

Recuperação de Informação em Infraestruturas de Dados Espaciais usando Mecanismos de Representação do Conhecimento

Odilon Corrêa da Silva, Vitor Cunha Fontes, Thiago Silva Miranda,
Jugurta Lisboa Filho, Alcione de Paiva Oliveira

¹Departamento de Informática - Universidade Federal de Viçosa (UFV)
CEP 36570.000 – Viçosa – MG – Brasil

odilon.correa@gmail.com, vcf_2004@hotmail.com,
thiagosilvamiranda@gmail.com, jugurta@ufv.br, alcione@dpi.ufv.br

Abstract. *Geographic information is a fundamental resource to public institutions to propose improvements to factors such as environmental protection, security and health. Spatial Data Infrastructure enables cooperation and sharing of geospatial data, however, the open and distributed nature of these infrastructures is a factor that increases the difficulty of retrieval and sharing of such data. This paper presents a proposal supported by artifacts of knowledge representation, which seeks to minimize problems related to the semantic information retrieval in metadata catalogs in a Spatial Data Infrastructure.*

Resumo. *A informação geográfica é um recurso fundamental para que órgãos públicos possam atuar com eficiência em áreas como proteção do meio ambiente, educação, segurança e saúde. Infraestruturas de Dados Espaciais possibilitam a cooperação e o compartilhamento de dados geoespaciais, mas o caráter aberto e distribuído dessas infraestruturas é um fator que aumenta a dificuldade de recuperação e compartilhamento desses dados. Este artigo apresenta uma proposta apoiada por artefatos de representação do conhecimento, que busca minimizar alguns problemas relacionados à semântica na recuperação de informação em catálogos de metadados em uma Infraestrutura de Dados Espaciais.*

1. Introdução

A preocupação com fatores relacionados a proteção do meio ambiente, questões de segurança e saúde são alguns desafios dos órgãos públicos. Para que estes órgãos possam propor melhorias e desenvolver tais fatores é interessante poder identificar as áreas mais necessitadas e formas de intervenção, além de avaliar os impactos e monitorar os resultados.

Para todas estas tarefas, a informação geográfica é fundamental. O planejamento sustentável de qualquer atividade requer, primeiramente, o conhecimento do meio ambiente em que esta atividade está inserida. Para a caracterização ambiental de uma área, por exemplo, é necessário o conhecimento dos seus recursos naturais e o entendimento da interação e correlação entre eles. Os Sistemas de Informação Geográfica (SIG), juntamente com técnicas e produtos de sensoriamento remoto, podem

auxiliar na busca desse conhecimento, bem como no gerenciamento e na atualização constante dos dados espaciais coletados [Burrough e McDonnell 1998].

Neste contexto, a informação geográfica não deve somente existir, mas estar disponível e complementada com informações sobre os formatos de codificação, armazenamento, parâmetros de qualidade, limitações de conteúdo, parâmetros de projeção cartográfica e, até mesmo, suas estruturas de dados. Para atender a estas necessidades foi instituído no âmbito do poder executivo federal, a Infraestrutura Nacional de Dados Espaciais (INDE).

Como descrito no decreto nº 6.666/08, a INDE tem como objetivo “promover de forma ordenada a geração, armazenamento, compartilhamento e disseminação dos dados geoespaciais pelos órgãos públicos em todas as suas esferas”. Outra preocupação é evitar a duplicidade de ações e o desperdício de recursos na obtenção de dados geoespaciais, por meio da divulgação de seus metadados. Para isso padrões e normas estão sendo homologadas pela Comissão Nacional de Cartografia (CONCAR)¹.

Em geral, a pesquisa/busca em uma Infraestrutura de Dados Espaciais baseia-se principalmente em consultas que usam palavras-chave, coordenadas espaciais, classificação temática ou temporal. Com o resultado da busca, o usuário pode acessar a fonte dos dados para visualizá-los ou obtê-los, analisar seus metadados ou iniciar uma nova busca. Essa abordagem apresenta algumas dificuldades, especialmente para os usuários inexperientes, que podem não saber quais palavras-chave usar, como preencher adequadamente o formulário de consulta ou definir quantos critérios serão utilizados [Hochmair 2005].

Este trabalho apresenta uma proposta para um *software* de busca que tem como foco principal a semântica, ou seja, o entendimento do contexto é priorizado. O *software* faz parte de um sistema de recuperação de informação em catálogos de metadados de Infraestruturas de Dados Espaciais, apoiado em artefatos de representação do conhecimento como tesouros, *topic maps* e técnicas visuais. Para isso foi realizado um estudo sobre os aspectos teóricos e técnicos necessários ao desenvolvimento do sistema. Como resultado foi desenvolvido um protótipo funcional do sistema.

O restante deste artigo está organizado como segue. A Seção 2 apresenta uma visão geral sobre Infraestruturas de Dados Espaciais e, em particular, sobre os aspectos da utilização de metadados geográficos. A Seção 3 descreve os conceitos relacionados à representação e recuperação da informação. A Seção 4 apresenta a especificação e aspectos de implementação do *software* proposto. A Seção 5 descreve a avaliação do *software* por meio um estudo de caso e os resultados alcançados. A conclusão e as considerações finais são vistas na Seção 6.

