

Protótipo de Serviço de Busca para a Web baseado em Contextos, Sintagmas e Correlações

Fábio Augusto Procópio de Paiva¹ and Cláudio Rodrigues Muniz da Silva²

¹ Universidade Federal do Rio Grande do Norte, Brasil, Natal/RN, fabiopaiva@yahoo.com

² Universidade Federal do Rio Grande do Norte, Brasil, Natal/RN crmsilva@dee.ufrn.br

Abstract

The exponential expansion of Internet resources in the latest years leveraged significantly the use of the concept of Search Engine to a relevance never seen before. Search Engines were designed with the main objective of helping users in the process of finding resources in the sea of information of the Internet. Since then, many improvements were made in the concept of search engine design and implementation in order to make it better in speed and in accuracy and get better satisfaction indexes. Even though, the web users still receive many responses in their searches that differ significantly from what they are really looking for. The IRiX (Information Retrieval in Context) technique approach contributes in this context, since it enable the users to improve their searches using relationships and contexts. This paper present a search engine web service prototype based in IRiX that is designed to work in several scenarios idealized to enable better filtering characteristics with the use of relationships among user profiles, linguistic gatherings and contexts. One of these scenarios is the use of the service by a human user and the other is the use of it by an Internet web service client application.

Key words: Web Search Engines, Context, Linguistic gatherings, Correlation, Web service.

1 Introdução

Durante uma conferência em 2004, Tim O'Reilly anunciou a segunda geração da Web (Web 2.0) como sendo um conjunto de tecnologias que permitiriam a utilização de diversos serviços web através de colaborações, compartilhamentos e da participação de diversos usuários em uma plataforma definida pela rede Internet [1]-[4]. Algum tempo depois, o jornalista John Markoff propôs uma terceira geração que tentaria ir mais além e visaria transformar a Web em uma grande base de dados global. Esta evolução da Web e sua crescente popularização influenciaram também o que conhecemos como Máquinas de Buscas. As máquinas de buscas surgiram com o objetivo de auxiliar os usuários na pesquisa de informação, de modo rápido, eficiente e organizado e, neste sentido, o conceito inicial evoluiu para tentar acompanhar as novas tecnologias usadas nas várias gerações da Web. No entanto, hoje ainda são freqüentes as situações em que os usuários web formulam consultas e recebem resultados que não satisfazem a sua intenção inicial de pesquisa.

Apesar de o objetivo das máquinas de buscas ser auxiliar os usuários nas pesquisas de informações, a literatura atual descreve uma série de dificuldades encontradas pelos usuários em utilizá-las eficientemente pelo fato das interfaces utilizadas não serem totalmente amigáveis ou pelo fato dos resultados apresentados nas respostas ainda estarem distantes das reais intenções das consultas [6]. Por exemplo, consideremos o caso de um estudante de Sistemas de Comunicações Móveis que formula uma consulta com a palavra "célula". Muito provavelmente entre os resultados obtidos pela máquina de busca usada, o estudante observará links para recursos que são da área de Biologia, ao invés de receber apenas os links de Telecomunicações desejados em sua área de estudo. A situação é tão evidente que em uma pesquisa realizada por uma empresa especializada em marketing em sites de busca revelou que apenas 12,80% dos participantes mostraram-se plenamente satisfeitos com os resultados obtidos pelas ferramentas de busca [7].

Este artigo apresenta um protótipo de uma ferramenta de recuperação da informação baseado na técnica IRiX (Information Retrieval in Context), que permite que informações relacionadas a um determinado tema – por exemplo, palavras-chave, títulos, frases ou parágrafos extraídos a partir de um texto – possam ser atribuídas à consulta formulada originalmente pelo usuário, possibilitando, desta forma, melhores resultados. O protótipo faz uso de contextos construídos através de um conjunto de sintagmas e de suas correlações que são definidas pelo usuário. Esse conjunto de informações, apesar de estarem inicialmente associados ao seu usuário criador, pode ser compartilhado também com agentes de software a fim de promover a socialização dos contextos.

