

Interoperabilidade em Repositórios Digitais: modelo de provedor de serviços interativo

Caio Saraiva Coneglian

Universidade Estadual Paulista Júlio Mesquita Filho – UNESP, Email: caio.coneglian@gmail.com

José Eduardo Santarem Segundo

Universidade de São Paulo – USP, Email: santarem@usp.br

RESUMO

O aumento do uso das Tecnologias da Informação e Comunicação, principalmente no que tange a Web, proporcionou a maior utilização de formatos digitais para a disponibilização de textos científicos. Nesse contexto, surgiram e vem se destacando a utilização de repositórios digitais como meio de arquivamento e de acesso das produções científicas temáticas e institucionais. Diversos protocolos e meios de acesso estão possibilitando uma melhor utilização dos repositórios em cenários distintos. A interoperabilidade e o protocolo OAI-PMH, vem colaborando em estabelecer meios de recuperar os documentos de diversas formas, mais eficientemente. Contudo, questiona-se como criar um sistema que possibilite o usuário a escolher as fontes informacionais que ele deseja utilizar em um sistema interoperável de repositórios digitais. Assim, esta pesquisa tem o objetivo de criar e de implementar um modelo interoperável de repositórios digitais, que possibilite o usuário cadastrar e escolher as fontes informacionais que serão utilizadas em suas buscas. A metodologia utilizada foi do tipo exploratória e aplicada, em que a partir da bibliografia de repositórios e interoperabilidade, criou-se e implementou-se um modelo interativo de repositórios digitais. Com esse trabalho, foi possível verificar que é possível desenvolver sistema interoperável, que permita o cadastro das fontes informacionais em tempo real, fornecendo ao usuário o papel de escolher as fontes informacionais que ele deseja utilizar. Com o aporte da tecnologia Apache Solr, os resultados obtidos de diversos repositórios foram mesclados, seguindo um sistema de relevância interno deste software. O sistema criado aproxima o usuário do programa computacional, permitindo que as necessidades informacionais dos usuários sejam melhores atendidas.

Palavras-chave: Repositórios digitais. Interoperabilidade. Protocolo OAI-PMH.

1 INTRODUÇÃO

O advento da Web viabilizou que pesquisadores disponibilizassem e acessassem às produções científicas de diversos locais do planeta. Paralelamente à expansão da Internet, houve um grande avanço tecnológico que permitiu que os computadores obtivessem um processamento automático dos dados com muita agilidade, além de um aumento substancial na capacidade de armazenamento digital das informações. Uma consequência dessa conjuntura, foi a ampla difusão e utilização de recursos informacionais digitais, possibilitando que atualmente, a maioria dos conhecimentos produzidos estejam disponíveis em meios digitais.

Nessa perspectiva, a colaboração científica tem ocorrido mais ativamente, aproveitando a evolução das Tecnologias da Informação e Comunicação sucedida no final do século XX e no começo do século XXI. A criação e expansão da Web foi o fato de maior impacto e que mais contribuiu para a ampliação da colaboração entre os pesquisadores, pois permitiu a formação de comunidades virtuais na Internet, além de promover uma maior facilidade no acesso às produções de cientistas de outras instituições e de outros países (BALANCIERI et al., 2005).

Neste cenário, os repositórios digitais vêm ganhando um notável destaque, pela facilidade com que estes promovem na difusão da produção intelectual, tanto de organizações, quanto de uma temática específica. Os repositórios digitais possibilitam o arquivamento de documentos e estão apresentando uma eminente notoriedade, devido as iniciativas de acesso aberto, em que os objetos digitais arquivados nestes sistemas, podem ser acessados por qualquer usuário sem qualquer custo (SANTAREM SEGUNDO, 2010).

Para expandir o alcance e o compartilhamento dos objetos digitais contidos dentro de um repositório, esta tecnologia foi construída com o intuito de ser interoperável. Assim, torna-se possível que alguns serviços sejam capazes de coletar, reunir e fornecer o acesso a objetos digitais pertencentes a diversos repositórios em um único ambiente.

Contudo, grande parte das implantações de repositórios digitais estão focadas em repositórios institucionais, ou seja, aqueles que reúnem a produção intelectual de uma instituição, tornando-se uma fonte de informação bastante importante, porém limitada, pois o usuário tem acesso somente a produção daquela instituição em questão.

Da mesma forma, os serviços que reúnem a produção de diversos repositórios, por meio da interoperabilidade, fixam-se nas normativas de suas implementações, em que são reunidos conjuntos limitados de repositórios que estarão presentes nestes serviços. Assim, caso um usuário deseje realizar uma busca em um conjunto de repositórios escolhidos por ele, as buscas deverão ser realizadas isoladamente em cada repositório, ou encontrar uma implementação de um serviço que reúna exatamente o conjunto escolhido, sendo esta opção, por vezes, inexistente.

Assim, o objetivo dessa pesquisa consiste na criação e implementação de um modelo para interoperabilidade de repositórios digitais, que proporcione aos usuários a escolha das fontes informacionais utilizadas durante o processo de busca, bem como a possibilidade de cadastrar novos repositórios que não estejam disponíveis no momento da realização da busca.

Para tal, a metodologia utilizada tem natureza qualitativa com caráter exploratório, ao buscar na literatura conceitos acerca dos assuntos tratados na pesquisa, e aplicada, pois realiza a implementação de um modelo teórico proposto no trabalho. A pesquisa foi dividida em quatro fases: revisão bibliográfica sobre repositórios digitais, interoperabilidade entre repositórios e padrão OAI-PMH; posteriormente, a partir da revisão bibliográfica, foi construído um modelo de interoperabilidade de repositórios digitais interativo; e por fim, foi realizada a implementação do modelo proposto, com o intuito de provar a viabilidade da proposta.

