

Proposta de um Serviço de Busca na *Web* por Documentos Relacionados aos Trabalhos Catalogados em uma Biblioteca Digital

Celeste Magela Campos da Silva¹ Guilherme Tavares de Assis^{1,2} Denilson de Oliveira Chaves¹

¹Centro Universitário de Belo Horizonte (Uni-BH)
{cmagela,dochaves}@gmail.com tavares@acad.unibh.br

²Info Educacional – Belo Horizonte MG
{gtavares@infoeducacional.com.br}

Resumo - Este artigo apresenta a proposta de um serviço de busca na *Web*, que utiliza resultados de consultas submetidas à máquina de busca Google Scholar, para encontrar materiais diretamente relacionados a algum trabalho científico catalogado numa biblioteca digital. Para comprovar a eficácia do serviço de busca proposto, o mesmo foi desenvolvido na forma de um utilitário para a Biblioteca Digital do Centro Universitário de Belo Horizonte (BDUni). Assim, experimentos práticos foram realizados no intuito de avaliar os benefícios que o serviço de busca na *Web* proposto pode trazer à utilização de uma biblioteca digital, e também de determinar qual é o mais eficiente critério de busca a ser utilizado no serviço para se obter os melhores resultados da *Web* para os usuários.

palavras-chave – bibliotecas digitais, busca na *web*.

Abstract - This paper presents an approach to a Web search service, which uses results of queries submitted to the search machine Google Scholar, in order to find documents directly related to any scientific work stored in a digital library. In order to demonstrate the effectiveness of the proposed Web search service, it was developed as a tool for the Belo Horizonte University Digital Library (BDUni). Thus, practical experiments were performed in order to evaluate the benefits that the proposed Web search service can bring to the use of a digital library, and also in order to determine which is the more efficient search option to be used in the service to obtain the best Web results for the users.

keywords – digital libraries, web search.

1 INTRODUÇÃO

Atualmente, a Internet possui um vasto repositório de documentos digitais em constante crescimento, tornando-se um importante ambiente para disseminação de publicações e serviços de informação. Neste contexto, as Bibliotecas Digitais (BDs) têm sido utilizadas como um mecanismo para organizar tais publicações, permitindo melhor acesso às mesmas.

BDs, segundo a *Digital Library Federation*¹, são sistemas eletrônicos de armazenamento de informações com capacidade para selecionar, estruturar, oferecer acesso intelectual, interpretar, distribuir, preservar a integridade e garantir a persistência, ao longo do tempo, das coleções de trabalhos digitais, para que estejam legíveis e disponíveis para uso por uma comunidade ou um conjunto de comunidades definidas. O conteúdo das BDs inclui metadados, podendo ser internos ou externos às bibliotecas. Esses metadados são atributos de dados ou documentos, normalmente descritivos (como, por exemplo, informações sobre autor e conteúdo de uma publicação), freqüentemente subdivididos em categorias e registrados de acordo com algum padrão. Geralmente, as BDs permitem aos membros de uma comunidade acadêmica a possibilidade de fazer pesquisas em cima de seu catálogo de documentos cadastrados e ainda a possibilidade de efetuar o *download* de algum trabalho em particular, possibilitando o estudo por completo do mesmo.

Com o intuito de melhorar o nível de satisfação de um usuário ao utilizar uma BD para recuperar documentos de interesse, surgiu a necessidade de se verificar a proposta de um novo serviço onde, a partir de um documento cadastrado numa BD, fosse possível, de forma automática, fazer na *Web* um levantamento bibliográfico preliminar, baseado no assunto desse documento. Com isso, o objetivo principal desse artigo é apresentar uma proposta de extensão das funcionalidades oferecidas por uma BD, a partir do desenvolvimento de um serviço responsável pela busca na *Web* por documentos diretamente relacionados ao assunto abordado por algum trabalho científico catalogado nesta BD.

Para comprovar a eficácia do novo serviço proposto, experimentos práticos foram realizados, envolvendo a construção de um serviço de busca na *Web*, anexado em uma BD. Tais experimentos foram feitos por usuários reais, verificando a satisfação dos mesmos em relação a

¹ Consórcio de bibliotecas e agências relacionadas que usa tecnologias de informação eletrônica para estender suas coleções e serviços.

cada um dos critérios definidos de busca na *Web* e em relação ao serviço proposto como um todo, de acordo com a qualidade das respostas pretendidas. Assim, após os experimentos, foi possível analisar os resultados obtidos, medindo os ganhos com a nova funcionalidade.

Este artigo encontra-se organizado da seguinte maneira: na Seção 2, é apresentada uma descrição de trabalhos diretamente relacionados ao assunto abordado por este artigo; na Seção 3, é descrita a proposta do serviço de busca; na Seção 4, são descritos os experimentos práticos realizados e os resultados obtidos; e, na Seção 5, são apresentadas as conclusões obtidas e as perspectivas de trabalhos futuros.

2 TRABALHOS RELACIONADOS

Com relação ao serviço de busca na *Web*, existem alguns trabalhos importantes na literatura [1,2,7]. Particularmente, a estratégia definida neste trabalho baseia-se nas propostas apresentadas em [1] e [2], sendo descritas resumidamente nas subseções 2.1 e 2.2, respectivamente.

