



Ministério da Saúde

FIOCRUZ

Fundação Oswaldo Cruz



Instituto de Comunicação e Informação
Científica e Tecnológica em Saúde

Especialização em Informação Científica e Tecnológica em Saúde

**IMPLEMENTAÇÃO DE UM TESAURO COMO PROPOSTA PARA MELHORAR A
RECUPERAÇÃO DE INFORMAÇÃO NO PORTAL FIOCRUZ**

Rodrigo Gomes Ferrari Cesar

Orientador: Carlos Henrique Marcondes

Rio de Janeiro, 2007



Ministério da Saúde
FIOCRUZ
Fundação Oswaldo Cruz



CURSO DE ESPECIALIZAÇÃO EM INFORMAÇÃO CIENTÍFICA E TECNOLÓGICA EM SAÚDE

IMPLEMENTAÇÃO DE UM TESAURO COMO PROPOSTA PARA MELHORAR A RECUPERAÇÃO DE INFORMAÇÃO NO PORTAL FIOCRUZ

por

RODRIGO GOMES FERRARI CESAR

Fundação Oswaldo Cruz (FIOCRUZ)

Projeto apresentado ao Instituto de Comunicação e Informação Científica e Tecnológica em Saúde da Fundação Oswaldo Cruz como requisito parcial para obtenção do título de Especialista em Informação Científica e Tecnológica em Saúde.

Orientador: Carlos Henrique Marcondes, doutor em Ciência da Informação

Rio de Janeiro, dezembro de 2007

SUMÁRIO

1. INTRODUÇÃO	4
2. JUSTIFICATIVA	6
3. REFERENCIAL TEÓRICO	10
4. OBJETIVOS	15
4.1. Objetivos gerais	15
4.2. Objetivos específicos	15
5. METODOLOGIA	16
6. RESULTADOS ESPERADOS	18
7. BIBLIOGRAFIA CONSULTADA	19
8. CRONOGRAMA	22
9. ORÇAMENTO	23
ANEXOS	24

1. INTRODUÇÃO

O Portal Fiocruz foi inaugurado em maio de 2005 tendo como objetivo possibilitar o acesso a informações qualificadas sobre ciência, tecnologia e saúde. Seu conteúdo está também disponível em inglês e espanhol. Embora, antes disso, a Fundação Oswaldo Cruz já mantivesse no ar um *site* na internet, este não contava com uma interface administrativa para publicação de conteúdo. Sua manutenção e atualização aconteciam de forma estática via FTP¹, sem um sistema de administração que permitisse a inserção e/ou alteração de informações.

Entre o primeiro e segundo ano do novo formato do Portal Fiocruz, o número de acessos dobrou. Isso impactou diversos setores dentro da instituição, dentre eles o canal Fale Conosco que teve um aumento de 229,15% em mensagens no mesmo período. Conseqüentemente, não apenas os funcionários da instituição, mas também diversos setores da sociedade puderam opinar, criticar e sugerir caminhos para esse produto, tão em evidência.

Com a incorporação do Portal Fiocruz à estrutura organizacional da fundação e com a chegada de novos profissionais através do concurso público do final ano passado, novos desafios e possibilidades se apresentaram, tais como: melhoria na usabilidade e acessibilidade e; integração com outros sistemas previamente desenvolvidos e em franco uso.

Através de estatísticas de acesso e do Fale Conosco, identificamos que o usuário que acessa o Portal tem dificuldades para encontrar informação em seu conteúdo, o que nos motivou a propor melhorias no sistema de organização do conhecimento do Portal Fiocruz, e dentre elas, a incorporação de um vocabulário controlado para recuperação dos conteúdos publicados no portal da Fundação Oswaldo Cruz.

Nesse sentido, por várias bibliotecas da Fiocruz utilizarem os Descritores em Ciências da Saúde em sua indexação, a adoção por parte do Portal Fiocruz

¹ FTP é a sigla para *File Transfer Protocol* (Protocolo de Transferência de Arquivos), e é usado para transferir dados de um computador a outro através da internet ou de outra rede.

destes mesmos descritores possibilitaria uma recuperação única mais abrangente.

Atualmente o Portal Fiocruz utiliza o software Publique![®], um gerenciador de conteúdo operado através da internet. Este ambiente conta com um sistema de autoria que segue um fluxo hierárquico de criação, aprovação e publicação de informações entre autores e editores, além de um sistema de busca com filtros por data, seção, autor, entre outros.

Tal ferramenta permite a adição de novas funcionalidades, chamadas “componentes”, e é o desenvolvimento e implementação de um componente para indexação e recuperação do conteúdo no ambiente *web* que este trabalho propõe.

Com isso visamos integrar as bases de dados do Portal Fiocruz e dos demais *sites* das unidades a outras já existentes, além de aumentar a qualidade de indexação dos conteúdos inseridos, facilitando a recuperação da informação e melhorando a experiência dos usuários.