2. Infraestrutura de Dados Espaciais

A disseminação e o compartilhamento de conjuntos de dados geográficos de origens diversas têm uma enorme projeção, se apoiado por um ambiente computacional em que os dados são livremente compartilhados de maneira integrada. A partir de algumas

¹ Trecho extraído do decreto Nº 6.666 publicado em http://www.planalto.gov.br/ccivil_03/_Ato2007-2010/2008/Decreto/D6666.htm, acessado em março de 2010

definições de padrões de intercâmbio, muitas agências nacionais de mapeamento criaram *Clearinghouses* de dados espaciais. Trata-se de um aparato computacional baseado na Internet que facilita o acesso a dados espaciais, pela criação de um local centralizado onde podem ser encontrados dados de diversas fontes e onde estão disponíveis serviços complementares, como busca, visualização, transferência e manipulação de dados espaciais [Crompvoets et al. 2004].

Com o foco no compartilhamento de dados para usuários e aplicações, muitas *Clearinghouses* evoluíram para o que Masser (1999) chama de “primeira geração de Infraestrutura de Dados Espaciais (IDE)”. Segundo Nebert et al. (2004), o termo IDE é usado para designar um conjunto de políticas, tecnologias, padrões e recursos humanos necessários (principalmente as organizações produtoras de cartografia) para que se promovam a disponibilidade e o acesso a dados espaciais. Os autores apresentam os aspectos necessários para se iniciar a construção de uma IDE, os quais incluem dados espaciais, metadados, catálogos, visualização de dados online, acesso aos dados, geoserviços, capacitação e políticas públicas.

Em ambientes onde há grande heterogeneidade nos dados tratados, o uso de metadados permite relacionar dados por meio de uma classificação que seja comum entre os diversos tipos de dados disponíveis, a fim de melhorar a integração e o compartilhamento destes dados. De acordo com Casanova et al. (2005), os metadados tratam a interoperabilidade em nível de gerenciamento da informação, facilitando a recuperação e o uso de um dado ou informação contida em um ambiente computacional.

Atualmente, existem diversos padrões de metadados geográficos que são utilizados para documentar coleções de dados geoespaciais. Dentre estes, as experiências do governo americano com o *Content Standards for Digital Geospatial Metadata* (CGSDM) do *Federal Geographic Data Committee* - FGDC [FGDC 1998] e a norma ISO/TC 211:19115 [ISO 19115 2003] são amplamente utilizados.

A norma ISO/TC 211:19115 foi desenvolvida pela *International Standards Organization* (ISO), órgão responsável pela definição de um conjunto estruturado de padrões de informações associadas direta ou indiretamente à localização espacial. Esta norma serviu de base para a especificação do Perfil de Metadados Geoespaciais do Brasil [Perfil MGB 2009], em uma iniciativa para ordenar a geração, o armazenamento, o acesso, o compartilhamento, a divulgação e o uso dos dados geoespaciais no território brasileiro.

3. Representação e Recuperação da Informação

A compreensão semântica pode ser um fator que dificulta a interação dos usuários humanos de uma IDE, já que as visões de mundo podem influenciar, por exemplo, na consulta a metadados. A compreensão da semântica deve ser um requisito para uma IDE que pretende minimizar os problemas de Recuperação de Informação (RI) [Craglia e Goodchild 2008], [Davis Jr. e Fonseca 2009]. De acordo com Davenport e Prusak (1998), para disponibilizar o conhecimento em um formato que seja acessível ao usuário é necessário definir uma codificação. Um aspecto fundamental para atingir eficientemente esses resultados é a forma como esses conhecimentos são representados.

O tesauro foi adotado na área de documentação, associado à forma de organização do vocabulário de indexação/recuperação, utilizado amplamente na área da Ciência da Informação [Dodebei 2002]. De fato, o emprego de tesouros nas tarefas de indexação e RI, busca resolver o problema da alocação de documentos em classes de assuntos, não só por sua capacidade de controlar o vocabulário, mas porque é um instrumento que relaciona os descritores/termos de forma mais consistente, apresentando uma estrutura sintética simplificada e uma complexa rede de referências cruzadas. Isto permite ao especialista, por exemplo, localizar com mais facilidade a palavra-chave requerida para uma busca.

Segundo Sigel (2000), a gestão de conhecimento busca organizar e otimizar repositórios de conhecimento para que o usuário consiga recuperar informações de forma eficaz. De acordo com Rath (2003), falar sobre estrutura de conhecimento é falar sobre *topic maps*. Conforme a ISO 13250 [ISO 2002], *topic maps* é uma norma internacional unificada para descrever estruturas de conhecimento e formalizar a sua associação com recursos de informações. Assim, *topic maps* podem ser vistos como um paradigma que permite organizar e recuperar informações armazenadas em IDE.

Ahmed (2003) apresenta um padrão de projeto para *topic maps*. Este padrão pode ser usado para representar tesouros em *topic maps* seguindo alguns princípios, propostos por Garshol (2004). O padrão de projeto proposto por Ahmed (2003) e adaptado por Garshol (2004) pode ser aplicado, por exemplo, na criação de um mecanismo que permita processar e indexar os termos de um tesauro e palavras-chave de metadados geoespaciais, gerando um *topic maps*. Essa abordagem deu base para as primeiras especificações do *software* proposto nesse trabalho.

Segundo Ferneda (2003), a RI, na maior parte das bases de dados, dá-se através dos mecanismos que empregam os princípios da álgebra booleana e da teoria de conjuntos. A presença dos mecanismos booleanos nesses sistemas não tem feito com que estes fiquem necessariamente mais fáceis de serem usados pelo usuário final. Este fato é demonstrado na pesquisa de Savoy e Picard (2001), que apontam a baixa utilização dos operadores booleanos pelos usuários das máquinas de busca da Web.