2.1 BUSCA PELO TEXTO COMPLETO DE ARTIGOS CATALOGADOS EM UMA BD

Esta Subseção descreve, resumidamente, o trabalho apresentado em [1], que tem como objetivo propor um processo que utiliza resultados de consultas a máquinas de busca para encontrar, na *Web*, o texto completo de artigos catalogados em uma BD, para os quais não existe um apontador para o texto completo. Assim, o processo proposto pode enriquecer as informações do catálogo de uma BD, através da recuperação de conteúdo relacionado aos artigos presentes nessa biblioteca, a partir da submissão de consultas a máquinas de busca na *Web*.

Segundo [1], de acordo com a Fig.1, primeiramente é selecionado do catálogo da BD um registro (ou seja, um artigo) que possui um conjunto de metadados **ma**. A partir dos valores desse conjunto de metadados **ma**, é formulada uma consulta **qa** que é submetida a uma máquina de busca **M** na *Web*. Essa máquina de busca retorna, então, um conjunto de páginas **Pa** que são os resultados da consulta. A partir de **Pa**, é extraída e gerada uma lista ordenada de resultados candidatos **Ca**; cada resultado candidato é formado por um título e uma URL que indicam, respectivamente, o título do documento retornado pela consulta e a URL que informa onde este documento está armazenado.

Para formular a consulta **qa**, são selecionados, do conjunto de metadados **ma**, o título do artigo e o sobrenome de seu primeiro autor. A consulta segue então a sintaxe (“título do artigo” “sobrenome do primeiro

autor”). As aspas duplas são usadas para dar prioridade no conjunto de resultados candidatos àqueles que possuam URLs para documentos com casamentos exatos com o título e o sobrenome. Por exemplo, espera-se que a consulta (“*On line or invisible?*” “Lawrence”) retorne URLs com o texto do artigo “*On line or invisible?*” de Steve Lawrence, documentos com metadados que o descrevem, texto de outros artigos que o citam ou outros conteúdos relacionados a ele.

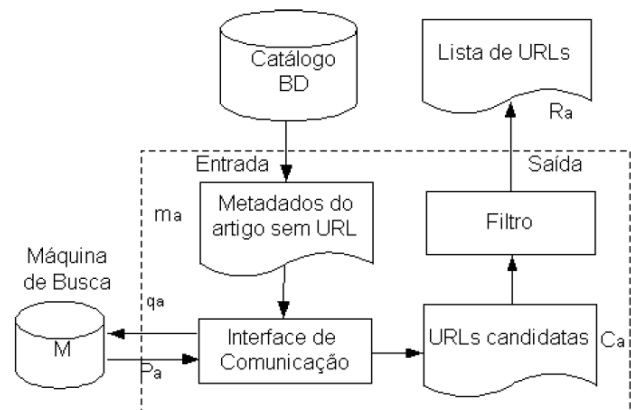


Fig.1 Processo usado para encontrar as URLs [1].

Segundo [1], o filtro que aparece na Fig.1 tem o intuito de retirar da lista de resultados candidatos **Ca** resultados pouco interessantes, por exemplo, documentos de currículos de autores, documentos relacionados a outros artigos que citam o artigo da consulta sem muita relação com o assunto nele tratado, etc. De uma forma geral, a estratégia adotada consiste em retirar resultados que possuam título pouco similar ao título do artigo da consulta. Também são excluídos os resultados cujas URLs apontem para documentos internos de outras BDs.

Para a realização de experimentos, foi utilizado um conjunto de 200 artigos selecionados de forma aleatória, 129 provenientes da BDBComp² e os restantes derivada de um subconjunto de dados da DBLP³, que foi nomeada de DBLP-Br. Para 116 deles (54 da BDBComp e 62 da DBLP-Br), foi obtida uma lista de URLs não-vazia em, pelo menos, uma das cinco máquinas de busca testadas, a saber: Scholar, Google, Citeseer, MSN e Yahoo. Para cada um desses 116 artigos, foi feita a união das listas recuperadas. Desse procedimento foi obtido um total de 818 diferentes URLs recuperadas (236 provenientes de testes com registros da BDBComp e as 582 restantes da

² Biblioteca Digital Brasileira de Computação da Universidade Federal de Minas Gerais -<http://www.lbd.dcc.ufmg.br/bdbcomp/bdbcomp.jsp>

³ Digital Bibliography & Library Project - <http://www.informatik.uni-trier.de/~ley/db/>

DBLP-Br). Então, as listas unidas foram divididas entre 23 especialistas para que eles, com o conhecimento dos dados presentes nos registros de metadados dos respectivos artigos, classificassem as URLs como relevantes ou não relevantes.

De acordo com [1], após avaliação, foi detectado que, para 95 das 116 listas obtidas (sendo 47 da coleção da BDBComp e 48 da DBLP-Br), pelo menos uma URL foi considerada relevante. A partir dessa classificação manual, foram gerados resultados comparativos para as máquinas de busca testadas. Para tanto, foi utilizada como métrica de comparação da eficácia das máquinas de busca, a precisão média nos documentos relevantes, onde intuitivamente, pondera a satisfação de um usuário ao encontrar documentos relevantes em uma lista ordenada de resultados, quando é examinada do início para o fim. Após ser tirada a média das precisões obtidas por cada máquina, foi calculado o ganho percentual da média de uma máquina de busca sobre a média de outra e, por fim, foi testado se o ganho comparativo dos testes pareados foi ou não significativo. Resumidamente, a partir das análises feitas, concluiu-se um ganho significativo do Scholar sobre o Google, percebendo que essas máquinas de busca possuem diferentes coberturas no problema tratado. Assim, aproveitando-se dessa conclusão experimental, a título de teste, os serviços do Scholar e do Google foram combinados da seguinte maneira: emitiu uma primeira consulta ao Scholar e uma segunda ao Google nos casos em que não há retorno de resultados na primeira consulta. Assim, nos casos em que o Scholar cobre o artigo procurado, permanecem os seus resultados. Nos outros, permanecem resultados retornados pelo Google. Essa combinação produziu um resultado significativamente melhor, ganhos de até 35%, do que o Scholar individualmente.