2 REPOSITÓRIOS DIGITAIS

As últimas décadas têm provocado um aumento substancial da produção científica, principalmente a partir da segunda metade do século XX. Dentre os motivos que causaram tal cenário, destaca-se o acesso aberto às produções, que contribuiu diretamente para este aumento (SANTAREM SEGUNDO, 2010).

Com o crescimento exponencial das publicações científicas, a recuperação de documentos que satisfaçam as necessidades dos indivíduos, passou a ser significativamente mais complexa.

Na busca de uma solução para este problema, os Repositórios Digitais foram criados para armazenar objetos digitais, especialmente de produções científicas. O Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT) define repositórios digitais como bases de dados que reúnem de uma maneira organizada a produção científica de uma instituição ou de uma área temática (IBICT, 2015).

Esta definição foi sendo aprimorada, e hoje, repositórios digitais tratam de sistemas de informação que se relacionam com o usuário, possibilitando existir uma interação entre os próprios usuários e entre estes e o sistema (SAYÃO et al., 2009).

Neste contexto, os repositórios digitais foram classificados em três tipos: 1) Repositórios institucionais: focados na produção intelectual de uma instituição; 2) repositórios temáticos ou disciplinares: voltados para comunidades científicas específicas, ou seja, da produção intelectual de áreas do conhecimento em particular e; 3) repositórios de teses e dissertações: tratando exclusivamente de teses e dissertações de cursos de pós-graduação (LEITE et al., 2009).

Os repositórios digitais como um todo se tornaram muito populares entre instituições e organizações com intuito de promover o livre acesso aos materiais científicos, disponibilizando a

todos as produções intelectuais produzidas por estas, ou as produções acerca de uma temática específica. A popularização dos repositórios digitais vai ao encontro de um conceito chamado de *Open Access Initiative* (OAI), que tem como premissa promover o acesso livre e irrestrito à literatura científica e acadêmica (SANTAREM SEGUNDO, 2010).

Outro ponto relacionado ao uso de repositórios trata da questão da facilidade do acesso à produção científico, pois com o uso dos repositórios, o acesso aos objetos digitais se torna mais fácil, aumentando a visibilidade das instituições e dos autores. Como consequência, há um aumento na oferta de acesso às produções científicas, fornecendo material para à produção de novos conhecimentos (SHINTAKU, 2015).

A utilização de repositórios digitais como ambientes colaborativos também pode contribuir para o acesso e participação de algumas comunidades em um ambiente digital. Os sistemas de repositórios podem ser desenvolvidos conforme as necessidades de cada usuário, conseguindo incluir públicos diversos. Estes ambientes colaborativos de repositórios digitais conseguem assim, ser uma ferramenta de inclusão, estimulando a inteligência coletiva com o intuito de construir uma rede de comunicação para a construção de novos conhecimentos (FERREIRA; VECHIATO; VIDOTTI, 2008).

São diversas as funcionalidades que os repositórios digitais podem ser inseridos, e, assim, contribuir para a transmissão e aquisição de conhecimentos. Nesse contexto, a utilização de repositórios digitais vem crescendo, sendo uma forma de divulgar produções científicas e de ser uma fonte de informação muito valiosa para as instituições, para grupos científicos e comunidades em geral, promovendo o livre acesso a informação.

Uma característica fundamental dos repositórios digitais, diz respeito a possibilidade de interoperabilidade que estes possibilitam. Tal característica apresenta grande importância pelo alto número de repositórios digitais existentes, em que meios que promovam a comunicação entre estas ferramentas podem auxiliar bastante aos usuários, no encontro aos objetos digitais que atendam suas necessidades informacionais. Na próxima seção será tratado com maiores detalhes sobre esta questão da interoperabilidade em repositórios digitais.

3 INTEROPERABILIDADE EM REPOSITÓRIOS DIGITAIS

Os repositórios digitais vêm ganhando um papel de destaque em questões relacionadas ao armazenamento e à preservação de documentos. Dessa forma, o processo de Recuperação da Informação dentro destes ambientes, necessitam de estudos e de novas ferramentas, para que o usuário consiga acessar aos documentos com maior eficiência.

A recuperação de um documento, contido em um repositório digital, funciona fundamentalmente por meio do acesso aos metadados descritivos desse documento, em que agentes computacionais acessam a tais metadados, com o intuito de verificar quais documentos atendem as necessidades informacionais do usuário.

Um meio de aprimorar os processos de recuperação em Repositórios Digitais perpassa, inevitavelmente, pela interoperabilidade, que diz respeito a comunicação entre repositórios distintos. Essa característica de interoperabilidade, possibilitou, entre outras coisas, que os usuários acessem a objetos digitais de diversos repositórios em um único ambiente.

Existem diversas ferramentas e protocolos para a realização da interoperabilidade de registros bibliográficos. Basicamente, estas formas podem ser divididas em duas alternativas: 1) busca descentralizada em diversos servidores; 2) busca em uma base de metadados centralizada. Destaca-se que, as duas opções promovem que o usuário interaja em uma única interface (MARCONDES; SAYÃO, 2001).

Na primeira alternativa, o sistema irá realizar consultas em diversas bases de dados que tenham uma interface capaz de comunicar com o sistema. Um exemplo é o protocolo Z39.50 que realiza buscas em catálogos automatizados de bibliotecas (MARCONDES; SAYÃO, 2001).

Este protocolo, foi proposto em 1984 pela *National Information Standards Organization* (NISO), com o intuito de criar um padrão para pesquisas em banco de dados bibliográficos. O Z39.50 é um protocolo que permite a comunicação entre uma máquina, chamada de cliente, que realiza chamadas para uma outra, chamada de servidor. O cliente, é assim chamado, pois solicita informações de um outro sistema, e é servido com tais dados. O servidor realiza o processamento solicitado pelo cliente. O protocolo Z39.50 define procedimentos e estruturas para a comunicação entre cliente e servidor, cujo o objetivo é a recuperação de registros em banco de dados identificados em uma busca (BORDIN, 2002).