O baixo uso dos recursos da álgebra booleana para a RI deve-se a vários fatores. A primeira limitação é que a recuperação booleana cria dois subconjuntos discretos, os que satisfazem a expressão de busca e aqueles que não a satisfazem. De modo semelhante, na recuperação booleana, não há mecanismos pelos quais o usuário possa atribuir a importância relativa dos diferentes termos da expressão, dado o fato de que todos os termos da busca booleana têm peso um ou zero na expressão. Outra limitação se relaciona à natureza ambígua da linguagem natural em que a necessidade de informação é expressa, implicando a possibilidade de recuperação de muitos registros irrelevantes [Allen 2001].

A busca pela informação começa com o problema do usuário. A lacuna entre o conhecimento que o usuário detém sobre o problema ou tópico e o que usuário precisa saber para resolver o problema constitui a necessidade de informação [Kuhlthau 1990]. Segundo Grand e Soto (2002), técnicas visuais de pesquisa facilitam a percepção do ambiente pelo usuário. Para os autores, o recurso de navegação é essencial, pois ajuda o usuário a construir sua própria representação do ambiente.

A árvore hiperbólica é um exemplo de tecnologia para visualização e navegação baseada na técnica foco+contexto [Lamping 1995]. Essa técnica pode ser utilizada para representação gráfica de *topic maps*, o que proporciona ao usuário maior facilidade de visualização e navegação pelo conteúdo informacional, sem que ele precise conhecer exatamente o termo desejado. Ao utilizar árvore hiperbólica para representar *topic maps*, o tópico posicionado pelo usuário no foco aparece maior. Na navegação, o aparecimento de tópicos periféricos cresce exponencialmente à medida que o *topic maps* é deslocado. Devido a estas características, a navegação hiperbólica, em conjunto com tesauros, *topic maps* e métodos de RI, apresenta uma boa alternativa para a visualização e recuperação de grandes conjuntos de dados, por exemplo, catálogos de metadados de uma IDE.

4. Especificação e Aspectos de Implementação

Como a proposta do trabalho é prover uma alternativa para as dificuldades em RI em catálogos de metadados em uma IDE, espera-se não apenas minimizar parte desses problemas, mas também associar mais semântica às operações de consulta a catálogos de metadados. A especificação da aplicação é voltada às necessidades da IDE de um projeto do contexto ambiental. Entretanto, sua especificação e desenvolvimento têm sido conduzidos de maneira que ele possa ser aplicado em outras IDEs. Diante das necessidades iniciais levantadas, um estudo de viabilidade técnica foi direcionado aos padrões, métodos, conceitos e tecnologias existentes.

Os serviços fornecidos pela aplicação de RI em catálogos de metadados de uma IDE se baseiam em gerar, consultar e avaliar uma base de dados. Para que os serviços possam ser implementados, foram especificados os componentes da arquitetura do sistema de RI em catálogo de metadados. O sistema é composto de três módulos que disponibilizam seus serviços por meio de classes de interface. Também fazem parte da arquitetura do sistema duas aplicações (uma para administração e outra de consulta), um módulo que controla a persistência e três repositórios de dados. Os componentes e seus relacionamentos estão ilustrados na Figura 1 e descritos a seguir:

- Base de recurso: repositório que armazena os recursos coletados no catálogo de metadados da IDE. O recurso é composto pela identificação do metadado e parte de seu conteúdo;
- Base semântica: repositório que armazena o conteúdo semântico de um tesauro. Sua estrutura foi desenvolvida seguindo a especificação ISO 5964 [ISO, 1985] para construção de tesauros multilíngue;
- Base indexada: repositório que armazena o resultado do processo de indexação entre as bases de recurso e semântica. Sua estrutura segue as especificações do ISO 13250 [ISO, 2002] para construção de *topic maps*;
- Módulo geração da base indexada: módulo responsável por implementar uma interface com a aplicação de administração e gerenciar o processo de geração da base indexada como um todo;
- Módulo consulta da base indexada: responsável por implementar a interface com a aplicação de consulta e gerenciar o processo de consulta à base indexada;

- Módulo avaliação da base indexada: responsável por implementar uma interface com a aplicação de administração e gerenciar o processo de avaliação da base indexada;
- Aplicação de administração: aplicação responsável por traduzir as funcionalidades dos módulos de geração e avaliação da base indexada para algo que o usuário consiga entender e assim interagir com o sistema;
- Aplicação de consulta: aplicação que apresenta a base indexada por meio de uma árvore hiperbólica, onde o usuário pode navegar, selecionar termos de seu interesse e, por fim, processar sua pesquisa.

