



## **Portal Semântico sobre Automedicação: um instrumento para prevenção e orientação à comunidade**

Divino Ignacio Ribeiro Jr.  
Alessa Fabiola Santos

**Resumo:** Automedicação é a decisão de consumir, sem avaliação médica, um medicamento sem conhecimento sobre como utilizá-lo, conforme Tourinho, Bucarechi et al. (2007) que reforçam a necessidade de uma política pública para a definição de intervenções e estratégias de promoção da saúde, e nesse sentido, a promoção de pesquisas e projetos voltados para a promoção da saúde com a educação sobre o uso de medicamentos. Assim foi proposta a construção de um Portal Semântico sobre publicações em periódicos especializados, sobre medicamentos, substâncias químicas e informações toxicológicas relacionadas e informações de origem institucional. A abordagem teórico-metodológica é multidisciplinar, com o vértice situado na área da Ciência da Informação e com aportes metodológicos na Ciência da Computação e Engenharia do Conhecimento. Essas áreas trouxeram significativas contribuições para a Ciência da Informação, corroborando Souza e Alvarenga (2004) quando afirmam que as tecnologias e métodos podem ser aplicados em projetos de novos motores de busca, na construção automática de vocabulários e tesouros, na indexação automática de documentos, criação de Portais Semânticos (JORGE, 2005), entre outras aplicações. Os processos de Engenharia do Conhecimento são baseados Gómez-Pérez, Fernández-López et al. (2004, p.107), usados para o desenvolvimento da ontologia de domínio do sistema. O projeto resultou num portal de consulta acessível pela Internet, composto por: a) interface de busca; b) mecanismo de busca semântica; c) ontologia de domínio; d) fontes de informação: documentos sobre substâncias químicas e artigos científicos obtidos por importação de suas bases de dados.

## **INTRODUÇÃO**

A OMS - Organização Mundial da Saúde (1985) caracteriza o uso racional de medicamentos (URM) quando “os pacientes recebem medicamentos apropriados para suas condições clínicas, em doses adequadas às suas necessidades individuais, por um período adequado e ao menor custo para si e para a comunidade”. Como exemplos de uso inapropriado de medicamentos podem ser citados: a poli farmácia; antimicrobianos



em posologias inadequadas ou para infecções não bacterianas; uso excessivo de medicamentos injetáveis, quando há disponibilidade de apresentações farmacêuticas orais mais apropriadas; prescrição em desacordo com diretrizes clínicas e, automedicação inapropriada .

Os resultados de uma pesquisa realizada por Tourinho, Bucarechi *et al.* (2007) reforçam a necessidade de uma política pública para a definição de intervenções e estratégias de promoção da saúde, visando: a) o uso racional de medicamentos; b) a prevenção da automedicação que possa trazer riscos aos usuários e à comunidade; c) o armazenamento e descarte seguros dos medicamentos; d) a prevenção do desperdício e e) a prevenção de acidentes domiciliares em crianças, deixando, de fato, os medicamentos fora do alcance destas, bem como, instituindo a obrigatoriedade legal das embalagens de proteção. Neste sentido, a promoção de pesquisas e projetos voltados para a promoção da saúde com a educação sobre o uso de medicamentos pode auxiliar prevenção da e os fatores a ela associados, sendo ferramentas úteis para colaborar com o uso racional de medicamentos.

É nesse contexto que foi definida a proposta desta pesquisa: criar e desenvolver um Portal Semântico sobre Automedicação, concebido como meio auxiliar no processo de prevenção à automedicação.

## **FERRAMENTAS E MÉTODOS UTILIZADOS NESTA PESQUISA**

O desenvolvimento de um Portal Semântico é fundamentado em métodos e técnicas aplicados no contexto da Web Semântica e da Engenharia do Conhecimento. Esta área trouxe significativas contribuições para a Ciência da Informação, conforme Souza e Alvarenga (2004), pois suas tecnologias e métodos podem ser aplicados em projetos de novos motores de busca, na construção automática de vocabulários e tesouros, na indexação automática de documentos, criação de Portais Semânticos (JORGE, 2005), entre outras aplicações.



Um Portal Semântico é composto pelos seguintes componentes: a) Ontologia de domínio, que é a representação do conhecimento sobre o domínio em questão, definido como uma especificação formal de uma conceitualização compartilhada, um modelo abstrato representando conceitos, propriedades, restrições, relações, axiomas e regras, conforme Borst (1997); Gruber (1993); Studer, Benjamins *et al.* (1998). Assim, se aplicam processos que definam uma atividade de engenharia, e não uma atividade informal e intuitiva, conforme Gómez-Pérez, Fernández-López *et al.* (2004, p.107); b) o mecanismo de busca semântica, que é um artefato computacional capaz de localizar conceitos da ontologia ou realizar algum processo de inferência para identificar uma determinada classe de conceito. Ele difere do tradicional por buscar elementos de significado do conceito desejado, e não apenas a localização de *strings*; c) repositório de conteúdos: é o banco com as fontes de informação mapeadas com a ontologia, como a localização de documentos, de repositórios de notícias, entre outros; d) a interface de busca semântica, que consiste no meio pelo qual ele irá realizar as pesquisas.