Vários trabalhos vêm sendo desenvolvidos a partir da técnica de recuperação da informação baseada em contexto. Alguns exemplos são: Inquirus 2, meta-máquina de busca que faz uso de informações de contexto para escolher a(s) máquina(s) onde será executada a consulta [8]; Autonomy's Kenjin lê o conteúdo de arquivos em edição ou em leitura e, a partir do conteúdo desses arquivos, sugere páginas ou arquivos locais que tenham o conteúdo semelhante aos que estão abertos [9]; Intellizap é uma meta-máquina de busca que constrói contextos a partir de trechos selecionados pelo usuário em um texto [10]. O trabalho proposto neste artigo se diferencia dos citados acima pelo fato de permitir que o próprio usuário se encarregue pela montagem e pela configuração dos seus perfis de pesquisa. Além disso, o protótipo permite que os contextos criados pelos seus usuários sejam compartilhados com outros usuários e outros softwares.

O artigo está estruturado da seguinte maneira: a seção 2, é dividida em subseções que descrevem os conceitos de máquinas de busca, serviços web, contextos, sintagmas e correlações; na seção 3, são analisados os requisitos necessários para a implementação do protótipo; na seção 4, é apresentada a implementação do protótipo e, por fim, na seção 5, são apresentadas as conclusões obtidas neste artigo.

2 Conceitos Iniciais

Com o crescimento do número de páginas web, sentiu-se a necessidade de criação de algum mecanismo que auxiliasse os usuários na tarefa de pesquisar informação nos conteúdos disponíveis na Internet. Para atender a esta demanda, foram desenvolvidas diversas ferramentas com o propósito de tornar menos complexo o processo de busca. As ferramentas de busca web são classificadas em 3 tipos: Diretórios por Assunto, Máquinas de Busca e Meta-máquinas de Busca. No entanto, neste trabalho, só abordaremos Máquinas de Busca.

2.1 Máquinas de Busca

Surgiram quando a quantidade de páginas disponíveis na Internet começou a dificultar a seleção manual e a navegacional dos recursos [11], como é feita nos diretórios por assunto. As páginas armazenadas em sua base de

dados são localizadas e analisadas por um programa chamado aranha. Porém, a análise dos recursos selecionados corre o risco de não ser tão bem realizada quanto se fosse feita manualmente.

As máquinas de busca possuem características particulares em relação à forma como os documentos são pesquisados, armazenados e manipulados. Apesar disto, existem tarefas que são comuns para a maioria das máquinas de busca:

- Descoberta – consiste em percorrer a Web em busca de recursos para armazená-los em sua base de dados;
- Indexação – criar índices de palavras que foram localizadas pelas aranhas com seu respectivo caminho de acesso além de outras informações;
- Pesquisa – permitir pesquisas de recursos armazenados em sua base de dados a partir de consultas de palavras ou de combinações.

Outro ponto em comum a se considerar são os elementos que compõem uma máquina de busca: aranha, indexador, motor de busca e interface. A Figura 1 representa, de uma forma genérica, uma máquina de busca. O controlador é um programa responsável por inicializar os serviços executados pelas aranhas e pelo indexador. Quando o controlador inicializa as aranhas, elas iniciam uma busca na Web à procura de recursos que comporão a base de dados. Quanto ao indexador, quando inicializado, ele efetua a indexação dos recursos identificados pelas aranhas e os armazena na base de dados. Quando o usuário deseja realizar uma pesquisa, ele formula uma consulta e a envia para a interface que, por sua vez, formata esta consulta enviando-a para o motor de busca. O motor de busca pesquisa na base de dados recursos que atendam àquela consulta. Depois de localizados, o motor de busca faz a ordenação dos recursos e os entrega à interface que monta uma listagem e apresenta ao usuário.

