

Geração de dicionários explicativos para melhorar a interação entre um SIG e seus usuários

Lirisnei Gomes de Sousa, Jair C Leite

DIMAp – Universidade Federal do Rio Grande do Norte (UFRN)
Natal – RN – Brasil

`lirisnei@consiste.dimap.ufrn.br, jair@dimap.ufrn.br`

***Abstract.** This work presents an approach to improve the interaction process between one GIS (Geographic Information System) and its users. To do this, one inference layer process (1) ontologies composed by the concepts that are manipulated by the SIG and (2) ontologies that represent the knowledge of the SIG users. We have done one case study with the SWIGG, one GIS that is used to environmental studies on the petroleum and gas productive areas.*

***Resumo.** Este trabalho apresenta uma abordagem para melhorar o processo de interação entre os usuários de um SIG (Sistema de Informações Geográficas) e a interface do mesmo. Para isto, uma camada de inferência gera um dicionário processando (1) ontologias compostas pelos conceitos manipulados pelo SIG e (2) ontologias que representam o conhecimento dos usuários que utilizam o sistema. Foi feito um estudo de caso com o SWIGG, um sistema que é utilizado para a elaboração de estudos ambientais em área produtoras de petróleo e gás-natural.*

1. Introdução

Os Sistemas de Informações Geográficas (SIGs) surgiram para auxiliar diversos tipos de estudos que fazem uso de informações georreferenciadas. Os SIGs têm a capacidade de manipular dados geográficos (temáticos, cadastrais, sensoriamento remoto, modelos numéricos de terreno e redes) e dados descritivos e alfanuméricos. Sendo possível acessar informações descritivas de uma entidade geográfica a partir de sua localização geográfica e vice-versa (Lisboa 2002).

Os estudos que fazem uso dos SIGs podem envolver profissionais de diferentes áreas. Por exemplo, em um Estudo de Impacto Ambiental (EIA) os agrônomos descrevem as mudanças causadas ao solo, os sociólogos descrevem as mudanças sociais provocadas e os biólogos descrevem os impactos sofridos pelos animais. Mas um dos grandes problemas no uso de um mesmo SIG por diversos profissionais está no fato de que cada um deles trabalha com as informações segundo a sua visão e necessidade. Isto torna a modelagem de dados geográficos um processo bastante complexo, pois um único conceito pode ser definido com diferentes termos. Por exemplo, um sociólogo pode usar o termo “pedra” para fazer referência ao que para um geólogo é uma “rocha”. Os conceitos também podem ser mais detalhados ou não, por exemplo, um geólogo pode diferenciar “rocha ígnea” de “rocha metamórfica” enquanto que para um engenheiro tudo pode ser simplesmente “rocha”. Estas diferentes formas de definir conceitos originam as barreiras semânticas (Davis 2004).

A solução adotada em muitos sistemas é definir uma representação conceitual que todos os sistemas possam entender. Uma representação conceitual tem o papel de funcionar como representação intermediária entre dois sistemas, de forma que os seus termos possam ser relacionados (Davis 2004). Já surgiram vários padrões para melhorar a representação conceitual da informação geográfica. No entanto, o excesso de representações leva ao mesmo problema da incompatibilidade.

A melhor idéia é possibilitar que os usuários utilizem o sistema com suas diferentes maneiras (termos) de definirem os conceitos, ou seja, permitir que os usuários utilizem o sistema utilizando vocabulários que lhes sejam familiares. Seguindo esta idéia, surgiram os SIGs baseados em ontologia. Neste contexto, uma ontologia é a descrição de uma certa realidade com um vocabulário específico, usando um conjunto de premissas de acordo com o sentido intencional das palavras do vocabulário (Fonseca 2000). O conjunto das ontologias que estão em um SIG fazem parte da camada semântica, e esta camada é responsável por prover conhecimento para que o SIG possa adaptar, filtrar ou explicar a informação, armazenada em sua base de dados, para os diferentes tipos de usuários. O fato de um SIG explicar para um sociólogo o significado de alguns conceitos geológicos, por exemplo, pode ser importante para que o sociólogo entenda a razão de determinadas decisões técnicas de um geólogo.