[®] Marca registrada pela empresa Fábrica Digital

2. JUSTIFICATIVA

A partir de meados de 2005, juntamente com a reformulação da arquitetura do conteúdo existente e a implantação de um gerenciador de conteúdo, o Portal Fiocruz passou a oferecer o Fale Conosco, um canal institucional de diálogo com a sociedade. Ao disponibilizar tal serviço, além de atender o usuário, a coordenação do Portal Fiocruz – atualmente sob minha responsabilidade – passa a contar com um importante instrumento para aferição das demandas da sociedade e, conseqüentemente, aquelas que não estão contempladas (ou pelo menos não de maneira adequada) no Portal. Por atender individualmente o usuário, recebendo suas críticas, sugestões e dúvidas, o Fale Conosco gera relatórios com dados qualitativos que nos permitem avaliar e fazer uma gestão mais adequada do conhecimento.

Em 2006 o Fale Conosco recebeu e respondeu 7178 mensagens, relacionadas às diferentes áreas da instituição. A área que apresenta a maior demanda é a de Ensino / Cursos, e no referido ano correspondeu a 21,26% das mensagens recebidas. Outra que se destaca é a área de Bibliotecas / Referências Bibliográficas, com 12,89%.

De acordo com a equipe responsável pelo Fale Conosco, cerca de 2/3 das mensagens dessas duas áreas solicita informações que já estão contempladas dentro do Portal Fiocruz.

Para ilustrar, listamos algumas das mensagens recebidas pelo Fale Conosco, com nossas respectivas respostas:

USUÁRIO: Prezados senhores, estou buscando o estatuto dessa Fiocruz, com interesse nas atividades da área de pesquisa.

FALE CONOSCO: Em resposta à sua solicitação, sugerimos que acesse o Portal Fiocruz (<http://www.fiocruz.br/>) e clique em “Conheça a Fiocruz”. Estão disponibilizados os *links* para o relatório de atividades, o relatório de gestão e o estatuto da instituição. Para obter o regimento interno, acesse http://www.fiocruz.br/media/regimento_interno.pdf.

USUÁRIO: Minha informação é bem singela. Gostaria de saber em quais estados da Federação encontramos unidades da Fiocruz. É só no Rio e Bahia? Não encontrei essa informação no *site*. Obrigada.

USUÁRIO: Vocês têm algum centro de pesquisa em Saúde Pública em São Paulo?

FALE CONOSCO: As unidades da Fiocruz estão localizadas no Rio de Janeiro, Salvador, Belo Horizonte, Manaus, Recife, Curitiba e Brasília. Para conhecer a Fiocruz, acesse <http://www.fiocruz.br/>, clique em "Conheça a Fiocruz" e em "Institutos".

USUÁRIO: Gostaria de saber como faço para receber as edições dos periódicos da Fiocruz.

FALE CONOSCO: Informamos abaixo os procedimentos para conhecer os periódicos publicados pela Fiocruz e fazer a assinatura:

- Acesse o Portal Fiocruz (<http://www.fiocruz.br/>);
- Clique no item “Comunicação e Informação”;
- Clique no item “Periódicos”. Você encontrará links para os periódicos Cadernos de Saúde Pública, História, Ciências, Saúde - Manguinhos, Memórias do Instituto Oswaldo Cruz e Trabalho, Educação e Saúde;
- Clique também nos itens “Radis” e “Canal Saúde”.

Concomitantemente, instalado na mesma época da inauguração do Portal Fiocruz, o analisador de *logs* AWStats monitora todo o acesso via HTTP² ao servidor, gerando estatísticas. Dentre as medições estão: o número de visitas e

² HTTP é a sigla em inglês para *HyperText Transfer Protocol* (Protocolo de Transferência de Hipertexto) utilizado para transferência de dados na rede mundial de computadores.

de *unique visitors*³, assim como a duração delas; dias e horários de maior tráfego; localidades dos provedores visitantes; sistemas operacionais e navegadores utilizados para acessar o Portal; acessos oriundos de ferramentas de busca, além das palavras nelas utilizadas para encontrar o Portal Fiocruz.

Se analisarmos as estatísticas de buscas no Portal Fiocruz, verificamos, por exemplo, que no ano de 2006 o termo “conjuntivite” foi repetidamente buscado pela grafia errada “conjutivite”. Curiosamente, por um outro erro de digitação, havia conteúdo inserido dentro do Portal com essa mesma grafia, fazendo com que o termo fosse encontrado. Portanto, a inexistência de um vocabulário controlado permitiu um duplo erro. Tanto a indexação do conteúdo foi feita incorretamente, como a busca do usuário recuperou apenas o texto que continha a palavra errada. Já os usuários que digitaram corretamente “conjuntivite” foram privados da recuperação completa dos conteúdos que tratavam sobre o tema.

