

13. Especificação de Mecanismo para Anotação Semântica de Dados Geoespaciais Disponíveis na Web

Autores: Carla Geovana do Nascimento Macário, Claudia Maria Bauzer Medeiros

A agricultura é uma atividade que contribui significativamente para o PIB brasileiro. Segundo [1] em 2005 o PIB atingiu R\$ 1.929 bilhões, sendo R\$ 537 bilhões provenientes de atividades agrícolas, o que corresponde a quase 30% do montante brasileiro. Desse valor, grande parte é proveniente do agronegócio. Diversos fatores contribuem para estes números. Um deles é a disponibilidade cada vez maior de sistemas que auxiliam no planejamento e no gerenciamento da produção. Em qualquer uma das fases da produção – seja ela antes, dentro e depois da porteira – a tecnologia de informação está presente como uma importante ferramenta para o aumento da produtividade e da qualidade dos produtos gerados. A rapidez no acesso às informações e, conseqüentemente, na tomada de decisões, é determinante na competitividade de uma empresa e seu sucesso a longo prazo.

Aplicações geográficas são aquelas que podem utilizar os SIGs em seu desenvolvimento para produzir algum tipo de informação geográfica, como é o caso da agricultura de precisão, que permite construir bancos de dados espaciais e temporais visando a otimização na utilização dos insumos agrícolas e o uso racional da terra, com conseqüente redução de custos e impactos ambientais.

Atualmente a web desempenha um papel importante neste contexto, funcionando, para esses sistemas, como um grande repositório de informações distribuídas pelo mundo – imagens de satélites, repositórios de dados, registros de coletas, mapas, entre outras. Tal fato tem provocado mudanças radicais nos sistemas no que diz respeito ao acesso e ao uso de informação geoespacial. Há 10 anos esses sistemas processavam dados locais e, portanto a informação era suficientemente livre de ambigüidade para aqueles que as usavam. Nos dias de hoje, os dados e métodos distribuídos podem ser recuperados e combinados em uma maneira *ad hoc*, de repositórios em qualquer lugar do mundo, extrapolando o seu contexto local. Como a semântica dos dados é geralmente descrita em uma linguagem não processável, o seu compartilhamento torna-se difícil. Além disso, os dados são ainda coletados e armazenados por organizações diferentes, usando seus próprios dispositivos, para propósitos distintos, geralmente em formato não padrão. Como aumento da demanda e do número de aplicações e usuários, não se sabe a priori quais os usos que cada dado pode ter. Normalmente, a busca por métodos e dados faz-se pelo seu conteúdo sintático, focando apenas no casamento de palavras-chave. Isso leva a vários problemas semânticos quando se trata de aplicações geográficas [2]:

- em nível de metadado, impede a descoberta de informação geográfica;
- em nível de esquema, impede a recuperação de informação geográfica;

- em nível de dados, impede a interpretação e integração de informação geográfica.

A interoperabilidade semântica é um pré-requisito para o acesso e busca efetiva de dados relevantes em diferentes contextos de aplicação. As soluções que vem sendo amostradas passam pela definição de padrões (para facilitar o intercâmbio de dados) e metadados (para permitir encontrar o que se deseja). Tem-se então um desafio: como viabilizar o uso de dados geoespaciais distribuídos e heterogêneos? Esta não é uma atividade trivial. Primeiro, pelo volume de informação gerado a cada momento. Segundo, pela dificuldade de prover um conteúdo semântico a estes dados de maneira a proporcionar seu uso efetivo pelas diversas aplicações geográficas.

Como resposta parcial a esse desafio, surgiu a idéia da Web Semântica, que busca agregar conteúdo semântico à informação. Para isso, propõe o uso de anotações sobre os recursos de informação – páginas web, imagens, mapas, etc... – para descrição de seu conteúdo, por meio de ontologias de domínio. Como resultado espera-se que as páginas de informação sejam anotadas em uma linguagem processável permitindo aos diversos agentes de busca proceder consultas mais eficazes [3].