2.2 SERVIÇO DE BUSCA DO TEXTO COMPLETO DE ARTIGOS CATALOGADOS NA BDBCOMP

Esta Subseção descreve, resumidamente, o trabalho apresentado em [2], que propõe a implementação de um serviço para busca na *Web* pelo texto completo de qualquer artigo, a partir dos seus metadados catalogados na BDBComp.

Segundo [2], a BDBComp possui um repositório central que contém metadados das publicações de diversos periódicos, conferências e eventos da comunidade brasileira de computação. Porém, dos 5138 artigos catalogados, 65,7% deles não possuem a URL para o texto completo dos artigos, o que significava que o usuário provavelmente teria que utilizar algum outro serviço disponível na *Web* para localizar esse texto pelo qual ele havia se interessado; possivelmente, uma máquina de

busca. Assim, a partir deste cenário, foi proposta em [2] a implementação de um serviço para buscar o texto completo na *Web* de qualquer artigo. Tal serviço, posteriormente, foi acoplado à BDBComp com o objetivo de fornecer aos usuários da biblioteca maior comodidade e cobertura de suas necessidades, implicando também na melhoria da qualidade dos recursos oferecidos. O serviço implementado utilizou o processo de metabusca descrito por [1] (vide Subseção 2.1), baseado nos resultados de consultas submetidas a diversas máquinas de buscas, para encontrar a URL do texto completo correspondente ou de qualquer outro material relacionado.

Para implementar o serviço proposto, de acordo com [2], primeiramente foi feita uma análise de impacto sobre a arquitetura da BDBComp. Os passos de execução do funcionamento de tal biblioteca, antes da implantação do serviço eram os seguintes:

- 1) o usuário acessa a interface da BDBComp e submete uma consulta;
- 2) o sistema processa a consulta submetida;
- 3) os metadados são recuperados do repositório;
- 4) o resultado é apresentado ao usuário.

Assim, a interface mostra os metadados para o usuário. Nessa etapa, caso o artigo descrito pelos metadados não possua a URL para o texto completo, não restava alternativa ao usuário a não ser abandonar a BDBComp e tentar localizar esse trabalho na *Web*.

A partir da implantação do serviço proposto por [2], o fluxo de execução da situação descrita passou a acontecer da seguinte maneira conforme apresentado na Fig.2:

- 1) o usuário acessa a interface da BDBComp e submete uma consulta;
- 2) o sistema processa a consulta submetida;
- 3) os metadados são recuperados do repositório;
- 4) a interface mostra os metadados para o usuário; nessa etapa, caso o artigo descrito pelos metadados não possua a URL para o texto completo, é mostrado na interface um comando denominado “Buscar na *Web*” que permite ao usuário buscar na *Web* o trabalho pelo qual se interessou;
- 5) o usuário, se for de interesse, aciona o comando para buscar o texto completo do artigo na *Web*;
- 6) no caso, a Interface envia os metadados para um componente *Web*;
- 7) esse componente *Web*, por sua vez, compacta os metadados em um arquivo XML e o envia para o metabuscador;
- 8) o metabuscador processa os metadados recebidos e submete as consultas para várias máquinas de busca na *Web*;
- 9) as máquinas de busca respondem às consultas submetidas e o metabuscador processa os resultados;

- 10) o metabuscador compacta os resultados retornados pelas máquinas de busca em um arquivo XML e envia para o componente *Web* da biblioteca;
- 11) o componente *Web*, por sua vez, processa o XML e envia as URLs retornadas para a interface da BDBComp;
- 12) a interface da BDBComp recebe as URLs relevantes e as mostra para o usuário.

Segundo [2], o metabuscador é o módulo responsável pela submissão das consultas às máquinas de busca e pelo processamento dos resultados de tais máquinas. Tal componente recebe os metadados de um determinado artigo e submete as consultas às máquinas de busca Google e Google Scholar. Essas consultas são constituídas

do título do artigo seguido pelo sobrenome do primeiro autor. Posteriormente, o metabuscador recebe um conjunto de URLs candidatas. Nesse conjunto, é aplicada uma filtragem que tem, por objetivo, desconsiderar resultados não relevantes. O filtro implementado se baseia no coeficiente de Jaccard, conforme descrito em [1] por ser computado de maneira mais simples e rápida que outras medidas. Tal coeficiente é usado para verificar a similaridade entre os títulos dos resultados retornados pelas máquinas de busca e o título do artigo pesquisado, sendo que, para o serviço implementado, foi configurado um valor mínimo para esse coeficiente, a fim de desprezar apenas os resultados irrelevantes. Por fim, é feita a ordenação dos resultados por nível de relevância.