Além disso, alguns termos são característicos de uma determinada região. “Dengue”, por exemplo, é também conhecida como “Febre Quebra-Ossos”. Uma indexação através de um vocabulário estruturado permitiria fazer um relacionamento entre os dois termos, de modo que um usuário que quisesse recuperar informação com apenas um deles obteria o conteúdo indexado de ambos.

Este conjunto de dados nos indica que, mais que uma simples percepção, o usuário do Portal tem dificuldades para (em diversos casos não consegue) encontrar conteúdo dentro do ambiente *web* da Fiocruz, motivando-nos a propor melhorias em seu sistema de organização do conhecimento; dentre elas, a incorporação de um vocabulário controlado para recuperação dos conteúdos publicados.

Como várias bibliotecas que compõe a Rede de Bibliotecas da Fiocruz utilizam a mesma base de descritores (DeCS), uma integração com o Portal Fiocruz permitiria a comunicação transparente entre os diferentes sistemas, possibilitando uma recuperação mais abrangente e única dos termos procurados (Figura 1). Tal

³ *Unique visitor* ou visitante único é uma estatística que descreve uma unidade de tráfego de um *site*, contando cada visitante apenas uma vez dentro de um período de tempo definido. Dessa forma, um *unique visitor* pode fazer várias visitas neste intervalo, sendo o acesso contabilizado uma única vez. Esta estatística é relevante como medida da verdadeira audiência, equivalente ao termo “*reach*” utilizado em Publicidade.

argumento é reforçado pela recente inauguração da Biblioteca Virtual Fiocruz, tendo como um de seus objetivos “desenvolver interfaces ‘amigáveis’, conciliando a identidade visual da BVS com a do Portal Fiocruz” (FIOCRUZ, 2007).

3. REFERENCIAL TEÓRICO

No contexto de um portal *web*, entende-se conhecimento como o conteúdo existente dentro dele, sendo um sistema de organização deste conhecimento a ferramenta que ajuda o leitor usuário a encontrar algo neste portal.

Muito se realiza no que diz respeito às ferramentas de gestão do conteúdo e a cada dia mais recursos são desenvolvidos para administrar as fontes de informação. Entretanto, os instrumentos de representação e recuperação desta informação não avançam na mesma velocidade.

Dentre os instrumentos de representação e recuperação, identificamos alguns (PARREIRAS, 2004):

- Pastas – Organizam os objetos de maneira hierárquica, servindo como uma árvore de navegação, além de ser bastante intuitivo, já que utilizam uma simbologia equivalente a dos gerenciadores de arquivos GUI⁴.
- *Links* – São instrumentos de relacionamento dos objetos, pelos quais se entende que existe uma relação, embora sem necessariamente se saber qual.
- Busca textual – Geralmente é feita nos metadados relacionados ao documento, assim como no próprio texto do documento.
- Associações por metadados – Associam um descritor a objetos diferentes, conferindo algo em comum entre eles.
- Vocabulário controlado – Ao se inserir um conteúdo, utiliza-se um dicionário que o associa a descritores.
- Tesouros – Conceitos estruturados em uma lista de termos relacionados entre si, ordenados de tal forma a permitir uma recuperação eficaz.

Embora auxiliem a organização do conhecimento, tais instrumentos não oferecem a flexibilidade e os recursos fornecidos pelas ontologias.

⁴ Sigla do inglês *Graphical User Interface* (Interface Gráfica do Usuário). É um mecanismo de interação homem-computador que permite selecionar símbolos e manipulá-los de forma a obter algum resultado prático.

- Ontologias – São uma formalização sistemática de conceitos, definições, relacionamentos e regras que capturam o conteúdo semântico de um domínio em um formato legível pela máquina. Pode ser utilizado por softwares de computador no gerenciamento automático de conhecimento. Tem como principais características: as relações semânticas entre termos são explicitamente declaradas e; as regras de inferência conectam declarações tais que deduções podem inferir – logicamente – outras.

Ainda segundo Parreiras (2004), a aplicação de ontologias apresenta outras formas de organizar o conhecimento, alterando positivamente a experiência do usuário.

Tesauros não chegam a ser ontologias: na ontologia, os termos, além de uma hierarquia, têm propriedades a partir das quais um software pode fazer inferências.

Embora um tesauro tenha algumas limitações como (MURAMAKI, 2005 *apud* SOERGEL, 2004):

- Cobertura semântica limitada – Não reflete relacionamentos conceituais que poderiam ser usados por um sistema para sugerir conceitos que expandam ou especifiquem.
- Falta de consistência – Como a semântica é imprecisa, eles podem criar ambigüidade.
- Processos automatizados limitados – Tradicionalmente eles são projetados para indexar e formular consultas por pessoas e não para processamento automatizado.

Há vários pontos positivos:

- Linguagem única – Vocabulário controlado para indexação e recuperação das informações e conteúdos (artigos, matérias, vídeos, áudios e documentos).