Fundamentalmente, as etapas do desenvolvimento de um Portal Semântico podem ser definidas como se segue: a) análise de Requisitos do Portal e do Projeto da Ontologia; b) desenvolvimento da Ontologia. Esta etapa é orientada pela metodologia METHONTOLOGY que, conforme Gómez-Pérez, Fernández-López *et al.* (2004), se resume na seleção dos termos do domínio, construção da taxonomia de conceitos, criação de diagramas e relações entre os conceitos, definição das características das classes e mapeamento das instâncias (elementos reais do domínio que farão parte da ontologia); c) desenvolvimento do Mecanismo de Busca Semântica e d) desenvolvimento da Interface do Portal integrada ao Mecanismo de Busca e o Repositório;

A etapa 'b' utiliza vocabulários controlados e tesouros existentes na criação dos conceitos, relações e outros itens da ontologia, pois estes recursos podem ser utilizados como fontes no seu desenvolvimento, por meio do uso da informação terminológica (conceitos e relações entre estes). Os tesouros são de especial importância na constituição terminológica de uma ontologia, pois são construídos objetivando a organização, acesso e recuperação de conceitos com termos de um determinado

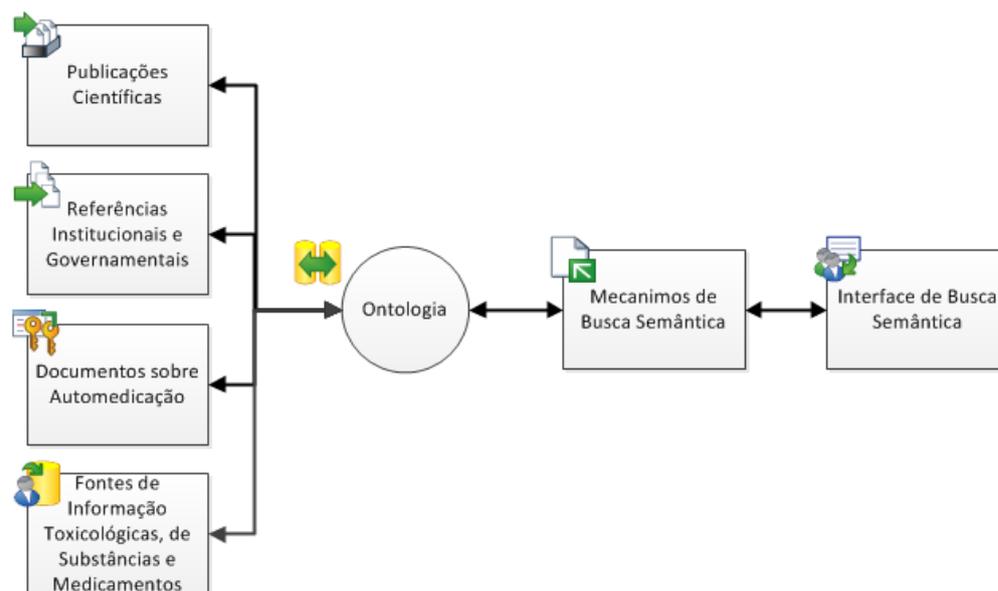


domínio. Porém, um tesouro não permite o armazenamento de regras de inferência e regras de definição de lógica, como por ex: conceitos interseccionados, logicamente disjuntos, equivalência lógica e agregação de conceitos.

## O PORTAL

O Desenvolvimento do Portal seguiu a arquitetura definida no projeto:

Figura 1 - Arquitetura do Portal Semântico sobre Automedicação



### Desenvolvimento da Ontologia de Domínio

Os artefatos desenvolvidos que compõe a ontologia são as estruturas conceituais das Monografias de Substâncias Químicas do Banco de Informações Toxicológicas – HSDB da NLM e o vocabulário DeCS completo:

- a) **O vocabulário DeCS** - Bireme (2013)- composto por mais de 31.000 descritores agrupados em várias categorias que incluem, além do vocabulário MeSH, as áreas de Saúde Pública, Homeopatia, Ciência e Saúde e Vigilância Sanitária. O



vocabulário DeCS foi obtido por extração do site da BIREME e está na forma de banco de dados relacional, a ser convertido para a forma de ontologia – OWL.

- b) **O Hazardous Substances Data Bank (HSDB®)** é um banco sobre toxicologia da NLM (2013) com informações de mais de 5400 itens: produtos químicos, higiene industrial, procedimentos de emergência, entre outras. Os dados são referenciados de livros, documentos governamentais, relatórios técnicos e periódicos primários selecionados. Todo o Banco HSDB foi importado e está na forma de banco de dados relacional. Foi extraído com processos de indexação automática um vocabulário para definir a ontologia inicial para vincular as monografias.

Os vocabulários controlados obtidos para construção da ontologia de domínio foram submetidos a um tratamento para identificar as relações conceituais entre eles. Isso potencializa a capacidade de recuperação semântica através da identificação de assuntos complementares que estão ligados pela estrutura conceitual da ontologia.