A segunda alternativa trata de uma base de dados contendo um conjunto de metadados de diversas fontes. Uma maneira, mais recente que tem sido bastante utilizada, é o processo de *harvesting* (coleta automática de metadados), em que os metadados são coletados automaticamente e periodicamente e as buscas são realizadas em um ambiente integrado. Um dos padrões mais utilizados e estudos do processo de *harvesting* é o *Open Archives Initiative Protocol for Metadata Harvesting* (OAI-PMH). (MARCONDES; SAYÃO, 2001).

3.1 PROTOCOLO OAI-PMH

A iniciativa de *Open Archives* (Arquivos Abertos) surgiu em 1999 nos Estados Unidos da América, e tinha o objetivo de desenvolver formas de interoperabilidade, com o intuito de promover a disseminação mais eficiente das informações (LAGOZE; VAN DE SOMPEL, 2001).

A base desta iniciativa é o protocolo OAI-PMH, que permite o compartilhamento de metadados. Este protocolo utiliza o Dublin Core, como padrão de metadados para a troca de informações¹. Tal padrão é composto de quinze elementos para a descrição dos recursos. A versão inicial do OAI-PMH foi lançada em 2001, e a versão 2.0, que é a última versão disponibilizada, foi lançada em julho de 2002 (GARCIA; SUNYE, 2003).

Com a utilização do protocolo OAI-PMH, um sistema pode obter todos os metadados dos documentos que um determinado repositório possui. Para tal, um sistema deve apontar para o endereço do repositório que se deseja obter os metadados e inserir as configurações da coleta, assim, será retornado uma lista contendo os metadados pertencentes aos objetos digitais contidos naquele repositório.

O protocolo OAI-PMH permitiu que houvesse uma grande facilidade na distribuição das informações científicas. Assim, por meio da interoperabilidade oferecida, ocorreu uma maior visibilidade e facilidade de acesso às produções científicas contidas em periódicos. Com isso, um editor de uma revista consegue obter os registros de documentos de suas produções espalhadas em outros locais, pois os dados estão abertos e podem ser compartilhados entre as organizações, além de que, as instituições de pesquisa podem criar mecanismos para divulgar e estimular os trabalhos produzidos (VIANA; ARELLANO; SHINTAKU, 2013).

¹ Dublin Core Metadata Initiative. Disponível em: <<http://dublincore.org/>>. Acesso em: 22 set. 2016.

Nessa perspectiva, um ambiente que utiliza o OAI-PMH tem basicamente dois atores principais, os provedores de dados e os provedores de serviços. Os provedores de dados são basicamente bases de dados que disponibilizam documentos eletrônicos para poderem ser acessados. Estes provedores, caso utilizem o protocolo OAI-PMH, permitem com que os seus metadados possam ser coletados por meio de um programa de *harvesting*. Os provedores de serviços fazem a interface para realizar o acesso aos metadados dos provedores de dados, sendo que este acesso pode ser feito a um ou mais provedores de dados (MARCONDES; SAYÃO, 2001).

Uma das questões principais envolvendo o protocolo OAI-PMH é o processo de *harvesting*. O *harvesting* é um processo unilateral em que os provedores de serviços realizam periodicamente uma busca coletando os metadados, podendo armazenar estes metadados em uma base de dados própria, para a exibição conforme as consultas realizadas pelos usuários (GARCIA; SUNYE, 2003).

Na próxima seção será demonstrado como, a partir dos conceitos de *harvesting*, OAI-PMH e provedores de dados e serviços, o presente trabalho propõe um modelo de criação de provedores de serviços interativos.

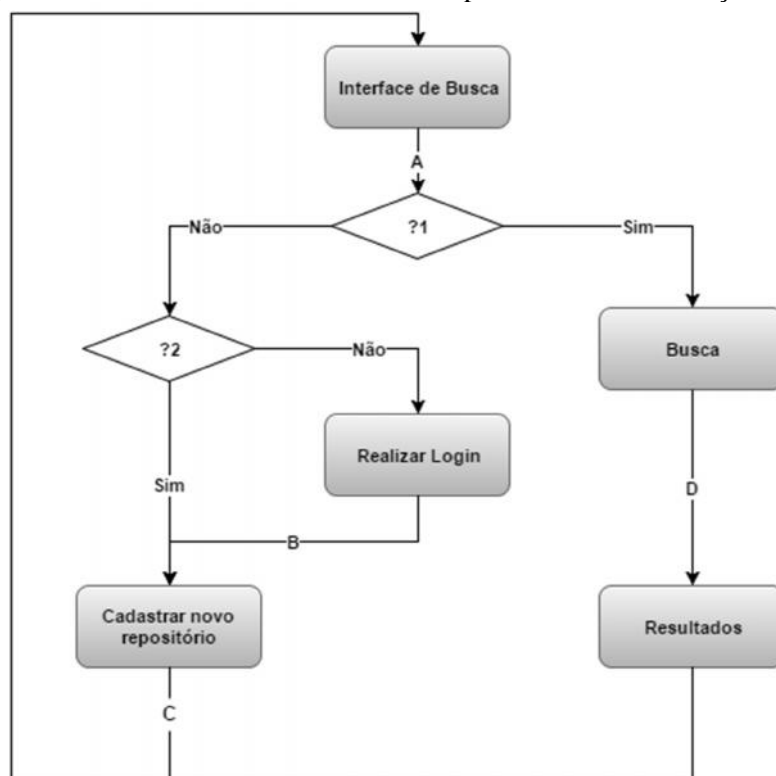
4 MODELO PROPOSTO DE PROVEDOR DE SERVIÇO INTERATIVO

O processo de extração e recuperação dos documentos em repositórios digitais, por meio do uso do protocolo OAI-PMH, tem sido objeto de estudo de diversas pesquisas, como Oliveira e Carvalho (2009) e Marcondes e Sayão (2001). Neste âmbito, inúmeras organizações têm utilizado este protocolo para a criação de provedores de serviço, que mantém acesso a diversas fontes informacionais. Portanto, é necessário que os provedores de serviços tenham mecanismos que colaborem para que as necessidades informacionais dos usuários sejam melhores atendidas.