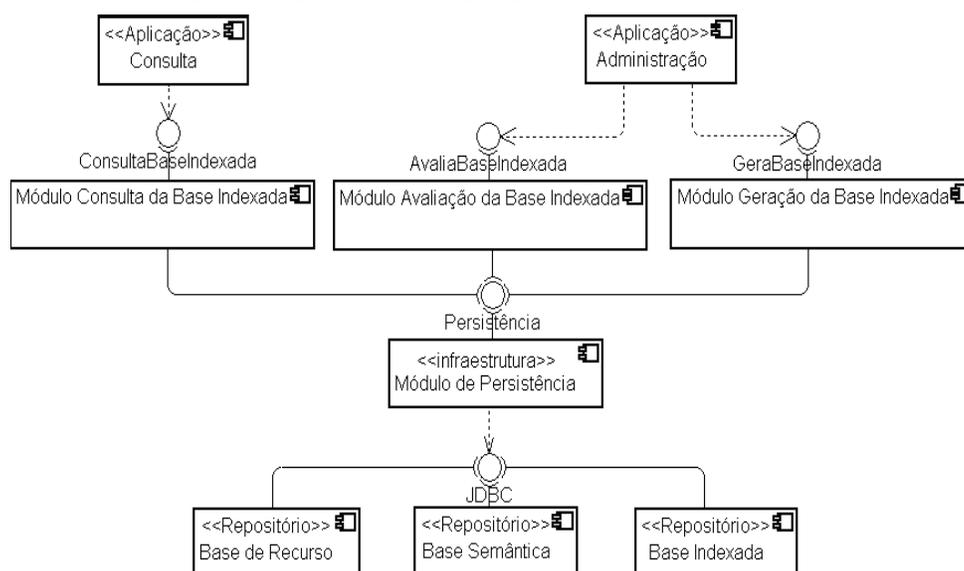


Figura 1. Diagrama de componentes do sistema

5. Estudo de Caso

A avaliação da arquitetura proposta foi realizada por meio de um estudo de caso que permitiu verificar sua viabilidade no contexto de uma IDE, que está sendo desenvolvida para atender as necessidades do projeto TERRANTAR².

O TERRANTAR é um dos sete grupos que compõem o Instituto Nacional de Ciência e Tecnologia da *Criosfera*, o qual tem como missão implementar um programa nacional de pesquisas sobre as massas de gelo e neve no planeta. O objetivo do projeto é a articulação nacional e internacional de projetos de pesquisa visando à caracterização dos ecossistemas terrestres da antártica. A gerência e o compartilhamento dos resultados dessas pesquisas são alguns dos seus objetivos específicos. Esses resultados/dados estão registrados e armazenados em diversos formatos, como dados geográficos vetoriais ou matriciais, teses, artigos, planilhas e fotos. A primeira coleção de dados que está sendo catalogada e disponibilizada é resultado de projetos sediados no Departamento de Solos da Universidade Federal de Viçosa (UFV).

² TERRANTAR – Ecossistemas terrestres da antártica - <http://www.terrantar.com.br>

5.1. Infraestrutura de Dados Espaciais do TERRANTAR

O principal objetivo da IDE TERRANTAR é apoiar o compartilhamento de dados geospaciais. Contudo, a compatibilidade tecnológica é sempre uma preocupação. Uma das iniciativas mais recentes de promover interoperabilidade foi coordenada pelo *Open Geospatial Consortium* (OGC), o qual tem proposto padrões para o uso de serviços no compartilhamento de informação geográfica.

A Figura 2 ilustra a estrutura da IDE do projeto TERRANTAR enriquecida com um Sistema de Recuperação de Informação (SRI), desenvolvido com a arquitetura proposta na seção 4. A figura também apresenta as ferramentas computacionais adotadas na implantação da IDE e seu relacionamento com elementos dessa estrutura. Por fim, pode-se observar que os dados espaciais são armazenados em diversos formatos.

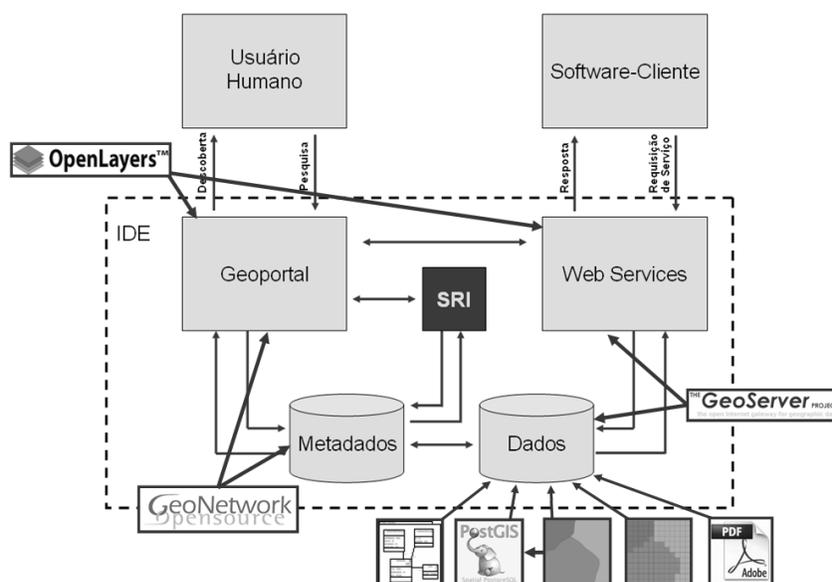


Figura 2. IDE do projeto TERRANTAR

Diante das necessidades iniciais levantadas no projeto, um estudo de viabilidade técnica foi direcionado aos padrões, métodos, conceitos e às tecnologias existentes que pudessem embasar o seu desenvolvimento. O enfoque do estudo foi a busca de ferramentas computacionais para otimizar e facilitar o desenvolvimento da IDE. Algumas ferramentas foram selecionadas para implementação da IDE, as principais são: OpenLayer³, GeoServer⁴, PostgreSQL⁵ e Geonetwork⁶.

A catalogação e acesso aos metadados são providos pelo Geonetwork. O repositório de metadados é mantido pelo SGBDR PostgreSQL. A implementação de serviços Web é provida pelo GeoServer. Essa ferramenta disponibiliza os dados geográficos em vários formatos, proporcionando interoperabilidade aos usuários.