O DeCS foi obtido a partir do Serviço DeCS/XML, que disponibiliza os descritores em XML obtidos por meio de um *crawler* que os capturam pelos seus códigos hierárquicos. Vejamos o exemplo:

Figura 2 - Fragmento do descritor 'Bioética' na forma XML.

```
<?xml version="1.0"?>
<decsvmx version="1.0" date="20130204 213124" query="Bioética">
<decsws_response service="" tree_id="K01.559.411.070">
  <tree>
    <self>
      <term_list lang="pt">
        <term tree_id="K01.559.411.070">Bioética</term>
      </term_list>
    </self>
    <ancestors>
      <term_list lang="pt">
        <term tree_id="K">CIÊNCIAS HUMANAS</term>
        <term tree_id="K01">Ciências Humanas</term>
        <term tree_id="K01.559">Princípios Morais</term>
        <term tree_id="K01.559.411">Ética</term>
        <term tree_id="N">ASSISTÊNCIA À SAÚDE</term>
        <term tree_id="N05">Qualidade, Acesso e Avaliação da Assist
        <term tree_id="N05.350">Ética</term>
        <term tree_id="SH">CIÊNCIA E SAÚDE</term>
      </term_list>
    </ancestors>
  </tree>
</decsws_response>
</decsvmx>
```

O processo de captura e conversão foi realizada em duas etapas: a) desenvolvimento do *crawler* para captura dos descritores, sinônimos, definições, taxonomia e relacionamentos horizontais, e cada descritor é salvo em um arquivo XML.



A validação da captura é realizada por amostragem das categorias, examinando-se o conteúdo obtido e a informação correspondente fornecida no site da Serviço DeCS/XML; b) conversão dos arquivos XML salvos, por meio de um *script* que popula um banco de dados previamente preparado para este fim. Um exemplo da conversão é o termo ‘Bioética’, apresentado na figura a seguir:

Figura 3 – Ilustração do Descritor ‘Bioética’ na forma relacional.

The screenshot shows a SQL query window with the following SQL statement: `SELECT * FROM decs_descriptor;` The result set is displayed in a table with the following data:

decs_id	lang	descriptor
K01.316.034	pt	Temas Bioéticos
K01.316.070	en	Bioethics
K01.316.070	es	Bioética
K01.316.070	pt	Bioética
K01.316.070.500	en	Ethics, Clinical
K01.316.070.500	es	Ética Clínica

O processo de mapeamento de relações entre o DeCS e o vocabulário CID-10 – Classificação Estatística Internacional de Doenças e problemas relacionados à Saúde, foi realizada por meio de outro *crawler* que obtém as informações de similaridade léxica entre os termos de ambos vocabulários. Esta captura é realizada com auxílio do mecanismo de busca por trigramas da Bireme que implementa o algoritmo de busca por similaridade léxica ‘Trigram Phrase Matching’. Este processo é realizado nas seguintes etapas: a) criação de um banco de dados com o CID-10; b) criação do *crawler* que captura as informações de similaridade léxica entre o CID-10 e o DeCS na forma XML; c) conversão do conteúdo em XML na forma relacional.

O uso deste algoritmo em vocabulários controlados permite recuperar termos que estão no mesmo contexto semântico, como por exemplo, os termos “Influenza [gripe] devida a vírus identificado da gripe aviária” (CID-10) e “Vírus da Influenza A” (DeCS - B04.820.545.405.400), como se observa pelo fragmento do arquivo XML:



Figura 4 – Arquivo com similaridade léxica entre os termos.

```

<?xml version="1.0" encoding="ISO-8859-1" ?>
- <serw version="1.0" date="20100303" time="173723">
- <request>
- <trigrams doc="1">
- <parms>
  <collection path="DeCSxp" size="100648" terms="16330" width="0" maxtv="50" maxrf=
    maxrel="30" minsim="0.400000" maxsim="1.000000">DeCSxp</collection>
  <document size="63" maxrf="0.500000" maxrel="5" minsim="0.450000"
    maxsim="1.000000">Influenza [gripe] devida a vírus identificado da gripe aviária
</parms>
- <similarlist>
  <similar rank="1" s="0.704708" c="7660" h="24201">Vírus da Gripe A Aviária</similar>
  <similar rank="2" s="0.704708" c="7660" h="24199">Vírus da Gripe A Aviária</similar>
  <similar rank="3" s="0.638580" c="4347" h="13931">Gripe Aviária</similar>
  <similar rank="4" s="0.586257" c="7660" h="24200">Vírus da Influenza A Aviária</sir
  <similar rank="5" s="0.581834" c="7660" h="24197">Vírus da Influenza Aviária</simil
</similarlist>
</trigrams>
- <jdlist doc="1" jds="19">
  <jd doc="1" rank="1" f="4" w="2.577507">B04.909</jd>
  <jd doc="1" rank="2" f="4" w="2.577507">B04.820.545</jd>
  <jd doc="1" rank="3" f="4" w="2.577507">B04.909.777.545.405.400</jd>
  <jd doc="1" rank="4" f="4" w="2.577507">B04.909.777.545</jd>
  <jd doc="1" rank="5" f="4" w="2.577507">B04.820.545.405</jd>
  <jd doc="1" rank="6" f="4" w="2.577507">B04.820</jd>
  <jd doc="1" rank="7" f="4" w="2.577507">B04.909.777</jd>
  <jd doc="1" rank="8" f="4" w="2.577507">B04.909.777.545.405</jd>
  <jd doc="1" rank="9" f="4" w="2.577507">B04.820.545.405.400</jd>

```

Um segundo processo de tratamento para enriquecimento semântico no vocabulário DeCS foi testado: a identificação de novas relações entre os seus descritores.