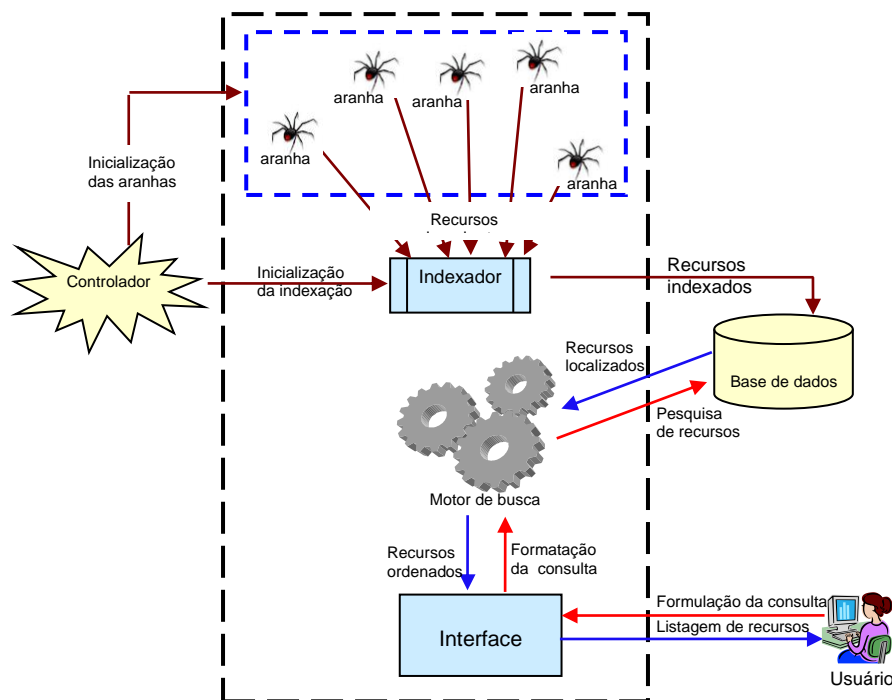


Figura 1: Representação genérica de uma máquina de busca

2.2 Serviços Web

Os Serviços Web, em inglês, Web Services, são uma tecnologia utilizada para realizar a integração de sistemas e permitir a comunicação entre diversas aplicações, mesmo que elas sejam desenvolvidas em plataformas diferentes. Normalmente, são APIs (Application Programming Interface) que podem ser acessadas via rede e executadas remotamente em um cliente requisitante de um serviço [12].

A arquitetura de um Web Service, representada na Figura 2, é formada por três componentes:

- Provedor de Serviços – responsável por disponibilizar os serviços para que possam ser utilizados pelos clientes;
- Registro de Serviços – local onde o provedor de serviços pode disponibilizar os seus serviços;
- Requisitante de Serviços – cliente que utiliza qualquer serviço disponibilizado pelo provedor de serviços.

São várias as vantagens em se utilizar web services [12]: Processamento distribuído, Ambiente distribuído, Processamento do lado do servidor e Padronização no desenvolvimento. Ver figura 2 a seguir:



Figura 2 – Arquitetura de um Serviço Web

2.3 Contextos

Atualmente, o termo contexto tem sido bastante empregado na área de máquinas de busca. Por conta disto, vê-se que existem várias definições para a mesma palavra, apesar da existência de pequenas diferenças entre cada uma delas [13].

Aqui, contexto é um conjunto de termos que possui uma relação com um determinado tema. Essas informações – que podem ser palavras-chave, títulos, frases ou parágrafos extraídos a partir de um texto – são atribuídas à consulta original do usuário.

Com o uso da técnica de recuperação da informação baseada em contextos, surgiu o termo “pesquisa contextualizada” que consiste em realizar pesquisas utilizando informações obtidas a partir de um conjunto de termos.

2.4 Sintagmas e Correlações

A maioria dos sistemas de recuperação da informação utiliza a palavra como forma de acesso à informação. Contudo, a palavra apresenta uma característica denominada polissemia. Essa característica determina que uma mesma palavra pode ter vários significados, permitindo, assim, que o resultado gerado pelos sistemas que usam esta forma de acesso à informação seja comprometido. Abaixo, são listados alguns exemplos de polissemia com o uso da palavra “aloha”:

- Um dos primeiros métodos de acesso para redes de computadores;
- Partido político americano;
- Saudação havaiana.

Percebe-se, então, que a palavra pode perder o seu real significado se for analisada fora do seu contexto original. Assim, para Kuramoto, a palavra analisada isoladamente não representa a melhor forma de acessar a informação. Por isso, como proposta para recuperar a informação com maior precisão, ele sugeriu a indexação e a recuperação da informação através de sintagmas nominais [14].

Formados por uma ou mais palavras, os sintagmas desempenham uma função determinada na frase. As palavras que formam essas frases não são posicionadas aleatoriamente, mas sim formando um conjunto (o sintagma) em torno de um núcleo. Dependendo do núcleo, o sintagma pode ser verbal ou nominal [15].

Para exemplificar, são extraídos 4 sintagmas nominais da seguinte frase “As características do meio ambiente do mundo dos negócios”:

- sintagma de nível 1: os negócios
- sintagma de nível 2: o mundo dos negócios
- sintagma de nível 3: o meio ambiente do mundo dos negócios
- sintagma de nível 4: as características do meio ambiente do mundo dos negócios

Outro conceito importante é o de correlação. As correlações representam uma relação mútua entre dois termos que podem conter mais de uma palavra. Elas são formadas por um conjunto de siglas, abreviaturas, acrônimos e símbolos relacionados aos seus respectivos significados.