Este trabalho apresenta uma abordagem, baseada em ontologias, para a geração automática de um dicionário que auxilia diferentes profissionais na interação com um SIG, a entenderem termos que não são de suas áreas de atuação. Foi feito um estudo de caso em um SIG que é utilizado para a elaboração de estudos ambientais em áreas produtoras de petróleo e gás-natural, o dicionário auxilia os profissionais ao interagirem com o SWIGG (Sistema Web de Informações Geográficas Geoambientais), um SIG desenvolvido no Laboratório de Geoprocessamento (GEOPRO) do Departamento de Geologia da Universidade Federal do Rio Grande do Norte (Souza 2004). Enquanto o usuário manipula as interfaces do SIG o dicionário é atualizado automaticamente, para explicar ao usuário somente os conceitos que o mesmo não entende, mas que está sendo visto por ele em cada momento. A seção 2 apresenta as ontologias desenvolvidas. A seção 3 apresenta a camada de inferência e o processo de geração do dicionário. Na seção 4 são feitas as conclusões.

2. As ontologias

Segundo uma visão filosófica, o termo ontologia é definido por Aristóteles como sendo uma concepção de tudo aquilo que pode “existir” ou “ser”. Mas na área da Informática, mais precisamente em Inteligência Artificial, uma ontologia é “a especificação de uma conceitualização” (Grubber 1992). Guarino (1998) já estende essa definição ao afirmar que uma ontologia é “uma especificação parcial e explícita que tenta, da melhor forma possível, aproximar a estrutura de mundo definida por uma conceitualização”. Assim, uma ontologia é uma descrição explícita de conceitos e relações referentes a um determinado domínio.

A abordagem proposta neste trabalho, para auxiliar a interação de diferentes profissionais com um SIG (Sistema de Informações Geográficas), consiste na geração automática de um dicionário que auxilia os profissionais a entenderem termos que não são de suas áreas de atuação.

A geração do dicionário é feita a partir do processamento de (1) ontologias que representam o conhecimento dos usuários e de (2) ontologias que representam o *conhecimento* do SIG. As ontologias que representam o *conhecimento* do SIG são formadas por conceitos manipulados pelo SIG. As ontologias que representam o conhecimento dos usuários são formadas por conceitos que podem ser apresentados aos usuários pela interface do SIG, e que os usuários tenham um entendimento claro do que os mesmos significam. A(s) palavra(s) que representam um conceito c_i forma(m) o tc_i , ou seja, o tc é uma, ou mais palavras, que remetem quem ler ou escuta ao conceito propriamente dito. Um conceito c_i pode ser relacionado com um conceito c_j através de uma propriedade p_k . Cada conceito c_i também pode ter uma descrição dc_i que consiste de um texto explicando o significado do conceito.

Um mesmo SIG pode manipular informações de diferentes áreas de conhecimento, o conhecimento de cada área deve ser representado por uma ontologia diferente. Por exemplo, se um SIG manipular informações sobre aspectos fisiográficos e sobre geologia, então são necessárias duas ontologias **Oaf** e **Og** contendo os conceitos manipulados pelos SIGs nas respectivas áreas de conhecimento, e estas ontologias formam a base de conhecimento do SIG **BSig**. Se este SIG for utilizado por um geólogo e por um sociólogo então são necessárias as ontologias **OGaf** e **OGg**, que formam a base de conhecimento **BG**, representando o conhecimento do geólogo nas duas áreas de conhecimento. Da mesma forma, para o sociólogo são necessárias as ontologias **OSaf** e **Osg**, que formam a base de conhecimento **BS**.