As etapas desse processo são sintetizadas a seguir:

- Identificar no texto da Definição eventuais termos que sejam Descritores.
- Mapear esses descritores, criando relações entre o Descritores identificado na definição e do Descritores em questão;
- Armazenar numa tabela as categorias do descritor (decs\_id) e os termos encontrados na definição.

Um script preparado para varrer os registros com descritores percorre e analisa cada um, observando os seguintes componentes: Descritores, texto da descrição e o código da categoria ao qual o descritores pertence.



Figura 5 – Pontos de varredura no vocabulário

Descritor Inglês: **Stroke**

Descritor Espanhol: **Accidente Cerebrovascular**

Descritor Português: **Acidente Cerebral Vascular**

Sinônimos Português: Derrame Cerebral, Ictus Cerebral, AVC, Apoplexia, Acidente Cerebrovascular, Apoplexia Cerebral, Apoplexia Cerebrovascular, Icto Cerebral, Acidente Vascular Encefálico, AVE, Acidente Vascular do Cérebro, Acidente Vascular Cerebral

Categoria: C10.228.140.300.775, C14.907.253.855

Definição Português: Grupo de condições patológicas caracterizadas por perda súbita, não-convulsiva, da função neurológica, devido a ISQUEMIA ENCEFÁLICA ou HEMORRAGIAS INTRACRANIANAS. O acidente cerebral vascular é classificado pelo tipo de NECROSE de tecido, como localização anatômica, vasculatura envolvida, etiologia, idade dos indivíduos afetados e natureza hemorrágica versus não-hemorrágica (Tradução livre do original: Adams et al., Principles of Neurology, 6th ed, pp777-810).

**Descritor** (pointing to the Portuguese descriptor)

**Categorias** (pointing to the category codes)

**Termos encontrados na Definição que também são Descritores** (pointing to terms like ISQUEMIA ENCEFÁLICA, HEMORRAGIAS INTRACRANIANAS, NECROSE, etc. in the definition)

O resultado obtido é armazenado em um banco de dados. Considerações iniciais sobre esse processo:

- Os termos identificados não são sinônimos;
- As relações criadas registram o vínculo contextual existente entre esses descritores;
- São relações ‘horizontais’, ou seja, ligam nodos da árvore do DeCS em pontos distintos sem passar pelos caminhos definidos pelas ligações hierárquicas;
- Essas relações acontecem em quaisquer pontos das categorias.
- Os tipos e a natureza dessas relações não estão mapeados, mas isso pode ser experimentado. Ex: Mapear entre os Descritores da Categoria Doenças (C) todos os termos encontrados em suas respectivas Definições que pertencem à Categoria Anatomia (A):



Figura 6 – Relações mapeadas no vocabulário

Doenças (Descritor)			Partes da Anatomia (termos encontrados na Definição que também são Descritores)	
decs_descriptor.decs_id	lang	descriptor	related_descriptor	decs_descriptor_1.decs_id
C01.252.200.100	pt	Abscesso Encefálico	CORAÇÃO	A07.541
C01.252.200.100	pt	Abscesso Encefálico	ORELHA MÉDIA	A09.246.397
C01.252.200.100	pt	Abscesso Encefálico	PULMÃO	A04.411
C01.252.200.100	pt	Abscesso Encefálico	SEIOS PARANASAIS	A04.531.621
C01.252.200.200	pt	Empiema Subdural	ORELHA MÉDIA	A09.246.397
C01.252.200.200	pt	Empiema Subdural	SEIO ETMOIDAL	A04.531.621.267
C01.252.200.200	pt	Empiema Subdural	SEIO FRONTAL	A04.531.621.387

Por meio deste processo é possível identificar, combinando categorias específicas, relações do tipo ‘doença acomete parte do corpo’ ou outras relações que fazem sentido no contexto do sistema.

## REPOSITÓRIO DE CONTEÚDOS

Foram mapeadas e importadas a Base de Monografias de Substâncias Químicas do Banco de Informações Toxicológicas – HSDB da NLM e os periódicos da Saúde disponíveis no SCIELO. Os artigos foram obtidos com um extrator através do recurso de artigo em formato XML do Scielo e reindexados com base no DeCS para vinculação conceitual com a Ontologia.