Uma sigla é formada pelas letras iniciais dos vocábulos fundamentais de uma denominação ou título (exemplo: FAB, Força Aérea Brasileira), apesar de algumas vezes ser interpolada por alguma vogal para facilitar a pronúncia. Uma abreviatura é uma parte de uma palavra que é interrompida por um ponto (exemplo: “ass.”, de assinatura), mas pode acontecer de as letras não seguirem uma sequência como “pte.”, de presidente. Um acrônimo é formado pela primeira letra (ou mais de uma) de cada uma das partes sucessivas de uma locução ou pela maioria dessas partes [16] (exemplo: “radar”, de radio detecting and ranging). E por fim, símbolo é um sinal que substitui um nome de uma coisa ou de uma ação [17] (exemplo: “cm”, de centímetro).

3 Definição de Requisitos

A definição de cenários é uma tarefa indispensável para se descobrirem os principais requisitos de um sistema. Quando bem definidos, os cenários auxiliam o projetista a compreender os objetivos do projeto, bem como identificar as visões individuais de cada ator que interage com o sistema.

Assim, foram descobertas e analisadas três visões que representam a interação entre os atores e o sistema:

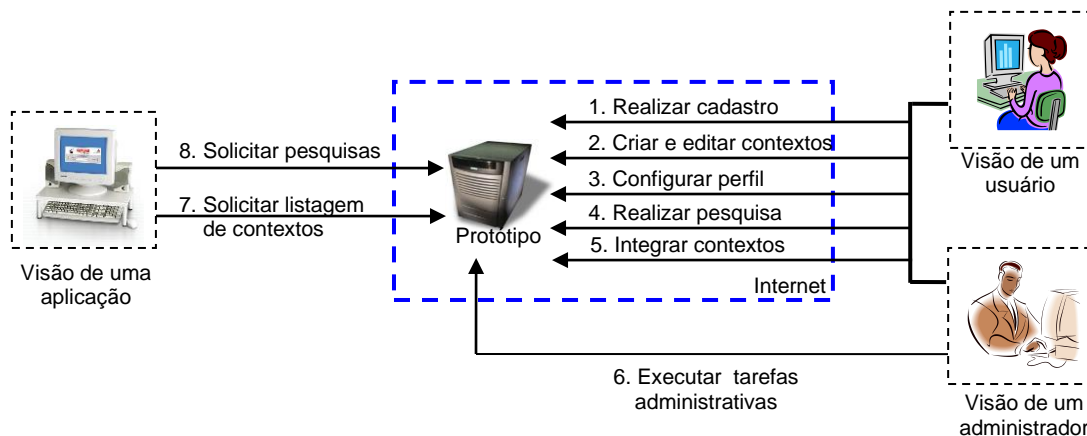


Figura 3: Cenários de uso

- **visão de um usuário** disponibiliza para o usuário as principais funcionalidades do protótipo. É nela que podem ser criados e configurados os perfis de usuários, além da execução de suas pesquisas.
- **visão de um administrador** – permite que o administrador configure as funções do protótipo. A configuração consiste na parametrização da execução de algumas funções, como também no cadastro de fontes de recursos e na criação de contas de administradores.
- **visão de uma aplicação** – possibilita que as aplicações requisitem o serviço oferecido de busca baseada em contexto e que possam obter contextos dos usuários cadastrados no protótipo.

Neste artigo só serão discutidos os cenários 3, 5 e 8, bem como os seus respectivos requisitos:

- **Cenário 3** – torna possível que o usuário configure os perfis utilizados em suas pesquisas. Para isso, é necessário que os requisitos 1, 2, 3 e 4 sejam satisfeitos.