As ontologias são formadas por um conjunto de conceitos, e todos os conceitos que compõem OSaf e OGaf devem estar em Oaf, da mesma forma todos os conceitos que compõem Osg e OGg devem estar em Og. Os termos (tc_i) que definem um conceito podem ser redefinidos, por exemplo, se em Og um conceito c_1 está referenciado por t_1 , então em Osg este conceito pode ser referenciado por t_5 . A figura 1 representa graficamente este caso.

Para validar a camada de inferência, que será apresentada na próxima seção, foram desenvolvidas ontologias de acordo com as características do SWIGG, o sistema cuja camada de inferência foi inserida. As ontologias que representam o conhecimento do SWIGG foram desenvolvidas a partir da análise da interface do sistema. Também foram desenvolvidas ontologias que representam o conhecimento de um geólogo e de um sociólogo.

O conhecimento do SWIGG foi dividido em quatro áreas de conhecimento: aspectos fisiográficos, geologia, mineração e socio-economia; portanto foram construídas quatro ontologias para representar o conhecimento do sistema nestas áreas.

Para desenvolver as ontologias que representam o conhecimento de um geólogo foram feitas entrevistas, baseadas na pesquisa qualitativa, com dois alunos do Programa de Pós-Graduação em Geodinâmica e Geofísica da UFRN que são usuários do SWIGG. Os entrevistados foram questionados sobre conceitos nas quatro áreas de conhecimento citadas acima, e foram criadas quatro ontologias para o geólogo. As ontologias que representam o conhecimento do sociólogo, também nas quatro áreas de conhecimento, foram desenvolvidas empiricamente, a partir de um consenso entre estudantes do curso de ciências da computação e geólogos.

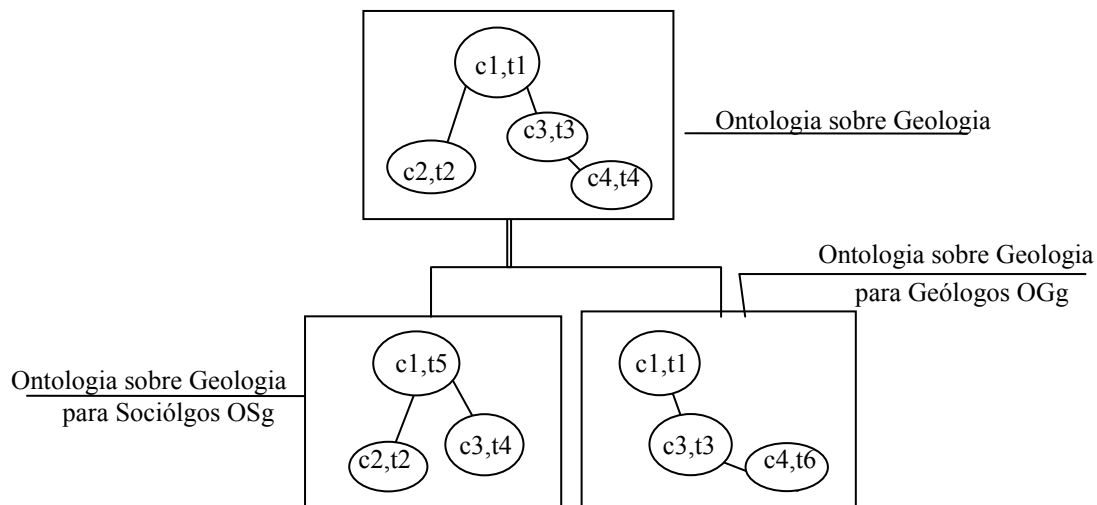


Figura 1. Relacionamento entre uma ontologia da Base de conhecimento B_{Si}g e da mesma ontologia nas bases que representam o conhecimento dos profissionais.

3. A camada de Inferência

O processo de geração do dicionário é realizado pela *camada de inferência*. Esta seção apresenta uma camada de inferência que foi desenvolvida em na linguagem de programação Java e aplicada no SWIGG. O SWIGG é um SIG desenvolvido para auxiliar à elaboração de estudos ambientais em áreas produtoras de petróleo e gás-natural na costa setentrional do estado do Rio Grande do Norte. Esta camada processa as ontologias da base de conhecimento B_{Si}g e das bases de conhecimento que representam o conhecimento dos usuários.