Quando se considera componente espacial, base em aplicações geográficas, o uso de anotações se torna mais complicado. Informação geográfica associa objetos e eventos a localidades, possuindo um vocabulário rico de nome de lugares e objetos geográficos, relacionamentos espaciais, padrões, fluxos e interações. Além disso, estes nomes variam com o tempo e até com o tipo de uso. A simples adoção de ontologias no processo de anotações não é suficiente. O uso de ontologias pode ser capaz de resolver questões de ambigüidades de nomes, mas é insuficiente para resolver relacionamentos espaciais, que podem ser especificados usando termos imprecisos como perto, próximo, longe, a $X km$. Por outro lado, coordenadas geográficas podem explicitar o local (onde), facilitando muitas consultas, mas não garante a semântica associada.

Esta limitação tem motivado pesquisa em Recuperação de Informação Geográfica (RIG), uma área em expansão que consiste principalmente em desenvolver ferramentas para recuperação de recursos de informação geográfica (dados, mapas, informação em geral) na Web para responder a questões do tipo *<tema, relacionamento espacial, localização>* [4]. Algumas das questões envolvidas são interpretar adequadamente os termos da consulta, resolvendo possíveis problemas de ambigüidade, e encontrar os documentos associados a relacionamentos geográficos estabelecidos, retornando-os segundo sua relevância.

O processo para criação totalmente automática de anotações representa um problema não resolvido. Várias abordagens semi-automáticas, que exigem a intervenção do usuário em algum momento do processo, vêm sendo propostas. A anotação semântica, ainda que semi-automatizada, aparece como uma boa solução para a interoperabilidade semântica, quando tratamos de informação textual. Mas quando se pensa em aplicações geográficas devemos considerar vários tipos de recurso além de texto - imagens de satélite, mapas, gráficos, bases de dados. Mecanismos para

anotação semântica desses recursos são praticamente inexistentes e com isso, grande parte de informação útil aos sistemas geográficos deixam de ser recuperadas.

O objetivo deste projeto de pesquisa de doutorado é propor um serviço de anotações semânticas geoespaciais – anotações para diferentes dados espaciais (textos, bases de dados, mapas, gráficos, imagens, workflows) – permitindo a recuperação de informação geográfica estratégica para a agricultura. Seguindo a área de RIG, pretende-se a adoção de ontologias como base para este processo, focando também na descoberta de conhecimento. Os principais pontos de pesquisa a serem cobertos são: uso de ontologias e componentes espaciais para anotação de recursos não textuais; uso combinado de diferentes fontes de dados para geração de anotações; documentação deste processo de geração, usando por exemplo, *workflows* científicos; armazenamento das anotações e formato a ser utilizado nas anotações. Espera-se que ao final as anotações geradas por meio deste serviço permitam aos usuários obter respostas a questões do tipo *Quais as culturas são adequadas para plantio na região X?* ou *Existem áreas de desmatamento na região Y?*.

Pretende-se que o serviço a ser desenvolvido seja integrado ao WebMAPS [5] - Sistema baseado na Web Semântica para Monitoramento Agrícola e Previsão de Safras, contribuindo para um de seus objetivos específicos: facilitar o gerenciamento e previsão de safras, e a tomada de decisão com base em dados confiáveis e em ferramentas computacionais que incorporem metodologias científicas ajustadas às condições geofísicas e climatológicas brasileiras.

Referências

- [1] CEPEA. *Centro de Estudos Avançados em Economia Aplicada (CEPEA)*. Esalq/USP, 2007. <<http://www.cepea.esalq.usp.br/pib/>>. Acesso em: 19 jun. 2007.
- [2] M. Lutz and E. Klien. Ontology-based retrieval of geographic information. *International Journal of Geographical Information Science*, 20(3):233–260, 2006.
- [3] V. Uren, P. Cimiano, J. Iria, S. Handschuh, M. Vargas-Vera, E. Motta, and F. Ciravegna. Semantic annotation for knowledge management: Requirements and a survey of the state of the art. *Web Semantics: Science, Services and Agents on the World Wide Web*, 4(1):14–28, january 2006.
- [4] C.B. Jones, A.I. Abdelmoty, D. Finch, G. Fu, and S. Vaid. The spirit spatial search engine: Architecture, ontologies and spatial indexing. In *Geographic Information Science: Third International Conference, Gi Science 2004*, pages 125 – 139, Adelphi, Md, USA, October 2004.
- [5] Claudia Maria Bauzer Medeiros. *The WebMaps Project*. LIS/IC/UNICAMP, 2003. Disponível em: <<http://www.ic.unicamp.br/~cmbm/webmaps.html>>. Acesso em: 8 dez. 2006.