Para colaborar nesta tarefa, este trabalho cria um modelo de interoperabilidade dentro de repositórios digitais, em que a principal contribuição da pesquisa ocorre na interação entre os usuários com as fontes informacionais. Tal interação visa fazer com que o usuário escolha as fontes utilizadas para a realização de uma busca. Para atingir esse objetivo, criamos um modelo que possibilita ao usuário definir em quais repositórios será realizada a busca, e se caso, um

repositório que o usuário deseje utilizar não faça parte do conjunto de repositórios disponíveis pelo sistema, permite-se que tal usuário insira esta nova fonte informacional neste conjunto.

Figura 1 - Fluxo das atividades executadas pelo usuário na utilização do modelo



Legenda:

A - O usuário deverá escrever a expressão de busca e escolher em quais repositórios deverá ocorrer a busca;

B - O usuário deverá estar logado para poder realizar o cadastramento de um novo repositório;

C - O usuário deverá cadastrar um novo repositório apontando o nome e a URL do mesmo, para que seja feito o harvesting dos metadados deste repositório escolhido;

D - Busca no conjunto de metadados dos repositórios escolhidos, aqueles que atenderam a expressão de busca. A tela de resultados contém um único conjunto de resultados que contempla os metadados dos diversos repositórios escolhidos, que atenderam a expressão de busca

?1 - Os repositórios que o usuário deseja utilizar estavam listados?

?2 - O usuário está logado no sistema?

Fonte: Elaborado pelos autores, 2016 .

Assim, o usuário tem a possibilidade de inserir novos repositórios, de modo que, este indivíduo deverá introduzir informações acerca do repositório em questão, e o sistema realizará uma coleta (*harvesting*) de todos os metadados que tal repositório possui. Para compreender com

mais clareza as interações realizadas pelo usuário durante a utilização do modelo proposto, construímos um diagrama de fluxo que expressa as diversas atividades executadas pelo usuário, ilustrado na figura 1.

Dentro das interações destacadas pela figura 1, destacamos que o usuário deverá estar *logado*² no sistema no momento de inserir um novo repositório. Tal processo mostra-se necessário para que exista um registro dos repositórios que as pessoas estão inserindo no sistema, além de ser possível que o usuário tenha possibilidades de configuração e personalização das fontes informacionais utilizadas em suas buscas.

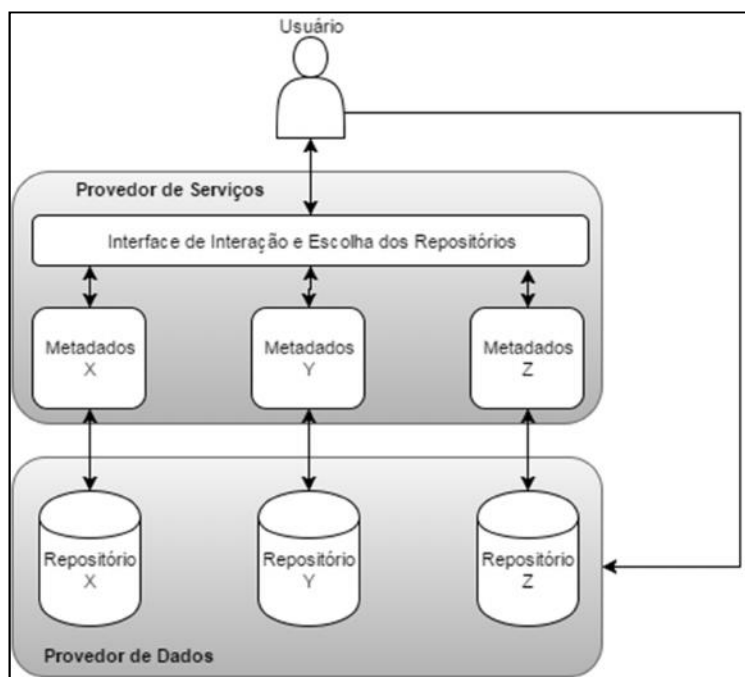
Essas possibilidades promovem um buscador (ou no contexto desta pesquisa, um provedor de serviços) em que o próprio usuário define suas fontes informacionais. Essa possibilidade, conduz a personalização da busca em um nível distinto do ocorrido atualmente, afirmação feita baseado no estudo de Monteiro (2009). Neste artigo, a autora discorre acerca das personalizações que os mecanismos de busca e outros *websites*, como de comércio eletrônico, tem possibilitado a seus usuários, como a escolha das variáveis de ordenação dos resultados e resultados baseados nos históricos de buscas ou de compras anteriores. Contudo, a autora não indica nenhum mecanismo ou *website* que fornece o controle ao usuário das fontes informacionais que serão considerados na busca realizada.

No contexto apresentado, elaboramos um modelo que contém os elementos necessários para o funcionamento de um esquema de interoperabilidade em repositórios digitais, bem como as interações demonstradas pelo fluxo de atividades. A figura 2 ilustra o modelo desenvolvido.

O modelo demonstrado na figura 2 apresenta duas divisões principais, o bloco de provedor de dados e o de provedor de serviços. Vale destacar que são apresentados no modelo três repositórios, no entanto o modelo não se limita a este número, podendo funcionar com somente um ou com diversos repositórios. Na sequência será explicado o modelo proposto detalhadamente.

² Pode-se compreender *logado* como um processo de autenticação em que o usuário deverá fornecer um usuário e uma senha já previamente cadastrados.

Figura 2 - Modelo de Interoperabilidade de Repositórios Digitais



Fonte: Elaborado pelos autores, 2016.