- **Requisito 1** – o usuário seleciona, a partir de uma lista, os contextos definidos e o protótipo deve fazer a associação entre o nome do contexto selecionado e o perfil que está sendo configurado;
 - **Requisito 2** – o usuário pode definir correlações associadas ao contexto selecionada no requisito 1. É possível a criação ou edição de correlações através do par código/descrição, em que o primeiro consiste num acrônimo, uma abreviatura, uma sigla ou um símbolo; e o segundo, num significado para o código;
 - **Requisito 3** – o usuário pode definir sintagmas relacionados ao contexto do requisito 1. O processo se dá através da criação ou da edição de palavras e/ou frases que são relacionadas ao contexto;
 - **Requisito 4** – o protótipo também pode acrescentar ao contexto selecionado no requisito 1 sugestões de usuários informando URLs, as quais possuem informações que possivelmente lhes serão úteis.
- **Cenário 5** – é nesse cenário que o usuário integra ao seu perfil os contextos que foram definidos por outros usuários. Para isso, é necessário que o requisito 5 seja satisfeito.
 - **Requisito 5** – a integração de contextos requer a existência de alguns deles criados e compartilhados por outros usuários. O contexto é definido como compartilhado no momento da sua criação. Satisfeito este requisito, o cenário 5 é atendido.
 - **Cenário 8** – oferece, por meio do protótipo, um serviço de busca que atende às requisições das aplicações. Faz-se necessário, no entanto, que o requisito 6 seja seguido.
 - **Requisito 6** – o protótipo deve ser capaz de responder às requisições de outras aplicações que solicitam o serviço de busca. Para isso, ele deve interpretar as requisições enviadas, processá-las, obter os resultados que atendam a elas e disponibilizar estes para a aplicação requisitante. Dessa forma, a disponibilidade do serviço de busca para aplicações é atendida.

4 Implementação

Na implementação do protótipo, foram utilizados alguns elementos que contribuíram com a operacionalização de algumas funcionalidades, conforme mostrado na Figura 4.

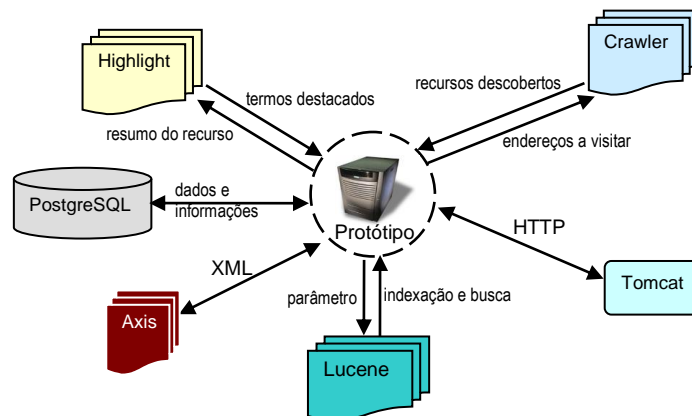


Figura 4: Relação entre o protótipo e os elementos utilizados na implementação

A biblioteca Crawler [18], realiza a tarefa de descoberta de recursos. A partir dos endereços informados pelo protótipo, a biblioteca percorre a Web em busca de novos endereços e, assim que novos endereços são descobertos, estes são enviados para a rotina de indexação do protótipo.

Lucene [19] é uma biblioteca que, baseada nos parâmetros informados para execução, permite a realização das tarefas de indexação e de busca de recursos, as quais foram discutidas na subseção Máquinas de Busca.

A biblioteca Highlight foi desenvolvida para dar suporte à Lucene. Ela, normalmente, é utilizada para extrair e destacar os trechos considerados mais relevantes para uma pesquisa requisitada.

O Axis, um framework de código aberto desenvolvido na linguagem Java e baseado no padrão XML, é utilizado para a construção de Web Service no padrão SOAP.

Para o armazenamento dos dados é utilizado o PostgreSQL e para o gerenciamento do tráfego de mensagens HTTP é utilizado o container web do projeto Apache, Tomcat.

A Figura 5 ilustra a página que permite a utilização de pesquisas baseadas nos contextos definidos no perfil do usuário que está conectado.