O SWIGG é composto de várias camadas, neste trabalho, para simplificação, estão sendo consideradas: a camada de apresentação, responsável pela geração da interface; a camada de aplicação, responsável pelo tratamento das requisições do usuário; e a camada de dados, responsável pelo armazenamento e processamento dos dados. A camada de inferência foi inserida entre a camada de aplicação e a camada de apresentação. A figura 2 mostra como a camada de inferência se relaciona com as demais.

O dicionário gerado é exibido em uma janela à parte da interface do SWIGG e contém descrições de conceitos (1) que estão na interface e (2) que o usuário não entende. A interface que o SIG gera, antes da inserção da camada de inferência, continua sendo gerada da mesma maneira, depois da inserção, ou seja, a interação entre o usuário e a interface do sistema não é modificada. A diferença é que uma tela adicional, a do dicionário, será exibida.

O processamento das ontologias ocorre da seguinte forma, quando, por exemplo, um sociólogo acessa uma interface contendo informações sobre geologia, a camada de inferência constrói um grafo $g(Og)$ representando Og e um grafo $g(OSg)$ representando OSg . Tendo $g(Og)$ e $g(OSg)$ em memória, os mesmos são analisados para saber que conceitos devem ser inseridos no dicionário. Os conceitos inseridos no dicionário são os conceitos c_i que pertencem a $g(Og)$ e não pertencem a $g(OSg)$, ou seja, o dicionário

conterá descrições de conceitos que o usuário não conhece, mas que estão na interface que está sendo visualizada pelo mesmo. Por exemplo, se o único conceito existente em $g(OSg)$ for *Rocha*, ou seja, o único conceito que um sociólogo entende sobre geologia é Rocha, então todos os conceitos c_j pertencentes a $g(Og)$ diferentes de Rocha serão colocados no dicionário. A descrição dc_i de cada conceito é inserida no dicionário, a descrição é um pequeno texto que explica qual o significado do conceito. Quando o usuário acessa o sistema, colocando nome e senha, o sistema identifica o perfil do usuário através de um arquivo que contém o cadastro do perfil de cada usuário.

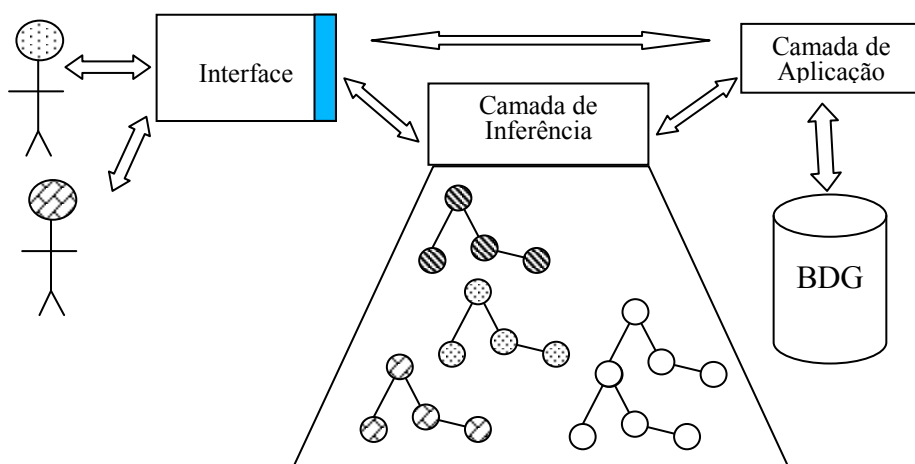


Figura 2. Arquitetura do SIG com a camada de inferência.