- Relacionamentos - Indicam relacionamentos semânticos entre os conteúdos.
- Hierarquia - Hierarquização em um sistema de navegação que ajuda os usuários a localizar objetos desejados.

A literatura dá indicações de que um vocabulário estruturado melhora a experiência do usuário.

Resultados utilizando o Tesouro Jurídico da Justiça Federal sugerem que expansão de consultas baseada em relacionamentos do tesouro melhora o resultado da busca. Tal estudo concluiu que a utilização de sinônimos e termos específicos acarreta em melhoria dos resultados obtidos (SILVEIRA, 2003).

O uso de tesouros é defendido como uma importante ferramenta para a busca de informação eletrônica, por organizarem termos e associarem conceitos em redes semânticas (BRASCHER, 2002 *apud* DOERR, 2001; HUNTER, 2001), provendo uma forma mais consistente de navegar no Portal Fiocruz.

Enfim, “o uso de vocabulários controlados melhora a qualidade e disponibilidade como suporte para buscas em várias bases de dados” (MURAMAKI, 2005 *apud* SHIRI, 2000).

Baseado nessas evidências, este projeto propõe a inclusão de um tesouro no Portal Fiocruz. Entretanto, ao invés de criarmos um tesouro próprio para a instituição, nossa proposta é utilizarmos o DeCS – Descritores em Ciências da Saúde – para tal fim, pelos seguintes motivos:

- O processo de construção de tesouros é complexo e de custo elevado, tanto de tempo quanto de recursos humanos (MURAKAMI, 2005);
- É estruturado e trilingue proporcionando um meio consistente e único para a recuperação da informação independentemente do idioma;
- Foi desenvolvido pelo Centro Latino-Americano e do Caribe de Informação em Ciências da Saúde (BIREME) para uso na indexação de artigos de

revistas científicas, livros, anais de congressos, relatórios técnicos, e outros tipos de materiais, assim como para ser usado na pesquisa e recuperação de assuntos da literatura científica;

- Além dos termos médicos originais do Medical Subject Headings (MeSH) da U.S. National Library of Medicine, foram desenvolvidas as áreas específicas de Saúde Pública, Ciência e Saúde e Vigilância Sanitária, fundamentais para a Fiocruz (DECS, 2006);
- Familiaridade dos usuários (profissionais e instituições) com a terminologia;
- É o mesmo vocabulário utilizado nas bases de dados dos acervos das bibliotecas da Fiocruz;
- O DeCS é a terminologia padrão utilizada por todos os componentes do Sistema Latino-Americano e do Caribe de Informação em Ciências da Saúde, em 36 países da região para o tratamento e disseminação de fontes de informação em saúde desde 1982;
- A Fundação Oswaldo Cruz integra o Comitê Consultivo Nacional da BVS Saúde Pública responsável pelo projeto de desenvolvimento e atualização da terminologia em saúde pública.

Além dos motivos listados acima, verificou-se junto à BIREME que o DeCS poderia se adequar à tecnologia RDF/XML. *Resource Description Framework* (RDF⁵) é uma sintaxe para modelagem de informação que utiliza “triplas”. Tripas são frases (em XML⁶) sobre os recursos na forma "sujeito-predicado-objeto", onde o sujeito é o recurso enquanto que o predicado, as características ou aspectos do recurso e expressa um relacionamento entre sujeito e objeto.

O RDF é um poderoso mecanismo em Web Semântica, onde um software pode armazenar, trocar e distribuir informação legível por máquina através da internet.

Assim, por exemplo, a tecnologia RDF permitiria que o sistema identificasse pessoas numa galeria de fotos no Portal Fiocruz usando informação de uma lista

⁵ Tecnologia endossada e recomendada pela W3C tendo como principais objetivos criar um modelo de dados, com uma semântica formal com suporte para o uso de XML.

⁶ *Extensible Markup Language* (XML) é uma linguagem de marcação que permite aos seus usuários definir suas próprias *tags*. Sua função principal é facilitar o compartilhamento de dados estruturados em diferentes sistemas de informação através da Internet.

de contatos; e automaticamente disparasse um e-mail para essas pessoas informando que suas fotos estão disponíveis na internet.

4. OBJETIVOS

4.1. Objetivo geral

Este projeto tem como objetivo principal servir como linguagem única para indexação e recuperação das informações e conteúdos existentes no Portal Fiocruz, permitindo um diálogo uniforme entre os *sites* da fundação e estimulando a disseminação e integração dos conhecimentos e práticas de saúde.

4.2. Objetivos específicos

- Integrar o DeCS - Descritores em Ciências da Saúde - ao gerenciador de conteúdos (Publique!®) utilizado no Portal Fiocruz e em diversos *sites* de unidades, departamentos e projetos da Fundação Oswaldo Cruz.
- Prover um vocabulário controlado para indexação e recuperação de artigos, matérias, vídeos, áudios e documentos;
- Promover a consistência na designação de termos de indexação;
- Indicar relacionamentos semânticos entre os conteúdos, de forma que uma matéria possa levar a outras que tratam de assuntos relacionados.