O processo de coleta dos dados dos artigos no Scielo utiliza o Serviço Personalizado ‘Artigo em XML’. Esse serviço retorna, de maneira estruturada, os metadados e o conteúdo do artigo que já está na versão HTML e PDF na base do Scielo, detalhado na seguinte forma: A URL do material no Scielo contém um código, passado por um parâmetro chamado PID, composto por:

- ISSN
- Ano da Publicação dentro do ISSN



- Número sequencial que define o Número do Fascículo dentro do Ano da Publicação (não considera o Volume)
- Número sequencial que define o Número da Publicação dentro do Número do Fascículo



Exemplos:

A Revista *Brazilian Journal of Infectious Diseases* Vol 15 pertence ao ano 2011, e tem 6 fascículos. Assim a sequencia dos códigos PID fica assim:

Fascículo 1: de pid=S1413-86702011000100001 até pid=S1413-86702011000100021  
 Fascículo 2: de pid=S1413-86702011000200001 até pid=S1413-86702011000200021  
 Fascículo 3: de pid=S1413-86702011000300001 até pid=S1413-86702011000300023

Baseado nisso, a **primeira etapa do processo** (que consiste na definição do conjunto de itens que serão coletados) é feita da seguinte maneira:

1. Acessar o fascículo inicial do primeiro volume do ano inicial desejado e copiar o primeiro e o último link dos itens desejados;
2. Repetir o procedimento para cada fascículo dentro do período desejado
3. Cada link copiado deve ser colado em uma lista, que pode ser elaborada num editor de textos ou planilha, um link por linha



Figura 7 - Estrutura do artigo disponibilizado na forma XML pelo Scielo.

```
<?xml version="1.0" encoding="ISO-8859-1"?>
<article xmlns:mml="http://www.w3.org/1998/Math/MathML" xmlns:xlink="http://www.w3.org/1999/xlink">
  <front>
    <journal-meta>
      <journal-id>1413-8670</journal-id>
      <journal-title><![CDATA[Brazilian Journal of Infectious Diseases]]></journal-title>
      <abbrev-journal-title><![CDATA[Braz J Infect Dis]]></abbrev-journal-title>
      <issn>1413-8670</issn>
      <publisher>
        <publisher-name><![CDATA[Brazilian Society of Infectious Diseases]]>
      </publisher>
    </journal-meta>
    <article-meta>
      <article-id>S1413-86702011000200001</article-id>
      <article-id pub-id-type="doi">10.1590/S1413-86702011000200001</article-id>
      <title-group>
        <article-title xml:lang="en"><![CDATA[Survey of Mycoplasma pneumoniae]]>
      </title-group>
      <contrib-group>
        <contrib contrib-type="author">
          <name>
            <surname><![CDATA[Hadi]]></surname>
            <given-names><![CDATA[Nahal]]></given-names>
          </name>
        </contrib>
      </contrib-group>
    </article-meta>
  </front>
</article>
```

O quadro a seguir relaciona os periódicos usados nos teses do projeto e as quantidades de artigos obtidos de cada um:

Quadro 1 - Periódicos usados para compor o repositório de conteúdos

ISSN	Revista	Qtde
0102-8650	Acta Cirurgica Brasileira	758
0365-0596	Anais Brasileiros de Dermatologia	285
0001-3765	Anais da Academia Brasileira de Ciências	381
0066-782X	Arquivos Brasileiros de Cardiologia	998
0004-2730	Arquivos Brasileiros de Endocrinologia & Metabologia	680
0004-	Arquivos Brasileiros de Oftalmologia	834




---

2749		
0004- 2803	Arquivos de Gastroenterologia	321
0004- 282X	Arquivos de Neuro-Psiquiatria	1686
1519- 6984	Brazilian Journal of Biology	497
1413- 8670	Brazilian Journal of Infectious Diseases	335
0100- 879X	Brazilian Journal of Medical and Biological Research	2084
1806- 8324	Brazilian Oral Research	167
0102- 311X	Cadernos de Saúde Pública	1809
1807- 5932	Clinics	133
1677- 5538	International braz j urol	289
1806- 3713	Jornal Brasileiro de Pneumologia	262
0021- 7557	Jornal de Pediatria	493
1678- 7757	Journal of Applied Oral Science	276
0074- 0276	Memórias do Instituto Oswaldo Cruz	1770
0102- 7638	Revista Brasileira de Cirurgia Cardiovascular	323

---



0034-7299	Revista Brasileira de Otorrinolaringologia	642
1516-4446	Revista Brasileira de Psiquiatria	454
0104-4230	Revista da Associação Médica Brasileira	508
0034-8910	Revista de Saúde Pública	1186
0036-4665	Revista do Instituto de Medicina Tropical de São Paulo	728
1516-3180	Sao Paulo Medical Journal	420
Total		18319

O processo de carga da base de monografias de substâncias HSDB é realizado de maneira mais simples: o Portal da NLM disponibiliza uma versão dos materiais na forma XML, e que pode ser convertida por meio de softwares de manipulação de XML para um banco de dados relacional.