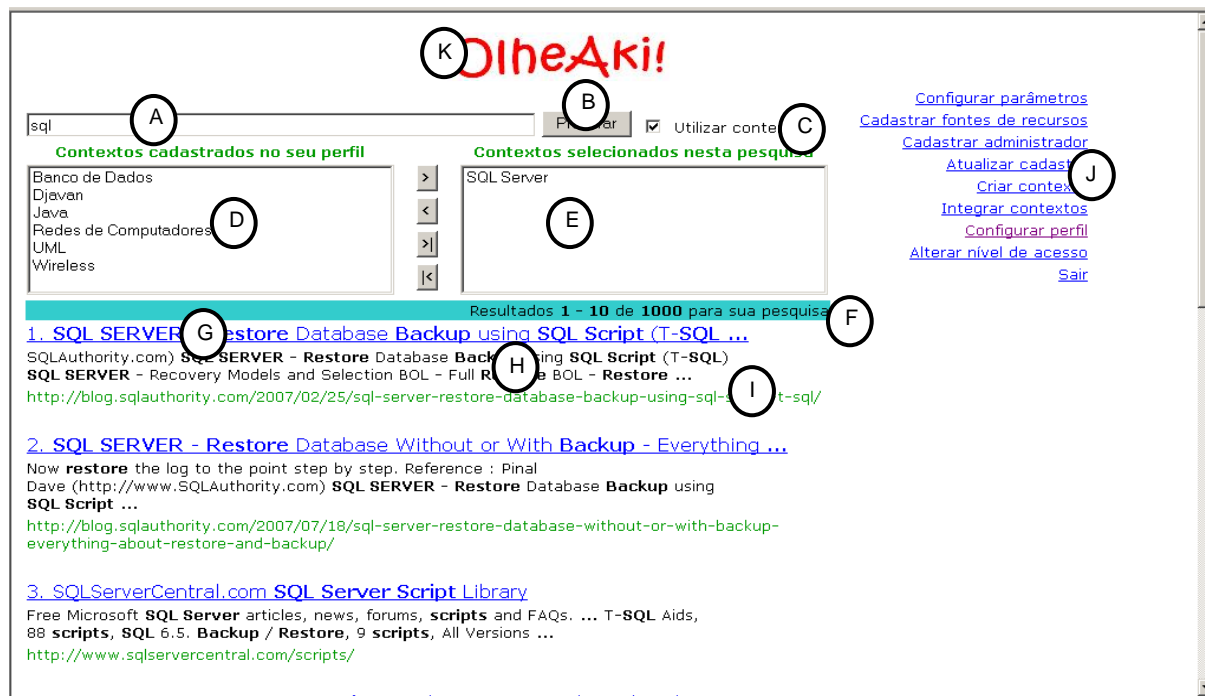


Figura 5: Página de pesquisa com a utilização de contextos

Em (A), o usuário informa as palavras-chave e/ou frases que serão utilizadas na pesquisa. No exemplo, foi informada a palavra SQL. Em (C), é informado o tipo de pesquisa que será realizado: baseada apenas em palavras-chave ou utilizando contextos.

Em (D), são listados todos os contextos configurados para o perfil do usuário, exceto SQL Server, uma vez que este foi selecionado para a consulta. Na figura, os contextos relacionados ao usuário conectado são Banco de Dados, Djavan, Java, Redes de Computadores, UML e Wireless. Já em (E), são listados os contextos selecionados para a pesquisa. No exemplo, apenas o contexto SQL Server foi selecionado. Informados esses dados, depois que o usuário pressionar o botão Procurar (B), o protótipo inicia o processo de pesquisa. Em (F), o protótipo apresenta um resumo da quantidade de documentos localizados na pesquisa. (G) representa o título do recurso localizado. No exemplo da figura, o título localizado é "SQL Server – Restore Database Backup using SQL Script (T-SQL...)". Em (H), é apresentado um pequeno resumo do recurso ilustrado em (G). (I) apresenta o endereço onde o recurso está disponível. (J) são os links que podem ser acessados a partir da página de pesquisa e (K) é o logotipo do protótipo. No exemplo ilustrado na Figura 5, a consulta enviada para o mecanismo de busca é a seguinte:

(conteudo: "SQL" AND "backup" AND "restore" AND "script" AND "SQL Server")
OR (titulo "SQL" "backup" AND "restore" AND "script" AND "SQL Server")

Listagem 1: Consulta formatada

A Listagem 1 mostra que, antes de a consulta ser enviada para o motor de busca, todas as correlações e todos os sintagmas atribuídos ao contexto SQL Server foram concatenados com a consulta original formulada pelo usuário, configurando-se, assim, uma consulta contextualizada

A Figura 6 representa o procedimento de integração de contextos compartilhados. Este processo consiste em adicionar contextos cadastrados por outros usuários ao perfil do usuário conectado ao protótipo. No exemplo, o contexto a ser importado está relacionado à área médica.