5. METODOLOGIA

Como metodologia, o projeto propõe:

- Análise tecnológica do DeCS e do formato RDF;
- Desenvolvimento de um protótipo da ferramenta que realizará a integração do DeCS ao Portal Fiocruz. Tal protótipo prevê os templates das telas de busca e de recuperação dos termos, além de uma interface de indexação no sistema de administração;
- Após a implementação, acreditamos necessário o treinamento dos profissionais responsáveis pela inserção do conteúdo no Portal Fiocruz;
- Avaliação da implantação do sistema através de relatórios e reuniões entre os participantes;
- Levantamento das estatísticas de acesso e do Fale Conosco;
- Divulgação do uso e impacto da tecnologia.

Propomos uma implementação modular do componente, a mais abrangente possível – independente de plataforma ou tecnologia – para que, a partir do momento em que esteja em funcionamento, possa ser incorporado sem dificuldades em qualquer outro gerenciador de conteúdo que porventura venha a ser adotado pela instituição. Além disso, deve visar outras implementações de novos tesouros.

Estudos preliminares nos mostraram que o mais indicado seria implementar uma solução baseada em RDF. Por isso, este projeto recomenda que o DeCS passe a utilizar tal *framework*, o que, futuramente, possibilitaria sua migração para uma ontologia.

Como projetos futuros oriundos da implementação do DeCS no Portal Fiocruz, acreditamos importantes:

- Estudar a possibilidade de uso de um indexador automático que sugira palavras-chaves. Avançar no estudo dos trigramas do DeCS;
- Criar um laboratório de usabilidade, para aferir o impacto na interação causado pela implementação de um tesouro no Portal Fiocruz;
- Verificar a necessidade de se ter outros vocabulários em paralelo ao DeCS. Isso será possível com a análise dos mesmos canais utilizados para a argumentação deste projeto, acrescido da proposta da criação do laboratório de usabilidade (item acima);
- Analisar periodicamente a qualidade de indexação do conteúdo.

6. RESULTADOS ESPERADOS

- Maior facilidade de acesso às informações no Portal Fiocruz com possibilidade de recuperação das informações contidas nele e em outros sites da fundação através de vocabulário controlado;
- Dinamização do acesso, incluindo integração com as Bibliotecas Virtuais em Saúde instaladas na Fiocruz e fora dela;
- Revisão do sistema de navegação do Portal Fiocruz, hierarquizando, de forma consistente, para que ajude usuários a localizar objetos de conteúdo desejados, facilitando a navegação.
- Versão diferenciada do gerenciador de conteúdo (Publique!®) oferecendo um vocabulário controlado integrado à ferramenta (uma interface possível está proposta na Figura 2 dos ANEXOS deste projeto);
- *Plug-in* (componente) portátil para utilização em outros sistemas.

7. BIBLIOGRAFIA CONSULTADA

ASSUMPÇÃO, R. T. **Recuperação de documentos jurídicos baseada em um tesouro**. 2001. 52f. Dissertação (Mestrado em Ciências da Computação). Departamento de Ciência da Computação, Universidade Federal de Minas Gerais, Belo Horizonte, 2001.

BERNERS-LEE, T.; HENDLER, J.; LASSILA, O. The semantic web. **Scientific American**, p. 35-43, mai. 2001. Disponível em: <<http://www.sciam.com/article.cfm?articleID=00048144-10D2-1C70-84A9809EC588EF21>>. Acesso em 21 out. 2007.

BIREME - CENTRO LATINO-AMERICANO E DO CARIBE DE INFORMAÇÃO EM CIÊNCIAS DA SAÚDE. **DeCS - Descritores em Ciências da Saúde**. Disponível em: <<http://decs.bvs.br/P/decswebp2006.htm>>. Acesso em 21 out. 2007.

BRÄSCHER, M. A ambigüidade na recuperação da informação. **DataGramZero**, Rio de Janeiro, v.3, n.1, fev. 2002. Disponível em <http://www.dgz.org.br/fev02/Art_05.htm>. Acesso em: 21 out. 2007.

CAZALENS, S. *et al.* A Web site indexing process for an Internet information retrieval agent system. In: WEB INFORMATION SYSTEMS ENGINEERING, 1., 2000, Hong Kong. **Proceedings of the First International Conference on Web Information Systems Engineering**. Washington: IEEE Computer Society, 2000. p.254 – 258.

DIAS, C. A. Portal corporativo: conceitos e características. **Ciência da Informação**, Brasília, vol. 30, n. 1, p. 50-60, abr. 2001.