³ OpenLayer - <http://www.openlayers.org>

⁴ GeoServer - <http://www.geoserver.org>

⁵ PostgreSQL - <http://www.postgresql.org>

⁶ Geonetwork - <http://geonetwork-opensource.org>

Finalmente, o OpenLayer permite que os usuários visualizem e analisem os dados geográficos em mapas interativos sem a necessidade de baixá-los. Os mapas são gerados utilizando-se o serviço de WMS (*Web Map Service*) disponibilizado pelo GeoServer.

Com a preocupação de melhorar a qualidade na descrição dos termos nos metadados foi especificado um procedimento para auxiliar os usuários produtores de dados espaciais da IDE. Por definição da equipe do projeto TERRANTAR, todo cadastro de metadado deve consumir termos existentes no tesouro da IDE. Caso o termo não exista, ele deve ser registrado no tesouro para que possa ser analisado e publicado. Esse procedimento possibilitou controlar os termos utilizados nas descrições dos metadados, fator que contribuiu para a geração da base indexada e eficiência no sistema de recuperação.

O aplicativo Geonetwork fornece uma solução para o portal de comunicação com os usuários. O Portal possui uma interface ilustrada pela Figura 3, que segue uma linha tradicional na recuperação de metadados. Segundo Hochmair (2005), como foi descrito na Seção 1, essa interface pode apresentar algumas dificuldades, especialmente para os usuários inexperientes, que podem não saber quais palavras-chave usar, como preencher adequadamente o formulário de consulta ou definir quantos critérios serão utilizados.



Figura 3. Interface de pesquisa da IDE do projeto TERRANTAR

5.2. Aplicação do Protótipo

Um protótipo da aplicação proposta foi implementado na linguagem Java⁷. O Sistema de Gerenciamento de Banco de Dados Relacional (SGBDR) PostgreSQL foi utilizado para armazenar parte dos dados necessários ao funcionamento do sistema. O acesso e manipulação do conteúdo de documentos XML (*eXtensible Markup Language*) são providos pela API (*Application Programming Interface*) DOM (*Document Object Model*). Para geração da árvore hiperbólica foi utilizado o pacote Java Treebolic⁸.

⁷ Java - <http://java.sun.com/>

⁸ Treebolic - <http://treebolic.sourceforge.net>

O funcionamento do protótipo se divide em três etapas, como ilustra a Figura 4. O primeiro passo consiste em importar o catálogo de metadados e tesouro para as bases de recurso e semântica, respectivamente. Em seguida, executar a rotina de geração da base indexada. Por último, a base indexada fica disponível para consulta.

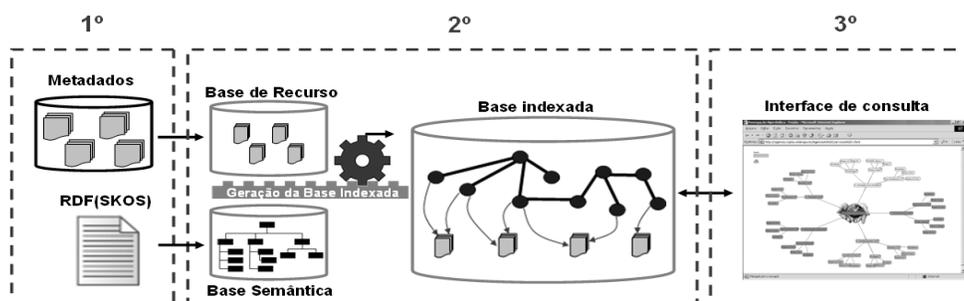


Figura 4. Etapas de funcionamento do protótipo

Os testes realizados neste estudo de caso resolvem problemas relativamente simples, porém suficientes para ilustrar as possibilidades em utilizar o sistema em cenários reais. No total, o catálogo da IDE possui 7 metadados e o tesouro utilizado no estudo de caso contém 65 termos.

A primeira etapa do teste do protótipo consistiu em gerar a base indexada. Antes de gerar a base indexada, é necessário importar os metadados e o tesouro para as bases de recurso e semântica, respectivamente. A Tabela 1 apresenta o conteúdo da Base de Recurso após a importação dos metadados, onde a coluna **Item** corresponde a uma referência ao identificador do metadado no catálogo da IDE e a coluna **Recurso** armazena suas palavras-chave.

Tabela 1. Conteúdo da Base de Recurso

Item	Recurso
1	"Perfil de Solo", "Amostra de Solo", "Península Potter"
2	"Península Potter", "Flora", "Fauna"
3	"Curva de nível", "Relevo", "Península Potter"
4	"Península Potter", "Solos Gelados"
5	"Potter Peninsula", "Permafrost"
6	"Península Potter", "Criossolos"
7	"Hidrologia", "Drenagem", "Península Potter"

A Tabela 2 ilustra como os termos e seus atributos são estruturados na base semântica. Os tipos de atributos são identificados por: **EQ** - termo equivalente; **TE** - termo específico; **TG** - termo genérico; **TR** - termo relacionado; **USE** - termo usado por.

Tabela 2. Conteúdo da Base de Semântica

Termos	Atributos
Solo	"TG: Geoambiente", "TE: Amostra de Solo", "TE: Criossolos", "TE: Perfil de Solo", "TE: Solos Gelados", "EQ: Soil"
Solos Gelados	"TR: Criossolos", "TG: Solo", "EQ: Permafrost"

A segunda etapa do teste do protótipo consiste em avaliar a Base Indexada. Como resultado dessa avaliação, pode-se observar que todos os recursos foram indexados. Isso se deve ao procedimento adotado na criação do metadado e descrito na nessa seção, em que todo metadado deve consumir termos do tesouro.