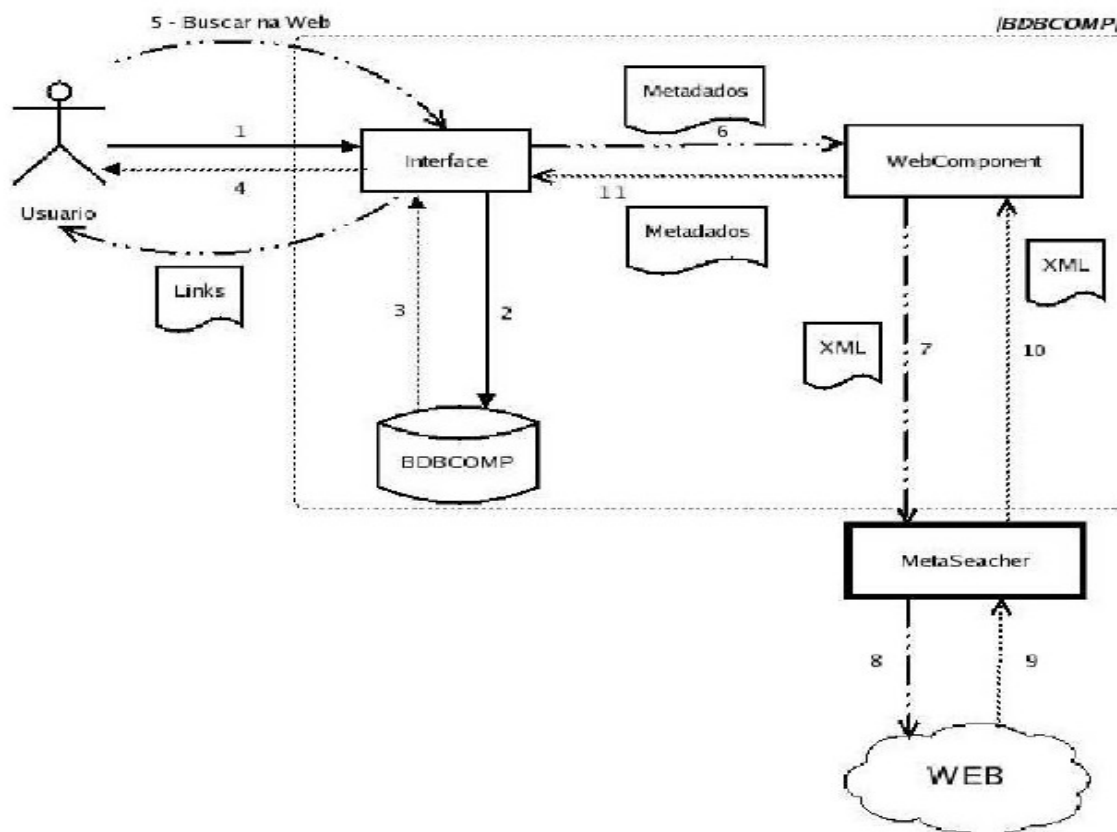


Fig.2: Fluxograma que representa a nova sequência de operação na BDBComp após a implementação do serviço proposto por [2].

Segundo [2], foram realizados inúmeros testes para verificar a eficácia do serviço implementado, que apontaram as seguintes situações:

- 1) o serviço localiza corretamente o texto completo do artigo na *Web*;
- 2) o serviço não localiza o texto completo, mas localiza trabalho(s) relacionado(s);

- 3) o serviço não localiza qualquer documento relevante na *Web*.

Após a fase de testes, [2] mediu o tempo entre o acionamento do comando "Buscar na *Web*", que marca o início da execução do serviço até o momento que são exibidos os resultados, e a média aritmética de espera da busca foi aproximadamente 6,3 segundos, o que

representa um resultado aceitável quanto ao benefício conseguido.

3 PROPOSTA DO SERVIÇO DE BUSCA

Segundo [1], um tipo comum de informação presente em BDs são metadados descritivos, ou seja, informações utilizadas para descrever os objetos do escopo da coleção de uma BD. Assim, no registro de metadados de uma BD voltada para publicações, é possível identificar facilmente informações sobre as mesmas, tais como o título, os nomes dos autores, as palavras-chave, o resumo e o nome do arquivo PDF referente armazenado.

Um fator, que pode aumentar o nível de satisfação dos usuários de uma BD, é a possibilidade da mesma de enriquecer as informações a respeito de uma determinada publicação armazenada, buscando-as automaticamente na *Web*. Tal situação é uma aplicação direta da abordagem do trabalho proposto: de posse dos metadados que descrevem um trabalho científico, realizar consultas a máquinas de busca e utilizar as respostas retornadas para sugerir apontadores para documentos que têm referência direta a tal trabalho catalogado na BD.

Dessa forma, pode-se dizer que a proposta desse artigo é apresentar um serviço de busca específico para uma BD, que se aproveita dos metadados catalogados de um trabalho científico, para expressar consultas sobre o vasto conteúdo indexado por máquinas de busca e assim selecionar, dentre as respostas retornadas, aquelas mais relevantes em relação ao trabalho desejado, aumentando assim o nível de satisfação dos usuários desta BD.

Segundo [1], máquinas de busca são serviços onde é possível expressar uma necessidade genérica de informação através de uma consulta, para a qual é retornado, como resposta, um conjunto de itens que apontam para documentos da *Web* que possam suprir a

necessidade de informação expressa pelo usuário. Conforme cita [7] tal serviço é uma ferramenta essencial para lidar com a quantidade de informação atualmente encontrada na *Web*.

As máquinas de busca oferecem serviços de busca em uma coleção de documentos sobre as quais não se tem controle devido à possibilidade de livre publicação na *Web*. Já em uma BD, a coleção tem foco específico e o seu conteúdo é totalmente controlado. O serviço de busca, a partir de uma BD, consiste em obter documentos relevantes na *Web* em relação a algum trabalho apresentado como resposta da consulta à biblioteca. Levando-se em consideração os experimentos feitos por [1] e as necessidades do serviço proposto, ao montar uma consulta a ser submetida a uma máquina de busca, as estratégias adotadas no serviço proposto são baseadas:

- 1) no metadado referente ao título do trabalho científico;
- 2) nas palavras-chave do trabalho científico;
- 3) no metadado referente ao(s) autor(es) do trabalho científico.