A partir das duas principais bases, as mesmas foram submetidas à processos de indexação automática, com a finalidade de mapear o conceitos encontrados nos vocabulários controlados que serviram de base para a ontologia de domínio do sistema. Dessa maneira foi possível automatizar o processo de mapeamento dos conceitos do repositório de conteúdos (artigos do Scielo e Base HSDB)

### **Processo de Indexação pelo sistema do MTI**

O sistema Medical Text Indexer – MTI foi desenvolvido pela National Library of Medicine – NLM, com “o objetivo da criação do projeto Iniciativa de Indexação é



investigar métodos de indexação automática que substituem parcial ou totalmente as práticas de indexação atuais”.

Em contraponto, a iniciativa veio agregar valor e auxiliar o trabalho do bibliotecário indexador na análise documentária e na identificação dos descritores pertinentes, mas não excluir o fator humano que se sensibiliza em diagnosticar mudanças a serem tomadas quando forem percebidas.

Para que o projeto tenha validade, conforme National (2012) “os métodos criados terão sucesso a partir do momento que sejam implementados para que resultem no desempenho da recuperação da informação”.

Os materiais enviados para o servidor da NLM são compostos pelos resumos e títulos de cada artigo, num arquivo de dados similar ao fragmento apresentado a seguir:

Figura 8 - Fragmento de arquivo com texto a ser processado

	0	10	20	30	40	50	60	70	80
1	S0100-879X2002000900005	Angiotensin and baroreflex control of the circulation#The							
2	S0100-879X2002000900009	Angiotensin-(1-7) improves the post-ischemic function in							
3	S0100-879X2002000900010	Cardiovascular control in experimental diabetes#Several s							
4	S0100-879X2002000900003	Chronic converting enzyme inhibition normalizes QT interv							
5	S0100-879X2002000900007	Dissociation between the circulating renin-angiotensin sy							
6	S0100-879X2002000900006	Gender differences in vascular expression of endothelin a							
7	S0100-879X2002000900002	Interaction between paraventricular nucleus and septal ar							
8	S0100-879X2002000900008	Myocardial antioxidant and oxidative stress changes due t							
9	S0100-879X2002000900011	Nitroergic modulation of vasopressin, oxytocin and atrial							
10	S0100-879X2002000900001	Recent advances in angiotensin II signaling#Angiotensin I							
11	S0100-879X2002000900004	Role of sensory nervous system vasoactive peptides in hyp							
12	S0100-879X2002000800001	A tobacco cDNA reveals two different transcription patter							
13	S0100-879X2002000800012	Adrenoceptors of the medial septal area modulate water in							

Cada linha corresponde a um artigo: o identificador (código DOI), o título e o resumo. O resultado do processamento nos servidores da NLM é retornado em outro arquivo de dados, como ilustrado a seguir:



Figura 9 - Arquivo de dados com a resposta do processamento de indexação

	0	10	20	30	40	50	60	70
1	S0100-879X2003000900012 1:0	Inducible·Nitric·Oxide·Synthase C1533698 97970 E						
2	S0100-879X2003000900012 2:1	Nitric·Oxide C0028128 37784 MH RtM·via:·Nitric·O						
3	S0100-879X2003000900012 3:2	Oxygen C0030054 35855 MH RtM·via:·Oxygen;Forced·						
4	S0100-879X2003000900012 4:3	NADPH·Oxidase C0068355 30482 MH RtM·via:·NADPH·O						
5	S0100-879X2003000900012 5:4	Nitrogen C0028158 29375 MH RtM·via:·Nitrogen;For						
6	S0100-879X2003000900012 6:5	Oxidative·Burst C0085416 27136 ET Entry·Term·Rep						
7	S0100-879X2003000900012 7:6	Hydrogen·Peroxide C0020281 23035 MH RtM·via:·Hyd						
8	S0100-879X2003000900012 8:7	Bacillus C0004587 13237 MH RtM·via:·Bacillus·<ba						
9	S0100-879X2003000900012 9:9	Mice,·C3H C0025920 10192 ET Entry·Term·Replaceme						
10	S0100-879X2003000900012 10:8	Macrophages C0024432 11589 MH RtM·via:·macropha						
11	S0100-879X2003000900012 11:10	NOS2·protein,·human C0669372 10111 NM TI MM						
12	S0100-879X2003000900012 12:11	Oxidation-Reduction C0030012 5601 MH RtM·via:·						
13	S0100-879X2003000900012 13:12	Cell·Culture C0007585 5127 ET Entry·Term·Repla						

Observe que cada artigo (nota-se o código DOI) recebe várias entradas de termos, com todos os dados necessário para o mapeamento dos elementos do DeCS na ontologia de domínio.

O funcionamento do MTI consiste em aplicar o texto enviado ao mecanismo indexador, quer irá operar realizando as seguintes etapas:

1. Identificar expressões (sequencias de texto) para processamento
2. Buscar no MeSH os termos, com uma taxa de coincidência para descritores em 100%, e termos relacionados para 75%, usando o algoritmo de trigramas
3. Limitar a exaustividade para 15 descritores (a exaustividade foi aplicada em etapa posterior)
4. Escolher a opção de saída de arquivo “Full Listing with Detailed”
5. Escolher a opção de formato de entrada “Single Line Delimited Input w/ ID”
6. O formato de cada linha deve ser: pid|texto a ser indexado. Exemplo:



## **Mecanismo de Busca Semântica**

O mecanismo de busca semântica, que é um artefato computacional capaz de localizar conceitos da ontologia ou realizar algum processo de inferência para identificar uma determinada classe de conceito.