DOERR, M. Semantic problems of thesaurus mapping. **Journal of Digital Information**, Southampton, v. 1, n. 8, mar. 2001. Disponível em: <<http://jodi.ecs.soton.ac.uk/Articles/v01/i08/Doerr/>>. Acesso em: 12 mar. 2001.

FUNDAÇÃO OSWALDO CRUZ. **Biblioteca Virtual Fiocruz**. Disponível em: <<http://bvsfiocruz.fiocruz.br>>. Acesso em 8 nov. 2007.

_____. **Portal Fiocruz**. Disponível em: <<http://www.fiocruz.br/>>. Acesso em: 23 out. 2007.

HEDDEN, H. A-Z Indexes to Enhance Site Searching. **Digital Web Maganize**, jan. 2005. Disponível em <http://www.digital-web.com/articles/a_z_indexes_site_searching/>. Acesso em: 21 out. 2007.

HUNTER, J. MetaNet: a metadata term thesaurus to enable semantic interoperability between metadata domains. **Journal of Digital Information**, Southampton, v. 1, n. 8, 2001. Disponível em: <<http://jodi.ecs.soton.ac.uk/Articles/v01/i08/Hunter/>>. Acesso em: 12 mar. 2001.

MINISTÉRIO DA SAÚDE. **Portal da Saúde**. Brasília, 2007. Disponível em: <http://portal.saude.gov.br/saude/>. Acesso em: 21 out. 2007.

MORAIS, E. F., SOARES, M. B. Web Semântica para Máquinas de Busca. **Curso de Pós-graduação em Ciência da Computação**. Belo Horizonte: UFMG, 2004. Disponível em <<http://homepages.dcc.ufmg.br/~nivio/cursos/pa03/seminarios/seminario7/seminario7.pdf>>. Acesso em: 21 out. 2007.

MORVILLE, P. **Ambient findability**. O'Reilly. 2005.

MORVILLE, P.; ROSENFELD, L. **Information architecture for the World Wide web - designing large-scale web sites**. O'Reilly. 2003.

MURAKAMI, T. M. **Tesauros e a World Wide Web**. 2005. 75p. Trabalho de conclusão de curso (Bacharelado em Biblioteconomia e Documentação). Escola de Comunicações e Arte, Universidade de São Paulo, São Paulo, 2005.

_____. Uso de Tesauros na World Wide Web. **Connexions Web Site**. nov. 30, 2005. Disponível em: <<http://cnx.org/content/m12999/1.15/>>. Acesso em: 21 out. 2007.

PARREIRAS, F. **Ontologias fazem portal corporativo avançar**. Webinsider. 1 jul. 2004. Disponível em: <<http://webinsider.uol.com.br/index.php/2004/07/01/ontologias-fazem-portal-corporativo-avancar/>>. Acesso em: 10 out. 2007.

RIBEIRO-NETO, B.; ASSUMPCÃO, R. T. Recuperação de documentos jurídicos baseada em um tesouro. In: XVI SIMPÓSIO BRASILEIRO DE BANCO DE DADOS, Brasília, out. 2001.

SHIRI, A. A.; REVIE, C. Thesauri on the Web: current developments and trends. **Online Information Review**, v. 24, n.4, p. 273-279, 2000.

SILVEIRA, M de L. Recuperação Vertical de Informação: Um Estudo de Caso na Área Jurídica. 2003. 114f. Tese (Doutorado em Ciência da Computação). Universidade Federal de Minas Gerais, Belo Horizonte, 2003. Disponível em: <<http://www.pbh.gov.br/prodabel/cde/publicacoes/2003/silveira2003a.pdf>>.

Acesso em: 21 out. 2007.

SIMÃO, J. B.; RODRIGUES, G. Acessibilidade às informações públicas: uma avaliação do portal de serviços e informações do governo federal. **Ciência da Informação**, Brasília, v. 34, n. 2, p. 81-92, mai./ago. 2005.

SOERGEL, D. *et al*, Reengineering Thesauri for New Applications: the AGROVOC Example, **Journal of Digital Information**, Southampton, v. 4, n. 4, 2004. Disponível em: <<http://journals.tdl.org/jodi/article/view/jodi-126/111>>. Acesso em: 19 jun. 2007.

SOUZA, R. R. Sistemas de recuperação de informações e mecanismos de busca na web: panorama atual e tendências. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 11 n. 2, p. 161-173, mai./ago. 2006. Disponível em: <http://www.scielo.br/scielo.php?pid=S1413-99362006000200002&script=sci_arttext>. Acesso em: 21 out. 2007.

NATIONAL LIBRARY OF MEDICINE. **UMLS Unified Medical Language System**. 2007. Disponível em: <<http://www.nlm.nih.gov/research/umls/>>. Acesso em: 21 out. 2007.

W3C. **OWL Web Ontology Language**. 2004. Disponível em: <http://www.w3.org/TR/owl-features/>. Acesso em: 21 out. 2007.