A terceira e última etapa do teste consistiu em realizar consultas com o protótipo e analisar seus resultados. Por fim, foram submetidas consultas semelhantes à interface padrão da IDE (ilustrada pela Figura 3), o que permitiu comparar os dois mecanismos. A Figura 5 ilustra a interface de consulta do protótipo.

Uma informação importante a ser exibida ao usuário é a nota de escopo do termo do tesouro. A interface exibe essa informação quando o usuário posiciona o mouse sobre o termo. Pode-se observar que os termos indexados (ex: “Solos Gelados” e “Perfil de Solo”) têm sua coloração alterada em relação a um termo não indexado (ex: “Vegetação”). Essa característica, além de evitar consultas desnecessárias, facilita sua identificação pelo usuário. A ligação entre os termos relacionados (ex: “Criossolos” e “Solos Gelados”) tem seu estilo pontilhado, diferenciando-se dos outros relacionamentos. Outra característica a ser destacada é o *multilíngue*, que foi levada em consideração no desenvolvimento da função de consulta. Essa característica não está visível na Figura 5, mas pode ser observada a relação de equivalência entres os termos “Permafrost” e “Solos Gelados”, na Tabela 2.

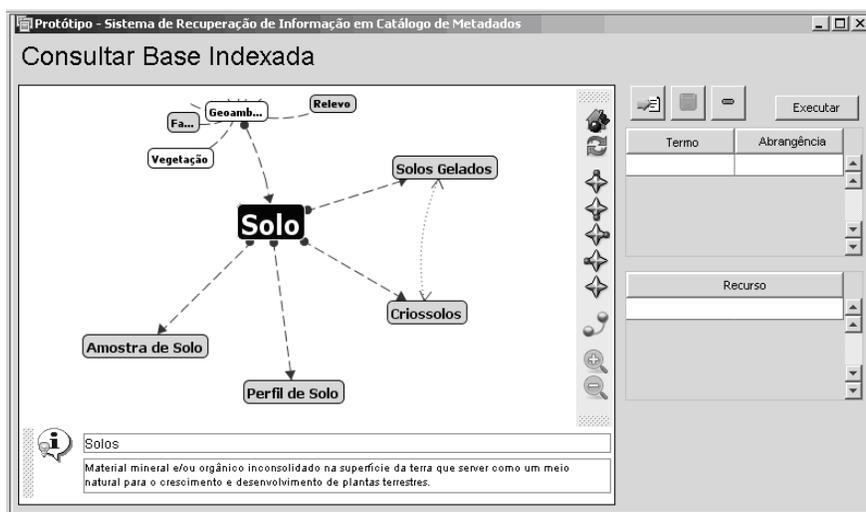


Figura 5. Tela: Consultar Base Indexada

5.3. Testes do Protótipo

O primeiro teste simula uma situação em que é informado apenas um termo e nenhum critério de abrangência. A Figura 6 ilustra os critérios dessa consulta e seu resultado. A consulta retorna dois recursos (Itens: 4 e 5 da Tabela 1). Analisando o resultado com o conteúdo da Tabela 1 e Tabela 2, é possível verificar que o fator *multilíngue* foi levado em consideração na busca. O mesmo termo passado como parâmetro na opção palavras-chave da interface da IDE (ilustrada pela Figura 3), retorna apenas um identificador (Item: 4).

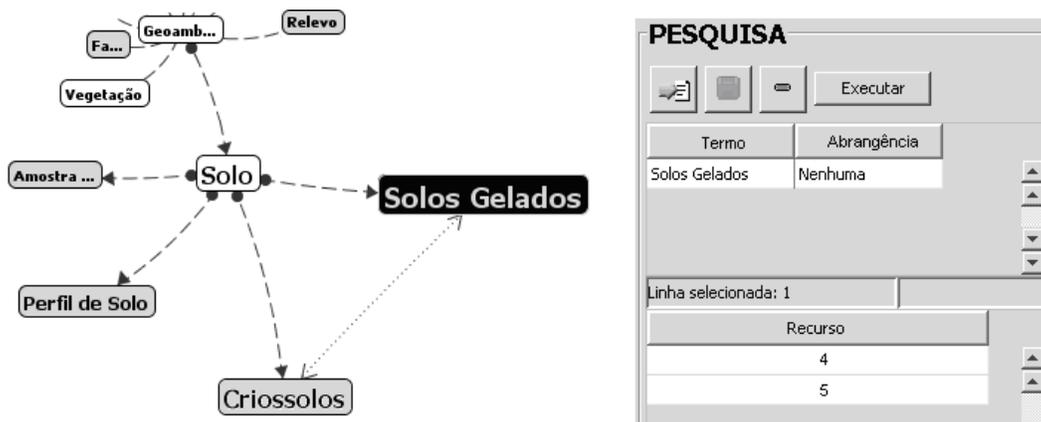


Figura 6. Teste I: consulta com termos

No teste II, utilizou-se o mesmo termo anterior e acrescentou-se como critério de abrangência a opção que faz referência a seus termos relacionados. A Figura 7 ilustra a consulta e seu resultado. A consulta retorna três recursos (Itens: 4, 5 e 6 da Tabela 1). Analisando o resultado com o conteúdo da Tabela 1 e Tabela 2, é possível verificar que o critério da abrangência é utilizado na busca. Como o termo não foi alterado, o resultado exibido na interface da IDE (ilustrada pela Figura 3) é o mesmo do Teste I.

