000200001&lng=en&nrm=iso>. Acesso em: 07 de Maio de 2018. <http://dx.doi.org/10.1590/S0100-19652002000200001>.

ANTONOV, A. Spring Boot Cookbook. Birmingham : Packt Publishing, 2018.

Centro Universitário das Faculdades Associadas de Ensino, UNIFAE, São João da Boa Vista – SP, Brasil. 13 institucional de artigos científicos

FURNIVAL, A. C.; 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. et al. Introducing Texture: An Open Source WYSIWYG Javascript Editor for JATS. In: JOURNAL ARTICLE TAG SUITE CONFERENCE (JATS-Con), 2017, Bethesda. Proceedings [Internet]. Bethesda (MD): National Center for Biotechnology Information (US), 2017. Disponível em: <<https://www.ncbi.nlm.nih.gov/books/NBK425544/>>.

GARNETT, A.; ALPERIN, J. P.; WILLINSKY, J. The Public Knowledge Project XML Publishing Service and meTypeset: Don't call it "Yet Another Word-to-JATS Conversion Kit". In: JOURNAL ARTICLE TAG SUITE CONFERENCE (JATS-Con),

2015, Bethesda. Proceedings [Internet]. Bethesda (MD): National Center for Biotechnology Information (US), 2015. Disponível em:
<<https://www.ncbi.nlm.nih.gov/books/NBK279666/>>.

HRADIL, J.; SKLENÁK, V. Practical Implementation of 10 Rules for Writing REST APIs. *Journal of Systems Integration*, v. 8, n. 1, p. 45-54, 2017.

➤ References