_____. **RDF Resource Description Framework**. 2007. Disponível em: <<http://www.w3.org/RDF/>>. Acesso em: 21 out. 2007.

WODTKE, C. Unraveling the Mysteries of metadata and taxonomies. **Boxes and Arrows**. 2002. Disponível em: <http://www.boxesandarrows.com/view/unraveling_the_mysteries_of_metadata_and_taxonomies>. Acesso em: 21 out. 2007.

8. CRONOGRAMA

ID	Macro-atividades	Duração (semanas)	Meses														
			1	2	3	4	5	6	7	8	9	1	1	1			
1	Formação da equipe e parcerias	3	■														
2	Revisão de literatura	7	■	■													
3	Análise tecnológica DeCS e RDF	9		■	■												
4	Implementação do DeCS no Publique	28				■	■	■	■	■	■	■	■				
5	Layout de busca e interface Publique	10									■	■					
6	Desenvolvimento dos templates	10									■	■	■				
7	Protótipo	0													x		
8	Testes e ajustes	6														■	
9	Treinamento	4														■	■
10	Entrega do componente	0															x

9. ORÇAMENTO

<i>Atividades</i>	<i>Valores (R\$)⁷</i>
Design (ambiente de busca completa, tela de resultado de recuperação, interface do sistema de administração)	4.700,00
Desenvolvimento client (templates)	4.200,00
Desenvolvimento server (base de dados e sistema)	19.900,00
Diárias (12) e passagens (6)	2.900,00
Custo total estimado	31.700,00

⁷ Valores para design e desenvolvimento baseados na Bolsa de Salários (Folha SP, 4/11/2007): webdesigner (R\$ 2667), programadores sistemas de informática júnior (R\$ 1914) e sênior (R\$ 2316). Para as diárias e passagens considerou-se o custo diário de R\$ 100,00 além de R\$ 289,50 para cada passagem Rio-São Paulo (TAM, 4/11/2007).