Também são construídas âncoras (links) para que o usuário possa navegar dentro do dicionário. Estas âncoras são construídas a partir do relacionamento entre os conceitos. Sendo c_i e c_j dois conceitos a serem inseridos no dicionário, se existe alguma propriedade p_k que relaciona c_i a c_j , por exemplo, se c_i estende c_j , então será construída uma âncora (em HTML) entre c_i e c_j no dicionário, p_k também será utilizada para incrementar o texto da descrição dc_i e dc_j .

No caso do conceito Rocha em $g(Og)$ existem três conceitos que o estendem, que são *Rocha Ígnea*, *Rocha Metamórfica* e *Rocha Sedimentar*. Neste caso, os três conceitos serão inseridos no dicionário e não serão construídas âncoras entre os mesmos e o conceito Rocha, pois este não será inserido no dicionário, mas a descrição dc_i de cada um destes será acrescentada com o texto “é um tipo de Rocha”. Como a ontologia implementada (Og) não possui descrições para estes três conceitos então somente este texto será a descrição do conceito no dicionário. Neste caso esta informação pode ser entendida pelo sociólogo, já que o mesmo entende o conceito Rocha.

O dicionário gerado é mostrado em uma janela, do navegador Web, diferente da que o usuário está interagindo com o SWIGG. A figura 3 mostra a interface do SWIGG no momento em que um usuário com o perfil de um sociólogo está interagindo com o sistema através de uma interface que possui conceitos geológicos. A figura 4 mostra o dicionário gerado, podemos ver na figura o texto “é um tipo de Rocha” na descrição dos conceitos que estendem Rocha. O conceito *Sedimento Quaternário*, não possui descrição mas, após o processamento recebe como descrição o texto “é um tipo de Sedimento”, neste caso, como Sedimento é um conceito que também está no dicionário,

então existe uma âncora (link) entre esta palavra Sedimento e a descrição do conceito Sedimento.

Como pode ser visto na Figura 4, o dicionário é composto de uma lista. Na frente do nome de cada conceito é colocada a sua descrição dc_i , incrementada ou não, como descrito acima. Abaixo de cada conceito são colocadas as instâncias do conceito com suas respectivas descrições. Por exemplo, o conceito *Rocha_Sedimentar* possui duas instâncias, que são *Beachrocks* e *Arenito*, estas instâncias, e suas respectivas descrições, foram inseridas no dicionário. Como a descrição dos conceitos, as descrições das instâncias também podem ter seus textos incrementados.

O conteúdo do dicionário é atualizado dinamicamente enquanto o usuário interage com o sistema. Se depois de interagir com a interface contendo conceitos sobre geologia, mostrada na Figura 3, o usuário (sociólogo) passar a interagir com uma interface com informações sobre mineração, então a camada de inferência gera um conteúdo, para o dicionário, contendo explicações de conceitos sobre mineração desconhecidos por um sociólogo, e atualiza automaticamente.