Para implementar o serviço de busca proposto e submeter as consultas, a máquina de busca utilizada foi a Google Scholar, por ter apresentado os melhores resultados nos testes feitos por [1] e por ser uma máquina de busca de propósito único e exclusivamente acadêmico, com chances maiores de atender aos propósitos do serviço proposto. Considerando, por exemplo, a consulta "Comparando modelos clássicos e estendidos de recuperação de informação", que corresponde ao título de um trabalho científico armazenado em uma BD, submetida à máquina de busca Google Scholar, uma lista de itens é retornada, contendo materiais diretamente relacionados à consulta submetida. Na Fig.3, é apresentado um item da lista de respostas retornada pela máquina de busca Google Scholar, onde se pode observar que possui um título, um pequeno resumo e a URL que aponta para o documento propriamente dito.

[\[PDF\] Tecnologias de Recuperação de Informações na Web](#)

JA Camacho-Guerrero, AA Macedo - 200.146.233.98

... Para manipular as coleções de documentos e para comparar as consultas ... Na Subseção 6.2.1.2 são apresentados modelos estendidos para cada ... Modelos Clássicos ...

[Ver em HTML - Pesquisa na web](#)

Fig.3: Item de resposta retornado pela máquina de busca Google Scholar.

Como já mencionado, o serviço proposto neste artigo tem o intuito de ajudar os usuários de uma BD a encontrar, na *Web*, URLs que levem a documentos que fazem referência direta a trabalhos científicos catalogados

nesta BD. O processo para prover o serviço foi idealizado a partir de [1]. Na Fig.4, o serviço é demonstrado esquematicamente.

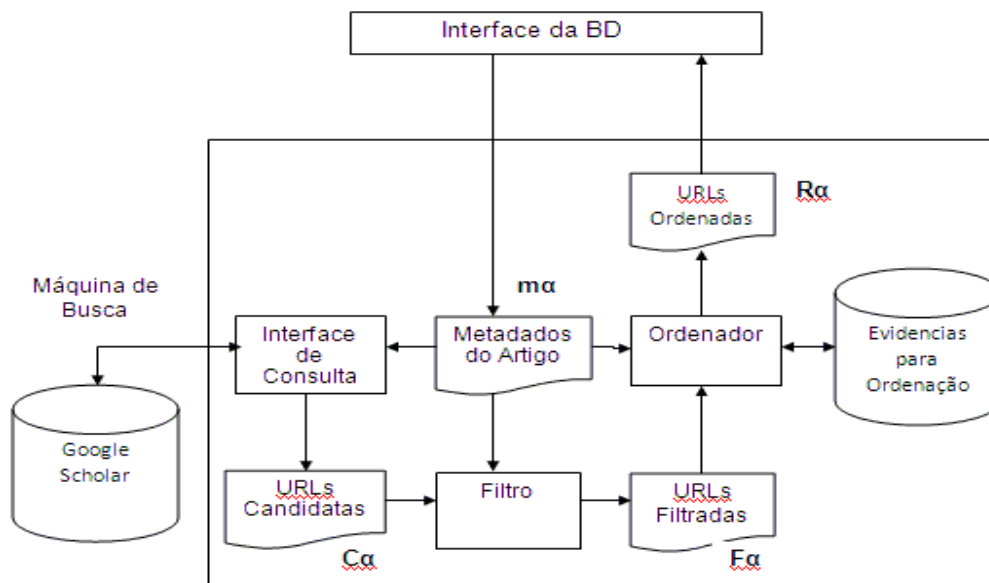


Fig.4: Arquitetura do Serviço de busca na Web.

De acordo com a Fig.4, o primeiro passo acontece quando o usuário interage com a interface da BD e, ao fazer uma consulta, tem como resposta os metadados de vários trabalhos científicos catalogados nesta BD. Esse passo acontece de forma independente do novo serviço implementado, e a partir dele, o usuário poderá escolher ou não utilizar o novo serviço. Assim, caso o usuário se interesse por algum dos trabalhos científicos retornados, ele pode escolher fazer uma busca na Web de forma automática por materiais diretamente relacionados a esse trabalho científico.

Para um melhor entendimento, considere que o trabalho científico desejado seja α e o conjunto de metadados do mesmo seja ma . Quando o usuário acionar o serviço proposto, ele deve escolher por qual critério deseja que a consulta na Web seja feita: pelo título de α , pelos autores de α ou pelas palavras-chave de α . De posse dessa informação mais ma , a interface de consulta da BD automaticamente gera e submete uma consulta para a máquina de busca Google Scholar, requisitando por documentos relacionados a α . Uma série de itens candidatos a terem referência com α são retornados pelo Google Scholar; esses itens candidatos são extraídos e compilados em uma lista $Ca=[C1,C2,...,Cn]$, onde cada resultado Ci é composto por uma dupla (Tc,Uc) sendo Tc o título e Uc a URL de um documento da Web retornado pelo Google Scholar. Foi estabelecido que Ca é composto, no máximo, pelos 20 primeiros itens de resposta retornados pelo Google Scholar, já que o mesmo retorna, em ordem, aqueles resultados que têm maiores chances de serem diretamente relacionados à consulta submetida.