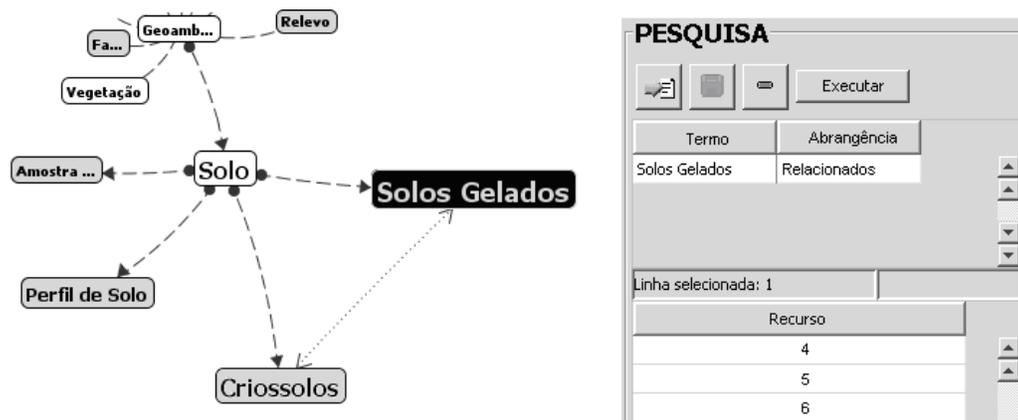


Figura 7. Teste II: consulta com termos relacionados

O terceiro teste simulou uma situação em que foi informado um termo que não foi indexado, junto com seu critério de abrangência que faz referência a seus termos especializados. A Figura 8 ilustra os critérios desta consulta e seu resultado. A consulta retorna quatro recursos (Itens: 1, 4, 5 e 6 da Tabela 1). Analisando o resultado, de acordo com o conteúdo da Tabela 1 e Tabela 2, é possível verificar que, além do critério da abrangência, o fator *multilíngue* foi utilizado na busca. Seguindo o procedimento dos testes anteriores, o termo foi informado na opção de palavras-chave da outra interface e retornou apenas um recurso (Item: 1 da Tabela 1). Não foi possível analisar a implementação do mecanismo de busca desta interface, mas por analogia com outros testes realizados, acredita-se que o mecanismo reconheceu a ocorrência do termo “Solo” nas palavras-chave “Perfil de Solo” e “Amostra de Solo”. O mesmo não aconteceu em

relação à palavra-chave “Solos Gelados”, caso essa palavra estivesse descrita no singular a mesma seria reconhecida.

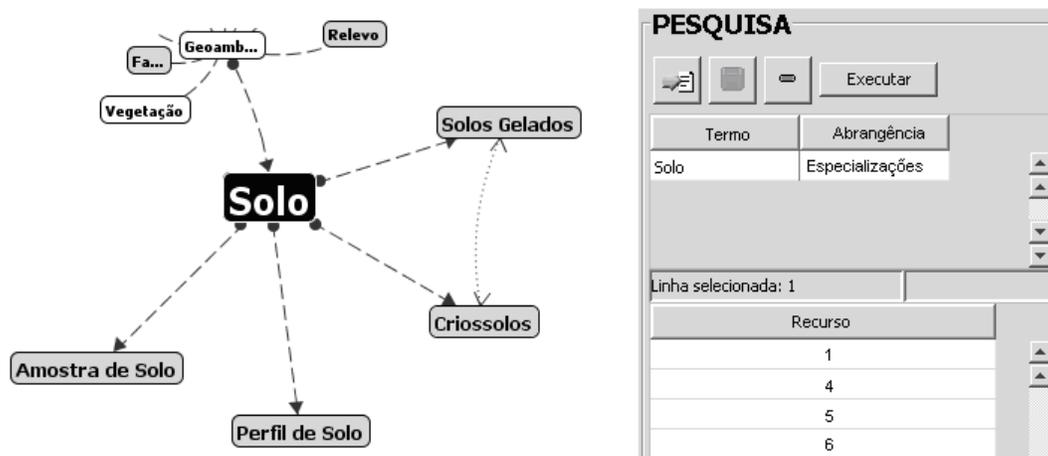


Figura 8. Teste III: consulta com termos e especializações

5.3. Comparação e avaliação dos testes do protótipo

Não é objetivo discutir os critérios relacionados à usabilidade das interfaces apresentadas, o que foge ao escopo deste trabalho e exigiria uma pesquisa das técnicas utilizadas na disciplina Interface Homem-Máquina. O que diferencia as duas interfaces e foi levado em consideração no desenvolvimento do protótipo foram os recursos que auxiliam e contribuem para o entendimento do contexto a ser consultado.

Por si só, o protótipo apresentado não garante a qualidade na recuperação da informação. Este fator é influenciado pela criação e catalogação do metadado, ou seja, seu conjunto de palavras-chave deve ser preenchido com termos do tesauro.

A avaliação do protótipo restringe-se a mostrar as possíveis soluções para algumas questões semânticas, que diferenciam essa proposta em relação a um mecanismo de busca semelhante ao adotado pela interface da Figura 3. Deve-se observar que esta mesma interface pode consumir um serviço Web que implementa a função de busca utilizada pelo protótipo e apresentar os mesmos resultados dos testes realizados.

6. Conclusões

A partir dos estudos e pesquisas realizadas, uma arquitetura para sistemas de recuperação de informação em catálogos de metadados foi especificada. A arquitetura proposta descreve um mecanismo de busca que faz uso de tesauro e *topic maps*. Esses artefatos semânticos contribuem para o entendimento do contexto a ser consultado e influenciam o mecanismo de busca. Esta abordagem permitiu minimizar alguns problemas relacionados à semântica no processo de RI.