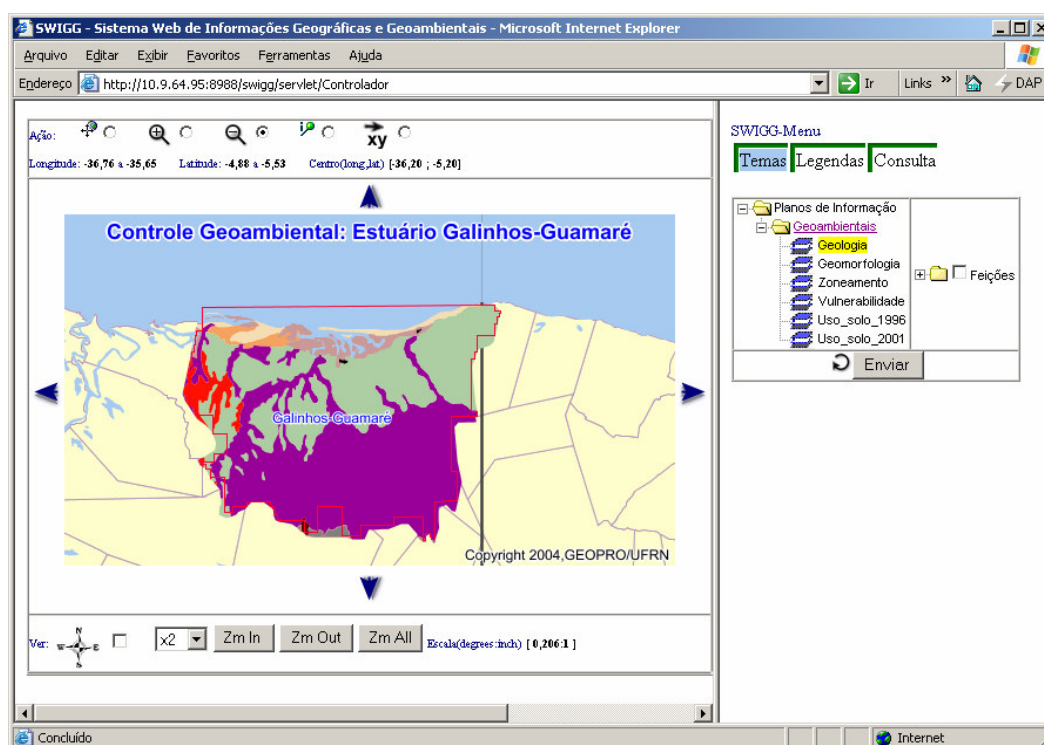


Figura 3. Interface do SWIGG.

Como pode ser visto na Figura 4, o dicionário é composto de uma lista. Na frente do nome de cada conceito é colocada a sua descrição dc_i , incrementada ou não, como descrito acima. Abaixo de cada conceito são colocadas as instâncias do conceito com suas respectivas descrições. Por exemplo, o conceito *Rocha_Sedimentar* possui duas instâncias, que são *Beachrocks* e *Arenito*, estas instâncias, e suas respectivas descrições, foram inseridas no dicionário. Como a descrição dos conceitos, as descrições das instâncias também podem ter seus textos incrementados.

O conteúdo do dicionário é atualizado dinamicamente enquanto o usuário interage com o sistema. Se depois de interagir com a interface contendo conceitos sobre geologia, mostrada na Figura 3, o usuário (sociólogo) passar a interagir com uma interface com informações sobre mineração, então a camada de inferência gera um conteúdo, para o dicionário, contendo explicações de conceitos sobre mineração desconhecidos por um sociólogo, e atualiza automaticamente.

