

Estudo e desenvolvimento de ferramenta de migração entre sistemas de publicação de teses e dissertações

Diego José Macêdo

Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT)

diegomacedo@ibict.br

Washington L. R. de Carvalho Segundo

Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT)

washingonsegundo@ibict.br

Tainá Batista de Assis

Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT)

taina@ibict.br

Milton Shintaku

Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT)

shintaku@ibict.br

Resumo

A evolução tecnológica requer que sistemas web que disponibilizam informações estejam alinhadas com as tendências, ofertando serviços mais condizentes com as necessidades dos usuários. Neste contexto, a Biblioteca Digital Brasileira de Teses e Dissertações (BDTD), que possui mais de 10 anos, iniciou por meio de uma pesquisa aplicada, uma atualização tecnológica em sua estrutura, culminando no desenvolvimento de uma ferramenta de migração entre sistemas de publicação de teses e dissertações em forma de Plug-in. Isso possibilita que esses documentos possam ser migrados, sem a perda de informação, de um sistema informatizado antigo, no caso o Sistema de Publicação Eletrônica de Teses e Dissertações (TEDE), para sistemas que utilizam outros softwares como o DSpace que se aplica aos repositórios institucionais. Desta maneira, possibilita à comunidade acadêmica brasileira o processo de disseminação de teses e dissertações eletrônicas no formato digital.

Palavras-chave: Migração, BDTD, Repositório Institucional, TEDE, DSpace.

Introdução

Em 2013, a Biblioteca Digital Brasileira de Teses e Dissertações (BDTD) agregava 97 bibliotecas digitais, sendo 81 desenvolvidas com o Sistema de Publicação Eletrônica de Teses e Dissertações (TEDE). No início de 2014, a atualização tecnológica da BDTD e a adoção de um novo sistema coletador, possibilitou a flexibilidade no processo de coleta. Assim, houve o agrupamento de um total de 104 bibliotecas à rede, sendo 83 implementadas em TEDE, 14 baseadas em DSpace e 7 com tecnologias diversas. Revela-se uma tendência na adoção do DSpace para depósitos de teses e dissertações (TDs), principalmente, quando se considera a implantação de repositórios institucionais (RI) nas universidades, como relataram Murakami e Fausto (2013). Desta forma, as teses e dissertações passaram também a integrar os repositórios institucionais construídos em DSpace. Nesse ponto, alinhando-se às tendências e com o compromisso de dar apoio às ações de disseminação da produção científica em acesso aberto, o Ibict desenvolveu um plug-in de migração de dados do TEDE, com formato de saída específico para o DSpace.

Objetivo(s)

O presente estudo apresenta o processo que culminou na construção do plug-in, objetivando, com isso, contribuir com discussões sobre conversão automática de metadados, normalização de metadados, migração de dados entre sistemas e adaptação de ferramentas, entre tantos tópicos afeitos à Ciência da Informação.

Metodologia

Este trabalho se alinha, portanto, ao que Pereira (1991, p.10) descreve como pesquisa prática, que tem por objetivo solucionar problemas de forma pragmática. O manual Frascatti (OCDE) destaca que a pesquisa aplicada pode ter como resultado produtos, processos ou sistemas que devem possuir característica estratégica. Alinhando-se a essa diretiva, o desenvolvimento de um software que possibilita a migração de dados entre sistemas distintos, TEDE e DSpace, é visto como um produto resultante de um estudo, e para tanto, foi estruturado em quatro etapas: levantamento dos requisitos, seleção do software, processo de ajuste e avaliação por meio de testes de execução. A conectividade entre sistemas dá-se por várias formas, sendo a manual a mais trivialmente implementável e, a automática, a que mais fácil é aceita pelos usuários. No entanto, a forma requer ser transparente às questões de migração. O estudo apresentou três possibilidades, duas automáticas (harvesting com OAI-ORE ou webservice) e uma manual (plugin de migração).

A adoção pelo processo manual deu-se depois da análise das opções, pelos seguintes motivos:

- As opções automáticas deixariam os arquivos integrais de forma aberta e qualquer provedor de dados poderia coletá-los, logo as instituições costumam apresentar restrições a esse procedimento;
- A execução manual permite processamento de dados durante a migração, possibilitando gerar conteúdo para outros campos, pela manipulação dos campos já existentes;
- O processamento manual gera arquivos para migração, que podem ser armazenados como forma de preservação ou posterior processamento.

Resultado(s) e discussão

Destacam-se assim os seguintes pontos nos resultados de desenvolvimento:

- A exportação dá-se organizando os itens exportados em pastas por Programa de Pós-graduação e grau;
- O conteúdo do campo citação é gerado com base nos campos existentes no banco, conforme norma ABNT (Associação Brasileira de Normas Técnicas) 6023;
- O conteúdo do campo assunto, quando se utilizado o vocabulário do CNPq, é recuperado de forma completa, com os valores intermediários;
- Há um algoritmo de correção de caracteres inválidos, os quais podem ter sido inseridos no processo de catalogação;
- Ocorre a normalização de campos como grau e padronização do campo de autor, conforme ABNT.

Os testes para refinamento do plug-in foram executados em diversas bases e com a exportação de mais de dez mil registros, sendo que menos de dez por cento apresentam problemas que não podem ser resolvidos automaticamente e são armazenados em pastas de triagem separadas, para posterior verificação e correção manual.

Conclusão

Teses e dissertações são as únicas produções científicas que são de total responsabilidade da instituição. Prescindem da figura do editor, e sendo, portanto, em geral, os tipos de documentos mais facilmente agregados aos repositórios institucionais e a outros sistemas de informação. Nesse caso, o desenvolvimento do plug-in fortalece a ação do Ibtic junto aos RIs e a atualização tecnológica das bibliotecas digitais, revelando a importância da disseminação das TDs em texto integral em seus formatos digitais. Apresenta-se a relevância de estudos aplicados que resultam em produto, visando o atendimento de uma

comunidade e alinhando-se às tendências de tecnologias de acesso aberto no contexto institucional acadêmico.

Referências

PEREIRA, K. L. A. (1991) – *A pesquisa em Música e Educação*. São Paulo: Edições Loyola. 121 p.

ORGANIZAÇÃO PARA A COOPERAÇÃO E DESENVOLVIMENTO ECONÔMICO (2007) – *Manual de Frascati*: metodologia proposta para a definição da investigação e desenvolvimento experimental. Coimbra: F-Iniciativas. 336 p. Disponível na Internet: <<http://www.uesc.br/nucleos/nit/manualfrascati.pdf>>.

MURAKAMI, Tiago Rodrigo Marçal; FAUSTO, Sibebe (2013) – Panorama atual dos Repositórios Institucionais das Instituições de Ensino Superior no Brasil. *InCID: Revista de Ciência da Informação e Documentação* [Em linha]. Vol. 4, Nº 2. [Consult. 26 Maio 2013]. Disponível na Internet: <<http://www.revistas.usp.br/incid/article/view/69327>>. ISSN 2178-2075