O protótipo foi desenvolvido com recursos que auxiliam e que buscam contribuir no entendimento semântico de um contexto. Em termos práticos, buscou-se simplificar a interação do usuário com a interface de busca. Entretanto, a lógica implementada no mecanismo de busca é a chave para resolver possíveis problemas de semântica na RI.

Por fim, mas não menos importante, algumas ações foram propostas para auxiliar os produtores de dados geográficos, também usuários da IDE. O procedimento adotado na catalogação dos metadados prima pela qualidade de suas informações. Esse procedimento é composto por atividades e permite uma evolução colaborativa do tesouro utilizado no estudo de caso. A adoção desse procedimento é fundamental para se continuar a obter resultados satisfatórios na utilização do sistema.

Agradecimentos

Projeto parcialmente financiado com recursos do MCT/CNPq/CT-Info e pela Fapemig.

Referências

- Ahmed, K. *Topic map design patterns for information architecture*. XML 2003. Disponível em: <<http://www.techquila.com/tmsinia.html>>. Acesso em: março/2010.
- Allen, B. L. Boolean browsing in an information system: an experimental test. *Information technology and libraries*. p. 12-20, Março, 2001.
- Burrough, P. A.; McDonnell, R. A. *Principles of geographical information systems*. Oxford: Oxford University Press, 1998. 333 p.
- Casanova, M. et al. Integração e interoperabilidade entre fontes de dados geográficos. In: *Banco de dados geográficos*. MundoGeo, 2005. p. 305-340.
- Craglia, M.; Goodchild, M. F. "Next-Generation Digital Earth." *International Journal of Spatial Data Infrastructures Research*, v. 3, p. 146-167, 2008.
- Crompvoets, J.; Bregt, A.; Rajabifard, A.; Williamson, I. Assessing the worldwide developments of national spatial data clearinghouses. *International Journal of Geographic Information Science*, v. 18, n. 7, p. 665-689, 2004.
- Davenport, T. H.; Prusak, L. *Conhecimento empresarial: como as organizações gerenciam seu capital intelectual*. Rio de Janeiro: Campus, 1998.
- Davis JR., C. A.; Fonseca, F. Understanding global change: the role of geographic information science in the integration of people and nature. In: NYERGES, T.; COUCLELIS, H.; McMASTER, R. B. *Handbook of GIS and society SAGE*, 2009.
- Dodebei, Vera Lúcia Doyle. *Tesouro: linguagem de representação da memória documentária*. Niterói: Intertexto; Rio de Janeiro: Interciência, 2002.
- Ferneda, E. *Recuperação da informação: análise sobre a contribuição da ciência da computação para a ciência da informação*. 2003. 147 f. Tese (Doutorado em Ciências da Comunicação) – Escola de Comunicação e Arte da Universidade de São Paulo, São Paulo, 2003.
- FGDC. Federal Geographic Data Committee. FGDC-STD-001-1998. Content standard for digital geospatial metadata. *Federal Geographic Data Committee*. Washington: D.C. 1998. 78p.
- Garshol, L. M. Metadata? Thesauri? Taxonomies? Topic maps! Making sense of it all. *Journal of Information Science*, v. 30, n. 4, p. 378-391, 2004.
- Grand, B. L.; Soto, M. Visualisation of the semantic Web: topic maps visualisation. In: *International Conference on Information Visualisation (IV'02)*, 6., p. 344-349, 2002.
- Hochmair, H. Ontology matching for spatial data retrieval from Internet portals. In: GEOSPATIAL SEMANTICS – Lecture Notes in Computer Science 3799, 2005, Mexico City. *Proceedings...* Mexico City, Mexico, 2005. p. 166-182.
- ISO 19115. *Geographic information – Metadata*. Geneva: International Organization for Standardization (ISO), 2003.

- ISO 13250. *Topic maps*. 2. ed. [S.l.]: International Organization for Standardization (ISO), 2002.
- ISO 5964. *Guidelines for the establishment and development of multilingual thesauri*. [S.l.]: International Organization for Standardization, 1985.
- Kuhlthau, C. Inside the search process: information seeking from the user's perspective. *Journal of the American Society for Information Science*, v. 42, n. 5, p. 361-371, 1990.
- Lamping, J.; Rao, R.; Pirolli, P. A. Focus+context technique based on hyperbolic geometry for visualizing large hierarchies, human factors in computing systems, CHI '95 Conference. *Proceedings...* ACM Press, 1995. p. 401-408.
- Masser, I. All shapes and sizes: the first generation of national spatial data infrastructures. *International Journal of Geographic Information Science*, v.13, n. 1, p. 67-84, 1999.
- Nebert, D. D. (Ed.). *Developing spatial data infrastructures: the SDI cookbook*, Version 2.0. 2004 (GSDI-Technical Working Group). Available at: www.gsdi.org/docs2004/Cookbook/cookbookV2.0.pdf. Acesso em: março/2010.
- Perfil MGB. *Perfil de Metadados Geoespaciais do Brasil*. Disponível em: http://www.concar.ibge.gov.br/arquivo/Perfil_MGB_Final_v1_homologado.pdf.
- Rath, H. H. *The topic maps handbook*. Gütersloh, Germany: Empolis, 2003.
- Savoy, J.; Picard, J. Retrieval effectiveness on the Web. In: *Information Processing and Management*, v. 37, n. 4, p. 543-569, 1998.
- Sigel, A. *Towards knowledge organization with topic maps*. [S.l.]: XML Europe, 2000.