Foi desenvolvido um protótipo de software que implementa essa busca, e uma vez encontrado um conceito ele recupera os documentos que possuem alguma ligação estabelecida pelos processos de mapeamento descritos anteriormente.

## **Interface do Portal**

A interface de busca semântica, que consiste no meio pelo qual ele irá realizar as pesquisas. Esta interface está em desenvolvimento e implementa apenas uma parte do mecanismo de busca. O primeiro protótipo pode ser acessado em [www.labtecgc.udesc.br/bds](http://www.labtecgc.udesc.br/bds).

O processo de Recuperação da Informação nesse portal é realizado com apoio de um mecanismo de busca capaz de localizar conceitos previamente organizados numa ontologia; o conhecimento representado nesta última poderá então ser disponibilizado na interface do usuário, através da qual este obtém documentos e informações.

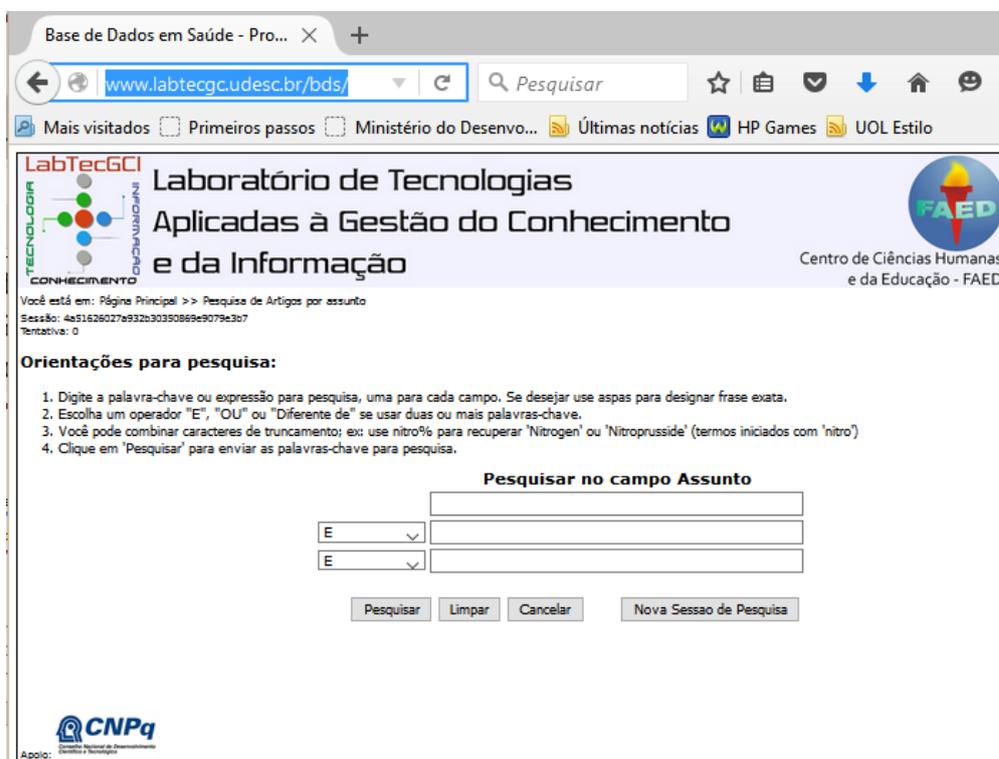


Figura 10 - Interface do Portal - versão Beta

## CONSIDERAÇÕES FINAIS

O portal terá seu software reformulado para receber melhorias e adequações aos padrões da W3C, como a disponibilização dos dados em RDF (Resource Description Framework) na forma de dados ligados.

Foram realizadas outras pesquisas a partir desse Portal, no sentido de compreender os benefícios e diferenças em relação aos sistemas de recuperação tradicionais. No entanto tais estudos ainda requerem maior desenvolvimento a fim de produzir resultados consistentes.

Ainda é necessário o desenvolvimento de produtos que contenham essas tecnologias semânticas. É bem comum encontrar iniciativas de pesquisa e trabalhos em uso em nível de laboratório; no entanto ainda é expressiva a quantidade de produtos de



bases de dados que utilizam tecnologias convencionais (ex: buscadores booleanos) em seu desenvolvimento.

## REFERENCIAS

BIREME. *DeCS/VMX: Health Sciences Descriptors*. Disponível em: <<http://decs.bvs.br/vmx.htm>>. Acesso em: 15 jul 2013.

BORST, WILLEM NICO. *Construction of Engineering Ontologies for Knowledge Sharing and Reuse*. (1997). 227 f. Tese (Doctoral thesis in Information Systems) - Dutch Graduate School for Information and Knowledge Systems, Universiteit Twente, 1997. Disponível em: <<http://doc.utwente.nl/17864/1/t0000004.pdf>>. Acesso: 21 jan 2011.