Posteriormente, ocorre um filtro (vide Fig.3) que seleciona de Ca aqueles itens com maiores chances de possuírem conteúdo fortemente relacionado ao trabalho científico α procurado, ou seja, o filtro tenta remover da lista Ca , URLs candidatas que correspondem a documentos de pouco ou nenhum interesse para o usuário, tais como currículos de autores, textos de artigos que citam α mas tratam de assunto pouco relacionado ao artigo α , etc. Além disso, a remoção de URLs de Ca reduz o custo de processamento nos passos subsequentes do processo. A estratégia adotada consiste em remover de Ca elementos que possuem um título Tc pouco similar ao título do trabalho científico α , tentando simular o comportamento de usuários que somente examinam resultados cujo título Tc seja minimamente similar ao título de α .

Para avaliar a similaridade de Tc referente a ta (título do trabalho desejado α), a proposta usa o coeficiente de Jaccard. Tal coeficiente funciona da seguinte maneira: dados dois conjuntos de termos A e B, o coeficiente de Jaccard entre eles é dado por $J(A,B) = |A \cap B| / |A \cup B|$. O título Tc será considerado pouco similar a ta se o coeficiente de Jaccard calculado sobre os termos Tc e ta ($J(Tc, ta)$) for menor que o valor do limiar que foi definido previamente de forma manual. O valor desse limiar é ajustado visando garantir uma relevância mínima entre o título do trabalho científico armazenado na BD e o título do item retornado pelo Google Scholar.

Finalmente, o ordenador (vide Fig.4) ordena a lista Ca resultante após o filtro descrito anteriormente e gera uma nova lista Ra de URLs ordenadas usando, como evidência adicional, o título do documento retornado. O ordenador

opera procurando colocar, no topo da lista ordenada, aqueles documentos com maior chance de satisfazer às necessidades do usuário. A lista **Ra** é então retornada para a Interface da BD e apresentada ao usuário interessado.

4 EXPERIMENTOS PRÁTICOS

Com o objetivo de analisar o funcionamento do serviço de busca na *Web* descrito na seção anterior, foram realizados experimentos práticos que consistiram na aplicação de um questionário a 10 estagiários de uma empresa vinculada à área de *business intelligence*. Os estagiários cursam algum curso superior na área de

informática, possuindo vivência com interfaces *Web*, mas não possuindo conhecimento sobre BDs e, tão pouco, sobre serviço de busca na *Web*.

Para a realização dos experimentos práticos, foi considerada a Biblioteca Digital do Centro Universitário de Belo Horizonte (BDUni), que armazena atualmente 51 publicações científicas. Tal biblioteca encontra-se detalhada em [3,4], e manutenções e novas implementações da mesma em [5,6].

O objetivo do questionário foi medir o índice de satisfação quanto ao uso do serviço de busca na *Web* proposto, criando, para cada participante, 3 consultas compostas por 2 atividades, conforme descrito na Tab. 1.

Tabela 1
TAREFAS COMPOSTAS NO QUESTIONÁRIO

Tarefa	Atividade 1	Atividade 2
Consulta 1	Fazer, a partir da interface da BDUni, uma pesquisa por título de trabalhos monográficos; foram criadas três consultas diferentes, com os seguintes termos: “recuperação de informação”, “redes neurais” e “sistema de informação”.	A partir dos resultados retornados por cada pesquisa, foi pedido que fossem feitas três buscas na <i>Web</i> , usando distintas estratégias, por materiais diretamente relacionados a algum dos trabalhos retornados pela biblioteca: a primeira, a partir do título; a segunda, a partir das palavras-chave; e a última, a partir dos autores do trabalho monográfico escolhido.
Consulta 2	Fazer, a partir da interface da BDUni, uma pesquisa por palavras-chave de trabalhos monográficos; foram criadas três consultas diferentes, com os seguintes termos: “banco de dados”, “ <i>data mining</i> ” e “ <i>data warehouse</i> ”.	A partir dos resultados retornados por cada pesquisa, foi pedido que fossem feitas três buscas na <i>Web</i> , usando distintas estratégias, por materiais diretamente relacionados a algum dos trabalhos retornados pela biblioteca: a primeira, a partir do título; a segunda, a partir das palavras-chave; e a última, a partir dos autores do trabalho monográfico escolhido.
Consulta 3	Fazer, a partir da interface da BDUni, uma pesquisa por autores de trabalhos monográficos; foram criadas três consultas diferentes, com os seguintes termos: “daniel”, “bruno” e “leonardo”.	A partir dos resultados retornados por cada pesquisa, foi pedido que fossem feitas três buscas na <i>Web</i> , usando distintas estratégias, por materiais diretamente relacionados a algum dos trabalhos retornados pela biblioteca: a primeira, a partir do título; a segunda, a partir das palavras-chave; e a última, a partir dos autores do trabalho monográfico escolhido.

O questionário teve como objetivo apresentar algumas perguntas a fim de avaliar as observações levantadas quanto à avaliação das estratégias que podem ser usadas pelo serviço proposto, que são: busca pelo título, busca pelas palavras-chave e busca pelos autores. Para cada estratégia avaliada, os participantes escolheram opções que variam de 1 a 5, sendo a opção 1 considerada totalmente insatisfeito, a opção 2 insatisfeito, a opção 3 indiferente, a opção 4 satisfeito e a opção 5 totalmente satisfeito. Por fim, foi feita aos participantes uma pergunta, onde questionava se o participante indicaria o serviço de busca na *Web*, pedindo para justificar sua resposta.