Em (A) são listados os contextos disponíveis para compartilhamento que foram cadastrados por outros usuários. Ao pressionar o botão Procurar (B), o protótipo apresenta as informações de correlações, de sintagmas e de fontes de recurso definidas inicialmente pelo usuário criador. (C), (D) e (E) apresentam, respectivamente, essas informações.

Ao pressionar o botão Integrar (F), o protótipo realiza, efetivamente, a integração do novo contexto ao perfil do usuário corrente.

O formulário 'Integrar contextos' possui uma barra de título azul com o texto 'Integrar contextos'. Abaixo, há um campo 'Contexto:' com uma lista suspensa (rotulada A) e um botão 'Procurar' (rotulado B). O formulário é dividido em três seções principais:

- Correlação:** Possui um cabeçalho com 'Código' e 'Descrição'. Abaixo, há uma lista de correlações: 'CAD : Cetoacidose diabetica' e 'IRC : Insuficiencia Renal Cronica' (rotulada C).
- Sintagma:** Possui uma lista de sintagmas: 'Hiperglicemia', 'Hipoinsulinemia' e 'Insuficiencia pancreatica' (rotulada D).
- Fontes de recurso:** Possui uma lista vazia para fontes de recurso (rotulada E).

Na base do formulário, há um botão 'Integrar' (rotulado F), um botão 'Voltar' e um botão 'Sair'.

Figura 6 : Integração de contextos

Como mostrado na Figura 3, existem dois cenários que oferecem serviços web: os cenários 7 e 8. As classes Java que realizam estas tarefas são publicadas sob a gerência do framework Axis e executadas a partir da solicitação de aplicações clientes. A Figura 7 equivale ao cenário 8, onde o protótipo de uma aplicação Java invoca o serviço de Pesquisa implementado pelo web service. A aplicação envia uma consulta e o web service executa a requisição e localiza recursos que satisfaçam à consulta.

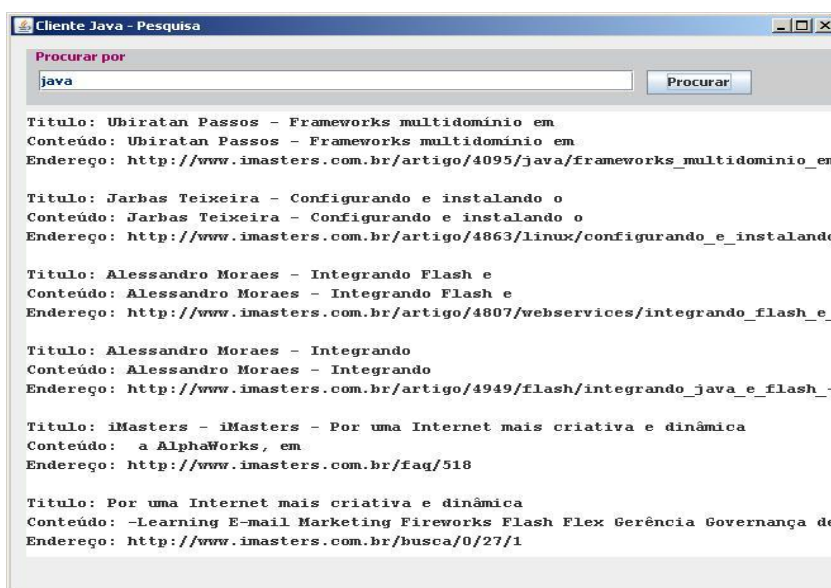


Figura 7: Aplicação cliente Java requisitando o serviço web Pesquisa

5 Conclusões

A implementação do protótipo de máquina de busca deste artigo é um passo importante para o estudo de novas tecnologias que procuram resolver os problemas dos atuais sites de busca da Internet. O protótipo apresentado é uma contribuição neste sentido, pois permite a integração de alguns conceitos próprios do que conhecemos como Web 2.0, tais como: a socialização, a gerência de perfis, o utilização de informações em XML e etc. Apesar de observarmos uma melhora das buscas em alguns contextos de uso, ainda não há garantias de que não haverá o retorno de informações redundantes ou irrelevantes. Muito provavelmente, a solução mais adequada para o problema envolve a agregação de mais inteligência na máquina de busca no sentido de interagir com outros sistemas que possam fornecer informações e meta-informações sobre os seus usuários, sobre as suas buscas e sobre as fontes para os recursos pesquisados, bem como a implementação de conceitos próprios de uma área que surgiu recentemente, conhecida como Gerência de Conhecimento. O desenvolvimento de novos trabalhos é de extrema importância por possibilitar a otimização de funções deficientes encontradas nesta versão inicial ou por viabilizar a construção de funcionalidades ainda não implementadas. Propomos algumas idéias que podem contribuir significativamente com este trabalho: indexação de recursos no formato de imagens, áudio e vídeo; localização de endereços indisponíveis; filtragem de recursos por formato e por idioma e definição automática da relevância dos termos em um contexto.