O usuário interage primeiramente com o provedor de serviços, em que este indivíduo poderá realizar uma busca, escrevendo uma expressão de busca, e poderá selecionar os repositórios digitais que serão as fontes informacionais utilizadas. Além dessa primeira interação, o usuário poderá interagir com o provedor de dados, ao passo que poderão ser realizadas inserções de novos repositórios como fontes de informação, assim, ao usuário agregar uma nova fonte de informação, o modelo deverá inserir este novo repositório, realizando o *harvesting* dos metadados do mesmo.

Estes processos de *harvesting* também são demonstrados no modelo, por meio das setas que ligam os blocos dos Repositórios (no Provedor de Dados) com os seus respectivos metadados (no Provedor de Serviços), em que a coleta é realizada utilizando o protocolo OAI-PMH. Como dito anteriormente, esse padrão permite que um programa realize a coleta de todos os metadados presentes em um determinado repositório. Deve-se, assim, realizar uma implementação de um programa simples, que realiza a extração dos metadados, utilizando o processo de *harvesting*.

Os metadados extraídos devem ser armazenados em uma base de dados única, em um determinado formato. O uso do OAI-PMH, auxilia nessa questão, pois o retorno dos metadados coletados ocorre em padrão Dublin Core, auxiliando na existência de um modelo único, para a realização da indexação e das buscas, além do modo como as informações serão apresentadas aos usuários.

Cabe ressaltar que uma das questões que deverão ser tratadas pelo modelo é a atualidade dos registros do repositório. Isto porque um repositório digital é atualizado frequentemente, necessitando haver verificações periódicas da existência de novos registros. A indexação computacional acontecerá automaticamente após a coleta de novos metadados, processo este que apresenta distinta importância, pois quando se trata de repositórios, há um grande volume de metadados que deverão ser indexados e buscados.

Por fim, o modelo indica um retorno com os resultados que atendam a expressão de busca construída pelo usuário. Tais resultados deverão ser apresentados em uma interface que possibilite uma interação com os resultados, podendo verificar os metadados que foram recuperados.

Como prova de conceito do modelo proposto, realizamos a implementação de um protótipo que visa validar a viabilidade do desenvolvimento deste modelo. Essa implementação será detalhada na próxima seção.

5 IMPLEMENTAÇÃO DO MODELO PROPOSTO

A implementação do modelo interoperável de repositórios seguiu o modelo descrito na seção 4, em que são identificados os pontos necessários para a implementação do modelo. Vale destacar que todo o processo descrito no presente trabalho foi desenvolvido pelos autores desta pesquisa.

Para a construção desse protótipo implementamos três serviços que se interagem: 1) “Interface Visual e Controle”, representando as interfaces gráficas de interação entre o usuário e o sistema, bem como o controle de solicitações e requisições feitas pelos usuários; 2) “Servidor Solrj”, responsável pela interação entre a “Interface Visual e Controle” e o Apache SOLR, em que o “Servidor Solrj” tem esse nome por utilizar como principal ferramenta um conjunto de códigos programados com a Linguagem Java, construído com o intuito de fazer a comunicação

entre esta linguagem e o Apache Solr e; 3) “Servidor Solr”, responsável pela indexação computacional, armazenamento e busca dos metadados dos repositórios digitais extraídos por meio do *harvesting*.

A utilização do Software Apache Solr³ para a realização das atividades descritas ocorreu, pois, essa ferramenta permite a realização da indexação computacional de diversos arquivos, para uma posterior recuperação de maneira eficiente e efetiva. Além disso, o destaque do Apache Solr é devido a velocidade com que é feita a indexação e a busca dos arquivos, ademais, a integração com outro *software*, o Apache Lucene⁴, abre uma gama de possibilidades para a realização das buscas, permitindo que sejam criadas diversas expressões, o que melhora a qualidade da recuperação dos registros indexados.

Outra característica do Apache Solr é que este sistema separa os dados inseridos nos chamados *core*, que são núcleos onde são inseridos os objetos digitais, como se fossem pequenas separações de uma base de dados. No presente trabalho, cada *core* contém os registros de um repositório digital.

A seguir, explicamos os diversos processos que foram implementados para a construção do protótipo, a sequência da explicação é feita de acordo com a ordem que ocorre os processos realizados pelo sistema.

O primeiro procedimento executado ocorre na interação do usuário com uma interface visual do sistema. Tal interação ocorre por meio de três tipos de tela, apresentadas na figura 3:

1. A tela contendo a Letra A na figura, permite com que o usuário possa construir uma expressão de busca e apontar quais repositórios serão utilizados como fontes de informação;
2. A tela com a Letra B, que, será utilizada caso o repositório que o usuário deseje utilizar como fonte, não esteja disponível no sistema, e tal usuário possa cadastrar um novo repositório, fornecendo o seu nome e a sua URL;
3. A tela de *login*, com a Letra C, necessária para que um usuário possa realizar o cadastro de um repositório, com o intuito de ficar registrado e ter o controle dos usuários que estão solicitando o cadastro de novos repositórios.

³ Apache Solr. Disponível em: <<http://lucene.apache.org/solr/>> Acesso em: 25 ago. 2015.

⁴ Apache Lucene. Disponível em: <<http://lucene.apache.org/>> Acesso em: 25 ago. 2015.

Figura 3 – Telas de interação do usuário

The figure displays three screenshots of a user interface for digital repository interoperability. Screenshot C shows a login form with fields for 'Usuário' and 'Senha' and a 'Fazer login' button. Screenshot B shows a form to add a repository with fields for 'Nome Repositório' and 'URL Repositório' and an 'Inserir Repositório' button. Screenshot A shows a search interface with a search bar containing 'correlação', a table of selected repositories (LUME and UNESP), a 'BUSCAR' button, and logos for unesp, GPnti, FAPESP, CAPES, USP, and a CD-ROM.

Fonte: Elaborado pelos autores, 2016.