Tais experimentos foram muito importantes, pois, obtendo a opinião de usuários reais quanto à interface do

serviço de busca na *Web* para a BDUni, pôde-se perceber que algumas melhorias precisariam ser feitas como, por exemplo, melhor tratamento dos caracteres especiais das respostas retornadas pelo Google Scholar antes delas serem exibidas ao usuário. Além disso, analisando as respostas dos participantes, notou-se o quanto as dúvidas e sugestões são fundamentais para o processo de melhoria da interação entre a aplicação e o usuário. Muitos usuários destacaram no questionário a importância e utilidade do serviço de busca na *Web*, pois facilita o trabalho de pesquisa e coleta de documentos científicos, na hora de realizar trabalhos científicos próprios.

A partir das respostas dos participantes, foi possível então avaliar o nível de satisfação dos mesmos em relação a cada estratégia utilizada pelo serviço de busca na *Web*.

As Fig.5, Fig.6 e Fig.7 apresentam os resultados da avaliação para cada critério de busca na *Web*: pelo título, pelas palavras-chave e pelos autores, respectivamente.

Avaliação da estratégia busca na web pelo título

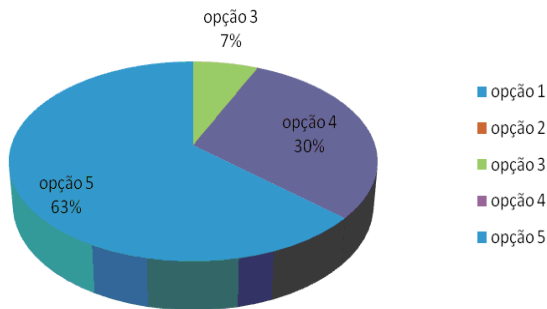


Fig.5: Gráfico de satisfação pelo critério busca na *Web* pelo título.

De acordo com a Fig.5, observa-se que o serviço de busca na *Web*, utilizando como critério de busca o título de algum trabalho monográfico catalogado na BDUi, obteve um alto grau de satisfação por parte dos participantes, onde 93% responderam como estando satisfeito ou totalmente satisfeito com as respostas retornadas pelo serviço proposto. Apenas 7% dos usuários consideraram indiferente as respostas retornadas pelo serviço de busca na *Web*, não havendo assim nenhuma resposta abaixo de 3 para o critério avaliado.

Avaliação da estratégia busca na web pelas palavras-chave

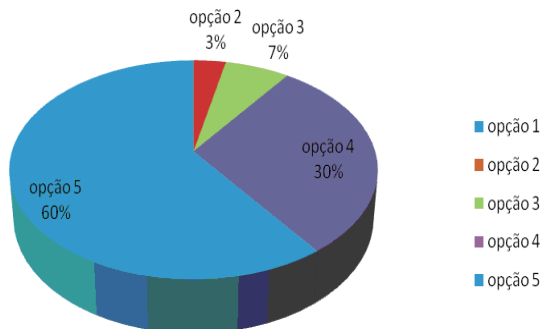


Fig.6: Gráfico de satisfação pelo critério busca na *Web* pelas palavras-chave.

De acordo com a Fig.6, pode-se observar que o serviço de busca na *Web*, utilizando como critério de busca as palavras-chave de algum trabalho monográfico catalogado na BDUi, também obteve um alto índice de satisfação por parte dos participantes, onde um total de 90% das

respostas indicaram usuários satisfeitos ou totalmente satisfeitos em relação as respostas retornadas pelo serviço de busca na *Web*. Apenas 7% responderam como estando indiferentes em relação as respostas retornadas e somente 3% das respostas indicaram a insatisfação dos usuários quanto as respostas retornadas, o que aponta uma excelente aprovação do critério.

Avaliação da estratégia busca na web pelos autores

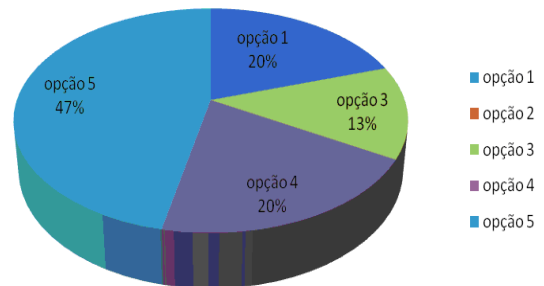


Fig.7: Gráfico de satisfação pelo critério busca na *Web* pelos autores.

De acordo com a Fig.7, observa-se que as respostas foram bem variadas para o serviço de busca na *Web*, utilizando como critério de consulta os autores de algum trabalho científico catalogado na BDUi: 67% das respostas indicaram os participantes estando satisfeitos ou totalmente satisfeitos com as respostas retornadas pelo serviço de busca na *Web*, 13% das respostas indicaram os participantes estando indiferentes em relação as respostas retornadas pelo serviço proposto, e 20% se disseram totalmente insatisfeitos em relação aos itens retornados. Considerando apenas este critério, pode-se observar que ele obteve alto índice de aprovação por parte dos participantes mas, se comparado com os dois critérios anteriores, percebe-se uma sensível diminuição no nível de satisfação por parte dos participantes.

A menor satisfação dos participantes em relação a busca na *Web*, a partir do critério autores, pode ser explicada pelo fato de o termo autores de um trabalho monográfico ser menos genérico na *Web* em relação ao seu título ou suas palavras-chave, o que resulta em respostas da máquina de busca que nem sempre serão diretamente relacionadas ao trabalho monográfico catalogado na BDUi. Na verdade, poucos são os autores de trabalhos monográficos que possuem outras publicações disponibilizadas na *Web*.

Para finalizar o questionário, foi feita a seguinte pergunta: você recomendaria o serviço de busca na *Web* da BDUi para localizar outros documentos relativos a publicações científicas? A Fig.8 apresenta o resultado para tal pergunta.