References

- [1] T. O'Reilly. (2005, September.). What is Web 2.0: Design Patterns and Business Models for the Next Generation of Software. O'Reilly [Online]. Available: <http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html>
- [2] Eric. (2006, September.). Web 2.0 History. Eric's Blog [Online]. Available: http://learningremix.net/f2006integ/enapier/2006/09/web_20.shtml
- [3] T. O'Reilly. (2006, October.). Web 2.0 Compact Definition: Trying Again. O'Reilly Radar [Online]. Available: http://radar.oreilly.com/archives/2006/12/web_20_compact.html
- [4] M. Glaser. (2006, April.). What is Web 2.0, and Should You Care? MediaShift [Online]. Available: http://www.pbs.org/mediashift/2006/04/jargon_watchwhat_is_web_20_and.html
- [5] T. Berners-Lee, J. Hendler and Ora Lassila (2001, May.). The Semantic Web: A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities. Scientific American.com [Online]. Available: <http://www.sciam.com/article.cfm?articleID=00048144-10D2-1C70-84A9809EC588EF21&pageNumber=1&catID=2>
- [6] H. Kuramoto, Sintagmas Nominais: uma Nova Proposta para a Recuperação da Informação, Ciência da Informação, vol 3, no. 1, pp. 30-43, 2002.
- [7] T. R. Bachinn, Pesquisa "O uso dos sites de busca no Brasil", Cadastra, Porto Alegre, Brasil, Technical Report 3-15, 2005.
- [8] E. J. Glover, S. Lawrence and W. P. Birmingham, Architecture of a Metasearch Engine that Supports User Information Needs, in Proceedings 8th International Conference on Information and Knowledge Management, Kansas, 1999, pp. 210-216.
- [9] E. Mulligen, M. Diwersy, B. Schijvenaars, M. Weeber, C. Eijk, R. Jelier, M. Schuemie, J. Kors and B. Mons, Contextual Annotation of Web Pages for Interactive Browsing, in Proceedings of the 11th World Congress on Medical Informatics, San Francisco, 2004.
- [10] L. Finkelstein, E. Gabriolovich and Y. Matias, Placing Search in Context: The Concept Revisited, in Proceedings 10th International WWW Conference, Hong-King, 2001.
- [11] B. V. Cendón, Ferramentas de busca na Web, Ciência da Informação, vol 30, no. 1, pp. 39-49, 2001.
- [12] R. A. D. Girardi, Framework para coordenação e mediação de Web Services modelados como Learning Objects para ambientes de aprendizado na Web, M.S. thesis, Pontifícia Universidade Católica, Rio de Janeiro, RJ, Brasil, 2004.
- [13] S. Lawrence, Context in Web Search, IEEE Data Engineering Bulletin, vol 23, no. 3, pp. 25-32, 2000.
- [14] H. Kuramoto, Uma abordagem alternativa para o tratamento e a recuperação de informação textual: os sintagmas nominais, Ciência da Informação, vol 25, no. 2, pp. , 1995.
- [15] M. A. Perini, Gramática descritiva do português. São Paulo: Ática, 1995.
- [16] A. B. H. Ferreira, Novo Dicionário Aurélio - Século XXI. Rio de Janeiro: Nova Fronteira, 1999.
- [17] P. A. Furaste, Normas Técnicas para o Trabalho Científico: Explicitação em Normas da ABNT. Porto Alegre: 2003.
- [18] L. Torunski. (2004, January.). Smart and Simple Web Crawler. JAVA.NET [Online]. Available: <https://crawler.dev.java.net>
- [19] O. Gospodnetic and E. Hatcher, Lucene in Action. Greenwich: Manning, 2005.
- [20] S. B. Palmer. (2001, September.). The Semantic Web: An Introduction. Infomesh.net [Online]. Available: <http://infomesh.net/2001/swintro>