Após o usuário escrever sua expressão de busca ou solicitar o cadastramento de um novo repositório, o serviço de “Interface Visual e Controle” envia para o “Servidor Solrj” um XML contendo ou a expressão de busca, com os repositórios que deverão ser utilizados na busca, ou os dados para ser realizado o cadastramento e *harvesting* de um novo repositório.

Para tratar estas duas tarefas, o “Servidor Solrj” apresenta duas funções principais detalhadas a seguir.

A primeira função trata de realizar o processo de *harvesting*, em que são coletados os metadados dos diversos repositórios. A implementação realizada permite que sejam coletados

metadados de quaisquer repositórios que utilizam o protocolo OAI-PMH em seus servidores. A realização do processo de *harvesting* pelo “Servidor Solrj” consiste das seguintes ações: conexão ao serviço OAI-PMH do servidor do repositório, solicitação dos metadados, coleta destes no padrão Dublin Core, e por fim, alteração de tais metadados no formato correto para a realização da indexação dos registros.

A conexão ao serviço OAI-PMH e a solicitação dos metadados ao servidor do repositório ocorre por meio da utilização de uma URL que contém as informações desejadas. Tal URL engloba em suma o sítio onde está hospedado o repositório, a indicação que se refere ao serviço de OAI-PMH e a requisição em si. Cabe destacar que, o usuário ao solicitar o cadastramento de um novo repositório, deverá inserir somente a sua URL, como por exemplo “http://repositorio.unesp.br” ou “http://www.lume.ufrgs.br”, e o sistema criará uma nova URL contendo as informações necessárias, em que nos exemplos relatados, as URLs ficariam: “http://repositorio.unesp.br/oai/request?verb=ListRecords&metadataPrefix=oai_dc” e “http://www.lume.ufrgs.br/oai/request?verb=ListRecords&metadataPrefix=oai_dc”.

Posteriormente a coleta dos metadados, o “Servidor Solrj” envia ao “Servidor Solr” um XML contendo os metadados obtidos por meio do processo de *harvesting*. A partir disto, o “Servidor Solr” cria um novo *core* em que serão armazenados os registros coletados. Configurou-se o Apache Solr para que permitisse com que a indexação computacional ocorresse baseado nos quinze elementos do Dublin Core.

Ao final, o “Servidor Solr” envia ao “Servidor Solrj” um XML confirmando a inserção dos registros de um repositório, finalizando assim a tarefa de inserção de um novo repositório como fonte informacional.

A segunda função consiste em realizar uma solicitação para o “Servidor Solr”, em que aponta a expressão de busca construída pelo usuário, bem como os repositórios que foram escolhidos para realizar a busca. Para realizar tal tarefa, este serviço faz diversas requisições, uma para cada fonte informacional escolhida pelo usuário. Assim, utilizando-se da linguagem Java, as requisições são construídas contendo duas variáveis, a expressão de busca e o repositório utilizado.

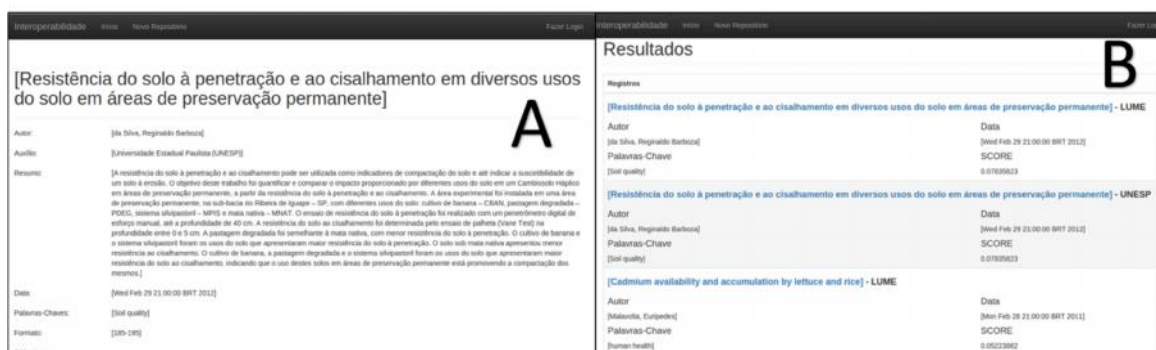
Desta forma, o processo de busca, executado pelo “Servidor Solr”, ocorre quando recebe do “Servidor Solrj” a expressão de busca construída pelo usuário e o indicativo de qual

repositório será feita a busca, conforme relatado. Assim, o Servidor Solr irá recuperar dentro de cada *core* dos repositórios escolhidos, os registros que atendam a expressão de busca construída pelo usuário.

Após ser realizada a consulta nos registros dos repositórios, o “Servidor Solr” deverá retornar os dados obtidos para o “Servidor Solrj”. Nesta fase, é enviado um XML contendo os documentos que atenderam a expressão de busca construída pelo usuário. Posteriormente a essa fase, o “Servidor Solrj” executa uma transformação do XML obtido, para formatos legíveis na “Interface Visual e Controle”.

Por fim, na “Interface Visual e Controle” se implementou as telas que retornam para o usuário os registros que atenderam a expressão de busca. Tais telas podem ser visualizadas na figura 4.