Você recomendaria o serviço de busca na web da BDUi para localizar outros documentos relativos a publicações científicas?

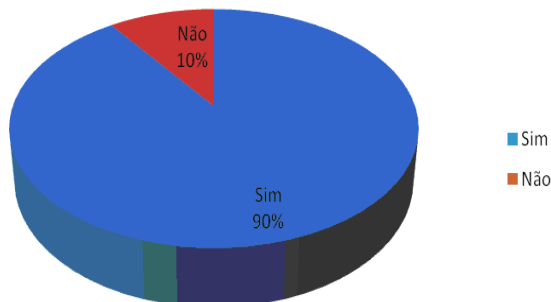


Fig.8: Gráfico de reposta sobre a avaliação final do serviço proposto.

De acordo com a Fig.8, pode-se observar que a aceitação dos participantes, em relação ao serviço de busca na *Web*, foi extremamente favorável: 90% dos participantes responderam que indicariam o serviço para fazer novas consultas, e apenas 10% responderam que não indicariam o serviço de busca na *Web*.

Alguns pontos positivos foram citados pelos participantes, como: o fato do serviço eliminar respostas retornadas pelo Google Scholar que não são referentes a trabalhos científicos (como matérias comerciais e currículos pessoais), as respostas retornadas serem diretamente associados ao assunto pesquisado, o serviço ser de fácil utilização e consistir em uma ótima forma de buscar trabalhos científicos. Existem algumas melhorias que foram sugeridas pelos participantes, tais como: tratamento de caracteres especiais, antes de mostrar os resultados, resultando em melhor entendimento das respostas retornadas pelo serviço proposto, e uma melhor forma de se fazer a busca pelo critério autores, que nem sempre retorna resultados satisfatórios.

5 CONSIDERAÇÕES FINAIS

O serviço proposto proporciona uma melhora significativa na qualidade dos serviços oferecidos por uma BD. Os usuários são beneficiados, pois possuem a oportunidade de buscar, de forma automática na *Web*, documentos diretamente relacionados aos trabalhos científicos catalogados em uma BD.

Assim, com a realização deste trabalho, foi permitido verificar a eficácia e os benefícios que pode trazer um serviço de busca na *Web*. Através dos experimentos realizados por meio da BDUi, constatou-se que o serviço, como um todo, possui um alto índice de aceitação por parte dos participantes. Todos os itens avaliados obtiveram sempre bons resultados.

A manutenção do serviço proposto de busca na *Web*, por meio de uma BD, é algo extremamente importante, já que está na sua versão inicial. Por isso, pretende-se, como um dos trabalhos futuros, a incorporação, no serviço de busca na *Web*, das técnicas de *stemming* que utilizam, ao invés das palavras inteiras, os seus radicais, ou seja, ao invés de utilizar “comparação” e “comparar” em uma dada busca, pode-se utilizar “compar”. A implementação dessa técnica poderá permitir a obtenção de resultados mais relevantes. Além disso, é sugerida também a técnica de remoção de *stopwords*, caracterizada pela retirada das palavras tais como “de”, “ou”, “e”, “para”, entre outras, no processo de busca na *Web*. Essa técnica, uma vez implementada, permitirá o aumento da precisão dos resultados retornados. Além das técnicas descritas acima, pode-se também pesquisar novos métodos de calcular a similaridade das respostas em relação ao trabalho científico desejado, sendo que, para isso, pode-se considerar outros critérios, como resumo dos artigos e descrições dos resultados das máquinas de busca.

6 REFERENCIAS BIBLIOGRÁFICAS

- [1] A. J. C. Silva. “Estratégias Para a Busca do Texto Completo de Artigos Catalogados em uma Biblioteca Digital”. Dissertação de Mestrado, Universidade Federal de Minas Gerais. Belo Horizonte, Minas Gerais, Setembro 2007.
- [2] H. S. Santos. “Implementação de um Serviço para a Busca do Texto Completo de Artigos Catalogados em uma Biblioteca Digital”. Projeto Orientado de Graduação, Universidade Federal de Minas Gerais. Belo Horizonte, Minas Gerais, Setembro 2007.
- [3] C. M. C. da Silva. “Construção de um protótipo para a Biblioteca Digital de Computação do Centro Universitário de Belo Horizonte”. Projeto Orientado de Graduação, Centro Universitário de Belo Horizonte. Belo Horizonte, Minas Gerais, Setembro 2007.
- [4] C. M. C. da Silva, G. T. de Assis. “Proposta de uma Biblioteca Digital para Trabalhos Monográficos”. Revista Científica do Departamento de Ciências Exatas e Tecnologia do UNI-BH. Belo Horizonte, Vol. 1, Nº1(2008).
- [5] M. S. Trindade. “Construção de um protótipo para a Biblioteca Digital do Centro Universitário de Belo Horizonte”. Projeto Orientado de Graduação, Centro Universitário de Belo Horizonte. Belo Horizonte, Minas Gerais, Setembro 2007.

[6] M. A. C. Ribeiro. “*Serviço de Auto Arquivamento para a Biblioteca Digital do Centro Universitário de Belo Horizonte*”. Projeto Orientado de Graduação, Centro Universitário de Belo Horizonte. Belo Horizonte, Minas Gerais, Setembro 2007.

[7] C. S. B. Gonçalves. “*Projeto e Análise de Sistemas de Busca na Web*”. Dissertação de Doutorado, Universidade Federal de Minas Gerais. Belo Horizonte, Minas Gerais, Fevereiro de 2007.