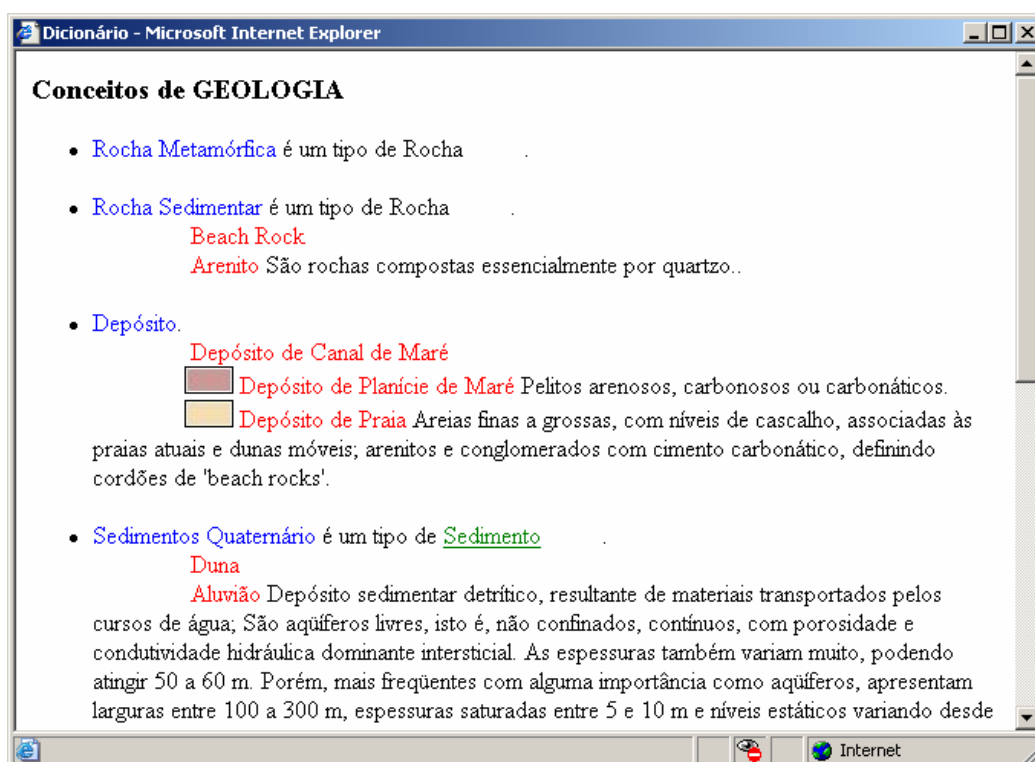


Figura 4. Interface do dicionário com conceitos sobre Geologia.

4. Conclusões

Este trabalho apresentou uma abordagem, baseada no uso de ontologias, para a geração automática de dicionário que auxilia os usuários do sistema a entenderem termos, recorrentes em estudos ambientais, que não são de suas áreas de atuação. Este trabalho introduziu uma camada de inferência, que possibilita a visualização das descrições dos conceitos visualizados nas telas da interface enquanto o usuário interage com o SWIGG. A camada de inferência gera um dicionário com a descrição dos conceitos que estão sendo visualizados pelo usuário e que não são familiares em suas áreas de atuação. A atualização do dicionário é feita para cada tela diferente que o usuário acessa no SWIGG.

A camada de inferência foi desenvolvida para funcionar o mais independente possível do sistema, isto faz com que esta camada de inferência possa ser facilmente embutida em outras aplicações. Se o SWIGG passar a manipular informações relacionadas a novas áreas de conhecimento, só será necessário desenvolver uma ontologia com os conceitos manipulados pelo sistema nesta nova área e ontologias representando o conhecimento dos usuários do sistema na mesma área de conhecimento.

A camada de inferência está pronta para suportar a inserção de novas áreas de conhecimento e de novos usuários no sistema. Atualmente estamos nos preparando para desenvolver ontologias para novos tipos de usuários.

Agradecimentos

Os autores deste trabalho agradecem a Agência Nacional do Petróleo pelo apoio, através do PRH 22.

Referências

- Davis, C., Câmara, G., Bancos de Dados Geográficos, INPE, 2004. livro disponível em <http://www.dpi.inpe.br/gilberto/livro/bdados/>
- Gruber, T. (1992) A Translation Approach to Portable Ontology Specifications Knowledge Systems Laboratory - Stanford University, Stanford, CA, Technical Report KSL 92-71.
- Guarino, N. (1998) “Formal Ontology and Information Systems” in: N. Guarino, (Ed.) Formal Ontology in Information Systems. pp. 3-15, IOS Press, Amsterdam, Netherlands.
- Fonseca, F. e Engenhofer, M. (2000) “Ontologias e interoperabilidade semântica entre SIGs”. III WORKSHOP BRASILEIRO EM GEOINFORMÁTICA, São Paulo.
- Lisboa Filho, J. (2002) “Estruturação e Modelagem de Bancos de Dados para GIS” Curso ministrado no GIS Brasil 2002 – Curitiba / PR.
- Souza, C. F., (2004) “Sistema Web De Informação Geográfica Para Gestão Ambiental: Monitoramento Das Áreas Costeiras Do Litoral Norte Do Estado Do Rio Grande Do Norte Sob Influência Da Indústria Petrolífera” Relatório de Graduação, Curso de Ciências da Computação, DIMAp - UFRN.