GÓMEZ-PÉREZ, ASUNCIÓN *et al.* *Ontological Engineering: with examples from the areas of knowledge management, e-commerce and the semantic web*. Londres: Springer-Verlag, 2004. 403 p.

GRUBER, THOMAS R. A translation approach to portable ontology specifications. *Knowledge Acquisition*, v. 5, n. 2, p.199-220, jun 1993. Disponível em: <<http://www.sciencedirect.com/science/article/B6WMS-45P67XR-D/2/b9a5cb273fc6497ae1b4fefad29b30b5>>. Acesso em: 01 nov 2010.

JORGE, MARCELO ALVIM. *Ontologias no suporte a portais semânticos*. (2005). 119 f. Dissertação (Mestrado em Ciência da Informação) - Escola de Ciência da Informação, UFMG, Belo Horizonte, 2005. Disponível em: <[http://dspace.lcc.ufmg.br/dspace/bitstream/1843/VALA-6KFNS2/1/mestrado\\_\\_\\_marcelo\\_alvim\\_jorge.pdf](http://dspace.lcc.ufmg.br/dspace/bitstream/1843/VALA-6KFNS2/1/mestrado___marcelo_alvim_jorge.pdf)>. Acesso: 01 mar 2011.

JULIANI, J. P. ; Adriano Pessini ; MARTINS, Alejandro ; Douglas Paulesky Juliani . Um modelo para a construção de base de conhecimentos sobre projetos nas organizações, suportado por ferramentas colaborativas. In: 39 Jaiio - Jornadas Argentinas de Informática, 2010, Buenos Aires. Anales del SSI 2010 Simposio sobre la Sociedad de la Información, 2010.

NLM, U.S. NATIONAL LIBRARY OF MEDICINE. *Hazardous Substances Data Bank (HSDB®)*. Disponível em: <<http://www.nlm.nih.gov/pubs/factsheets/hsdbfs.html>>. Acesso em: 13 jul 2011.

OMS, (ORGANIZAÇÃO MUNDIAL DA SAÚDE). *Promoting rational use of medicines: core components*. Genebra, set 2002. Disponível



em:<<http://apps.who.int/medicinedocs/collect/medicinedocs/pdf/h3011e/h3011e.pdf>>.  
Acesso em: 16 mai 2008.

OMS, (ORGANIZAÇÃO MUNDIAL DA SAÚDE). The Rational use of drugs. In: Report of the Conference of Experts, 25-29 nov, Nairobi. 1985, Conference Proceedings. Place. Published: WHO Publications Center USA 1985.

RIBEIRO JUNIOR, D. I. ; Rigo, Wanderson ; Fileto, Renato ; Oliveira, Vinícius de Araújo ; Pereira Jr. Vilmar César ; Scotti, Haline . Interfaces Web baseadas em Conhecimento para Anotação de Recursos de Informação e Gerenciamento de Repositórios. In: Simpósio Brasileiro de Informática na Educação, 2010, João Pessoa. Brasileiro de Informática na Educação (SBIE). João Pessoa : UFPB, 2010.

RIBEIRO JUNIOR, D. I. ; Cabral, Rodrigo Bittencourt ; Wangenheim, Aldo ; TOURINHO, F. S. V. ; SAVARIS, A. . Modeling and Creation of an Ontology to Knowledge related to Toxicology. In: 8th International Information and Telecommunication Technologies Symposium, 2009, Florianópolis. Proceedings of the 8th International Information and Telecommunication Technologies Symposium. Florianópolis : Fundação Barddal de Educação e Cultura, 2009. v. 8. p. 175-178.

SOUZA, RENATO ROCHA; ALVARENGA, LÍDIA. A Web Semântica e suas contribuições para a ciência da informação. *Ciência da Informação*, Brasília, v. 33, n. 1, p.132-141, jan/abr 2004. Disponível em:<<http://revista.ibict.br/index.php/ciinf/article/viewFile/50/50>>. Acesso em: 15 jun 2011.

STUDER, RUDI *et al.* Knowledge engineering: Principles and methods. *Data & Knowledge Engineering*, v. 25, n. 1-2, p.161-197, mar 1998. Disponível em:<<http://www.sciencedirect.com/science/article/B6TYX-3SYXJ6S-G/2/67ea511f5600d90a74999a9fef47ac98>>. Acesso em: 22 jan 2011.

TOURINHO, FRANCIS SOLANGE VIEIRA *et al.* Automedicação em crianças e adolescentes. *Jornal de Pediatria*, Porto Alegre, v. 83, n. 5, p.453-458, 2007. Disponível em:<<http://www.scielo.br/pdf/jped/v83n5/v83n5a10.pdf>>. Acesso em: 02 mai 2011.

---

## Informações dos autores

### Divino Ignacio Ribeiro Jr.

Universidade do Estado de Santa Catarina

Email: [divinoirj@gmail.com](mailto:divinoirj@gmail.com)



**Alessa Fabiola Santo**

UDESC

Email: [alessafds@gmail.com](mailto:alessafds@gmail.com)