Figura 4 - Tela de Pesquisa

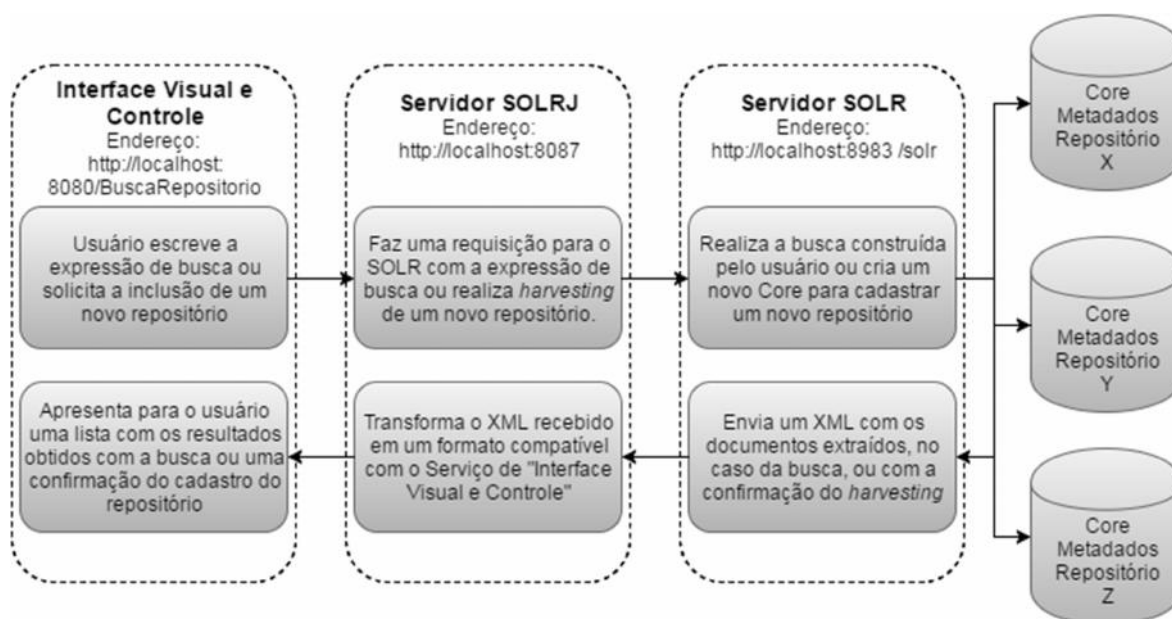


Fonte: Elaborado pelos autores, 2016.

Na figura 4 são apresentados dois tipos de telas: a tela com a Letra A mostra ao usuário os metadados de um artigo e o link para ser acessado o documento PDF, que está armazenado no provedor de dados, enquanto a tela com a Letra B ilustra como os resultados são retornados ao usuário após a realização de uma pesquisa. Desta forma, finaliza-se esta segunda função, tratando do processo de busca dos registros.

Para demonstrar os diversos serviços, que foram implementados para o desenvolvimento do protótipo apresentado neste trabalho, criamos a figura 5, que ilustra como tais serviços se relacionam.

Figura 5 – Relacionamento dos serviços do protótipo



Fonte: Elaborado pelos autores, 2016.

A figura 5 sintetiza a explanação realizada anteriormente, bem como apresenta detalhes práticos acerca da implementação realizada, como o endereço local que os serviços ficaram disponíveis durante a implementação.

Destaca-se também que os serviços construídos no presente trabalho se deram como serviços Web que recebem e enviam solicitações. A integração entre a visualização ocorreu por meio de *framework* que permite com que as páginas HTML estejam integradas com linhas de código construídas em linguagem de programação Java. Dessa forma, quando o usuário realiza uma solicitação no *website* construído, o sistema de controle resolverá as solicitações do usuário, se comunicando com os diversos outros serviços contidos na proposta.

6 CONSIDERAÇÕES FINAIS

Os repositórios digitais estão se tornando uma fonte de informação essencial dentro de instituições de ensino, órgãos governamentais e comunidades específicas, exigindo que sejam estudadas e desenvolvidos novos meios de acesso aos objetos digitais disponíveis nestas ferramentas.

Nesta perspectiva, a interoperabilidade nos repositórios digitais pode oferecer um meio bastante efetivo para promover uma recuperação da informação mais eficiente nestes ambientes digitais. Criar um provedor de serviços que acesse a diversos provedores de dados, para reunir as produções de uma instituição, órgão, comunidade ou nação, oportuniza que usuários tenham acesso à uma grande quantidade de documentos relevantes, sem existir a necessidade do usuário realizar buscar em diversos serviços, para que ele tenha em mãos todos os documentos pertencentes aquele determinado domínio.

Assim, o uso do protocolo OAI-PMH permite que a interoperabilidade seja realizada, por meio do acesso aos metadados dos provedores de dados, sendo que, por ser uma tecnologia bastante estudada e utilizada, e devido a facilidade para a realização do processo de *harvesting*, a viabilidade do emprego deste conceito de provedores de serviço acessando diversos provedores de dados, se torna alta.

O modelo proposto nesta pesquisa avança no uso de interoperabilidade em repositórios digitais, ao permitir interatividade na escolha das fontes informacionais, deixando a cargo do usuário a definição dos repositórios utilizados para a realização da busca. Assim, tal modelo possibilita ao usuário um nível de personalização da busca, que transcende os graus de personalização encontrados atualmente nos principais ferramentais de busca.

Por meio da implementação, verificou-se que a proposta é possível de ser construída, bem como mostrou ser factível um sistema que permite o cadastramento dinâmico de novas fontes informacionais (repositórios digitais). Outra característica validada pela implementação, foi a escolha pelos usuários dos repositórios que serão utilizados para a realização da busca, no momento da execução da mesma.

O uso do Apache Solr se mostrou bastante eficaz, tanto para a realização da indexação computacional, quanto o processo de busca em si, possibilitando que os metadados coletados fossem indexados com facilidade, além de fornecer subsídios para uma busca eficiente.

Portanto, o modelo proposto e sua implementação demonstraram uma nova forma de construir provedores de serviços, focados nas fontes informacionais que o usuário deseja utilizar em suas buscas. Assim, o uso de repositórios digitais poderá passar a ter uma maior utilização, pois os usuários encontrarão um novo meio de construir suas consultas, que poderão ter como base produções intelectuais de distintas instituições e temáticas. O presente estudo apresenta como trabalhos futuros a incorporações de outras fontes informacionais, como revistas científicas publicadas em plataformas de *Open Journal Systems* (OJS), que permitirá uma maior abrangência no conjunto de fontes que poderão ser utilizadas.