ANEXOS

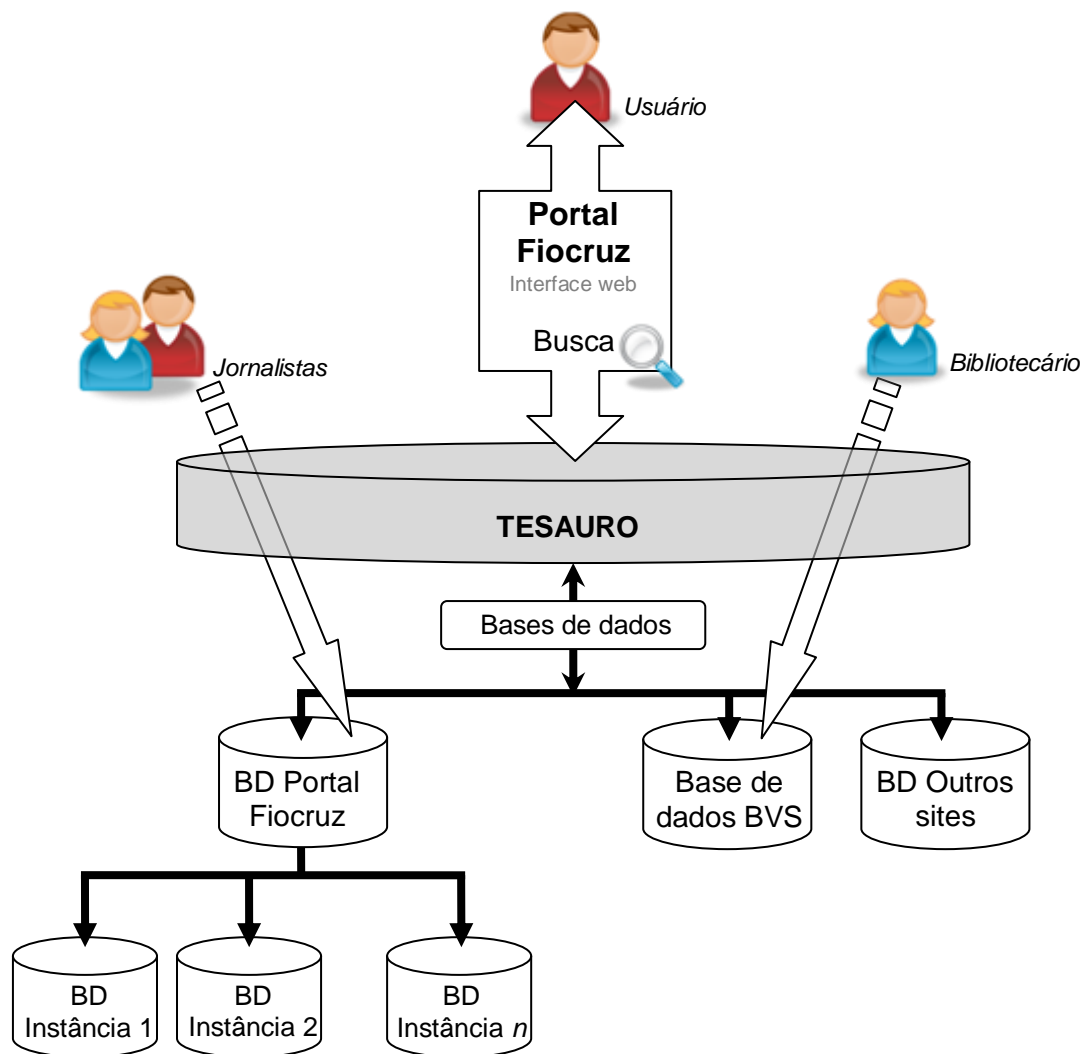


Figura 1 – Esquema de fluxo de inserção e recuperação em bancos de dados compartilhando um mesmo tesauro

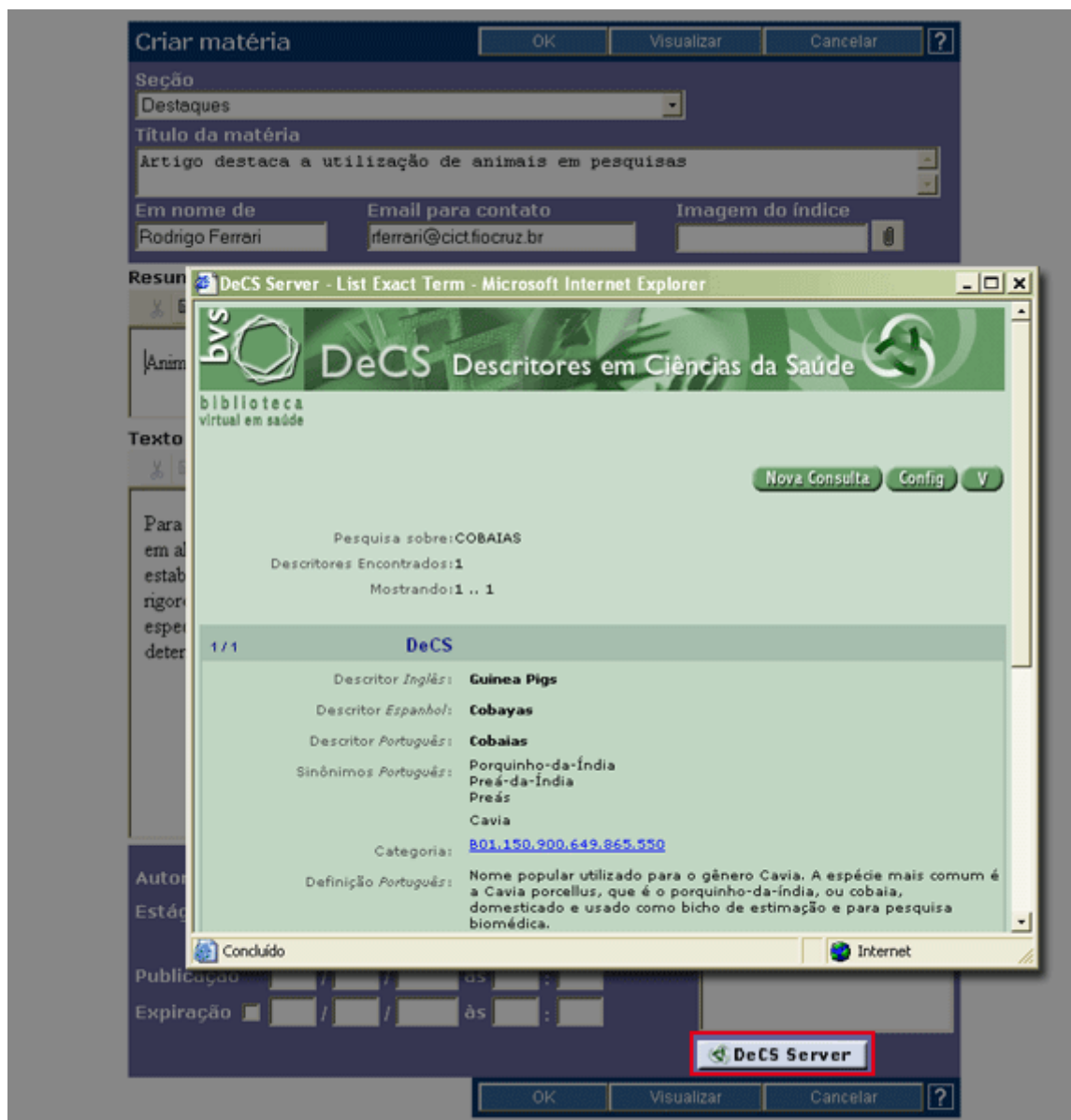


Figura 2 – Possível interface entre o DeCS e o gerenciador de conteúdo Publique!