Interoperability in Digital Repositories: model of service provider interactive

ABSTRACT

The increased use of information and communication technologies, especially regarding the Web, provided the increased use of digital formats for availability of scientific texts. In this context, have emerged and has been increasing the use of digital repositories as a means of archiving and access of thematic and institutional scientific production. Several protocols and means of access are enabling better use of repositories in different scenarios. Interoperability and OAI-PMH, has been working to establish ways to retrieve documents in many ways, more efficiently. However, the question is how to create a system that allows the user to choose the informational sources that he wants to use in an interoperable system of digital repositories. Thus, this research aims to create and implement an interoperable model of digital repositories, which allows the user to register and choose the informational sources that will be used in their searches. The methodology used exploratory and applied, where based on the bibliography of repositories and interoperability, was created and implemented an interactive model of digital repositories. With this work, we found that it is possible to develop interoperable system allowing the registration of informational sources in real time, providing the user the role of choosing the informational sources that he wants to use. With the contribution of Apache Solr technology, the results from several repositories have been merged, following a system of internal relevance of this software. The system created near the user's computer program, allowing that the information needs of users are best met.

Keywords: Digital repositories. Interoperability. OAI-PMH.

REFERÊNCIAS

BALANCIERI, R. et al. A análise de redes de colaboração científica sob as novas tecnologias de informação e comunicação: um estudo na Plataforma Lattes. **Ciência da Informação**, v. 34, n. 1, p. 64-77, 2005.

BORDIN, A. S. **Extensão do protocolo Z39. 50 para busca de servidores de bibliotecas digitais**. 2002. 96 f. Dissertação (Mestrado em Ciência da Computação) – Centro Tecnológico, Universidade Federal de Santa Catarina, Florianópolis. 2002.

IBICT. **Sobre Repositórios Digitais**. Instituto Brasileiro de Informação em Ciência e Tecnologia. 2015. Disponível em: <<http://www.ibict.br/informacao-para-ciencia-tecnologia-e-inovacao%20/repositorios-digitais>> Acesso em: 15 de jun. 2016.

FERREIRA, A. M. J. F. C.; VECHIATO, F. L.; VIDOTTI, S. A. B. G. Inclusão digital e social de indivíduos da terceira idade por meio do uso de tecnologias de informação e comunicação: o papel dos ambientes colaborativos. In.: ENCONTRO NACIONAL DE PESQUISA DA ANCIB, 9., 2008, São Paulo. **Anais...** São Paulo: ENANCIB, 2008. Disponível em: <<http://enancib.ibict.br/index.php/enancib/ixenancib/paper/viewFile/3067/2193>>. Acesso em> 20 maio 2016.

GARCIA, P. D. A. B.; SUNYE, M. S. O protocolo OAI-PMH para interoperabilidade em bibliotecas digitais. In: **CONGRESSO DE TECNOLOGIAS PARA GESTÃO DE DADOS E METADADOS DO CONE SUL**, 2003, Ponta Grossa. **Anais...** Ponta Grossa: CONGED, 2003. Disponível em: <http://conged.deinfo.uepg.br/~iiconged/2003/Artigos/artigo_09.pdf>. Acesso em: 17 jul. 2016.

LAGOZE, C.; VAN DE SOMPEL, H. The Open Archives Initiative: Building a low-barrier interoperability framework. In: ACM/IEEE Joint Conference on Digital Libraries, 2001, Roanoke. **Proceedings...**, Roanoke: ACM. 2001. Disponível em: <<http://www.elbib.ru/index/index.phtml?page=elbib/eng/journal/2001/part6/LS>>. Acesso em: 19 jun. 2016.

LEITE, F.C.L. et al. **Como gerenciar e ampliar a visibilidade da informação científica brasileira**: repositórios institucionais de acesso aberto. IBICT, Brasília. 2009. Disponível em: <<http://livroaberto.ibict.br/bitstream/1/775/4/Como%20gerenciar%20e%20ampliar%20a%20visibilidade%20da%20informa%C3%A7%C3%A3o%20cient%C3%ADfica%20brasileira.pdf>>. Acesso em: 10 jul. 2016.

MARCONDES, C. H.; SAYÃO, L. F. Integração e interoperabilidade no acesso a recursos informacionais eletrônicos em C&T: a proposta da Biblioteca Digital Brasileira. **Ciência da Informação**, v. 30, n. 3, p. 24-33, 2001.

MONTEIRO, S. D. As múltiplas sintaxes dos mecanismos de busca no ciberespaço. **Inf. Inf.**, Londrina, v. 14, n. esp, p. 68 – 102. 2009. Disponível em:

<<http://www.uel.br/revistas/uel/index.php/informação/article/viewArticle/2027>>. Acesso em: 14 abr. 2016.

OLIVEIRA, Renan Rodrigues; CARVALHO, Cedric Luiz. Implementação de Interoperabilidade entre Repositórios Digitais por meio do Protocolo OAI-PMH. **Goiânia: UFG, 2009.**

SANTAREM SEGUNDO, J. E. **Representação Iterativa: um modelo para Repositórios Digitais.** 2010. 224 f. Tese (Doutorado em Ciência da Informação) – Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília. 2010.

SAYÃO, L. et al. **Implantação e gestão de repositórios institucionais.** EDUFBA, Salvador. 2009.

SHINTAKU, M. **Federação de Repositórios Científicos:** identificação, análise e proposta de modelo baseado nas tendências tecnológicas e da Ciência. 2015. 268 f. Tese (Doutorado em Ciência da Informação) – Universidade de Brasília, Brasília. 2015.

VIANA, C. L. M.; ARELLANO, M A. M; SHINTAKU, M. **Repositórios institucionais em ciência e tecnologia:** uma experiência de customização do DSpace. 2013. Disponível em: <<http://eprints.rclis.org/7168/1/viana358.pdf>>. Acesso em: 16 jun. 2016.