

Uma Representação Ontológica da Geografia Física de Portugal

Catarina Rodrigues, Marcirio Chaves, Mário J. Silva

Universidade de Lisboa, Faculdade de Ciências

PALAVRAS-CHAVE: Informação Geográfica, ontologia geográfica, geografia física

INTRODUÇÃO

A informação geográfica assume um papel cada vez mais preponderante na nossa sociedade, evidenciado pelo surgimento de novas formas e metodologias de representação e utilização dessa informação. Paralelamente a este fenómeno, ocorre a diversificação das formas de acesso à Internet. Os dispositivos portáteis, por exemplo, permitem acesso dinâmico a dados distribuídos por diversas fontes de informação. A informação de âmbito geográfico passará a ser essencial em novos serviços acedidos a partir destes dispositivos. A conjugação dos dois fenómenos acima referidos resulta numa crescente procura de serviços de localização geográfica, serviços que podem aparecer incorporados numa grande diversidade de aplicações de processamento de dados espaciais.

As ontologias constituem a forma usada pela web semântica para representação de conhecimento partilhado, pelo que surge a oportunidade de considerar a criação de ontologias geográficas, com a definição dos conceitos geográficos comuns a cada comunidade. Essas ontologias poderão então ser reutilizadas em múltiplas aplicações que necessitem de comunicar com os utilizadores exprimindo nomes, tipos de lugares ou relacionamentos geográficos.

O grupo XLDB da Universidade de Lisboa disponibiliza desde 2005, gratuitamente, uma primeira ontologia de Portugal, designada por Geo-Net-PT01 (Chaves et al., 2005). Contudo, essa ontologia contém informação geográfica restrita ao âmbito administrativo do território português.

Para complementar essa ontologia é necessário inserir informação sobre a geografia física de Portugal, o que nos motivou a estudar os modelos de representação de conhecimento existentes sobre a geografia física. Após a identificação desses modelos, duas tarefas mais complexas surgem: a integração entre si dos modelos físicos existentes e, em seguida, a integração com o modelo administrativo já representado na Geo-Net-PT01.

O projecto de extensão da ontologia geográfica de Portugal com informação do domínio da geografia física envolve as seguintes fases:

- Identificação sistemática dos conceitos, através do levantamento das entidades associadas, como: planeta, continente, lago, rio, ribeira, planície ou serra. Esta é a fase da construção da base de todo o trabalho subsequente, onde apenas serão tratados os conceitos que forem identificados nesta fase;
- Identificação das instituições com dados relativos ao domínio físico. A pesquisa será orientada de modo a conhecer quem tem autoridade para estabelecer a definição de cada elemento a incluir na ontologia. Exemplos de instituições com dados no domínio físico são o Ministério do Ambiente e o Instituto Geográfico Português;
- Estabelecimento de relações geográficas entre os conceitos. Pretende-se definir relações lógicas de inclusão e adjacência, e, posteriormente, outras mais difusas usadas em linguagem natural, como “perto de”.
- Extensão da base de conhecimento GKB (*Geographic Knowledge Base*), utilizado para integração da informação geográfica administrativa que originou a ontologia Geo-Net-PT01. Torna-se necessária a revisão do modelo conceptual utilizado para gerir a informação resultante da incorporação da informação de domínio físico;
- Carregamento dos dados do domínio físico a obter das fontes de informação seleccionadas na base de conhecimento;
- Produção de uma nova ontologia geográfica de Portugal, com os dados e o conhecimento geográfico do domínio físico integrados no GKB. Tal como na Geo-Net-PT01, será produzida no formato OWL padronizado para a web semântica.

O presente artigo está estruturado da seguinte forma: apresenta, em primeiro lugar, uma síntese dos conceitos relativos a ontologias, suas aplicações e relações com a web semântica. De seguida, apresenta um modelo para representação ontológica da geografia física do planeta, com ênfase no território de Portugal. Descreve posteriormente a metodologia seguida para identificação e integração do vocabulário no GKB e geração da ontologia. Finalmente, serão apresentadas as conclusões.

ONTOLOGIAS

O termo ontologia é frequentemente definido como uma especificação explícita de uma conceitualização de um domínio particular (Gruber 1993), especificando que entidades ou conceitos podem existir neste domínio, quais as relações entre elas e, eventualmente, quais as operações que suportam. Uma ontologia pode assim ser definida como a especificação e conceitualização de um modelo de conhecimento de uma área de conhecimento determinada.

As ontologias são frequentemente formadas por pelo menos cinco componentes que devem ser parte desta independentemente da sua finalidade:

- Entidades (ou classes, ou conceitos, ou *feature types*)
- Ocorrências (ou objectos, ou *instances*, ou *features*)
- Definições das entidades,
- Relacionamentos entre entidades e entre *features*,
- Metadados.

As ontologias são uma nova forma de representação da informação, que cada vez assume um papel de maior destaque. A importância crescente das ontologias advém do facto destas permitirem:

- A modelação formal de domínios de aplicação;
- A Compreensão do conteúdo de bases de dados;
- A partilha de informação, ou seja a interoperabilidade de dados, essencial para a comunicação entre sistemas de vários tipos, como os de informação cadastral ou os motores de busca espaciais.

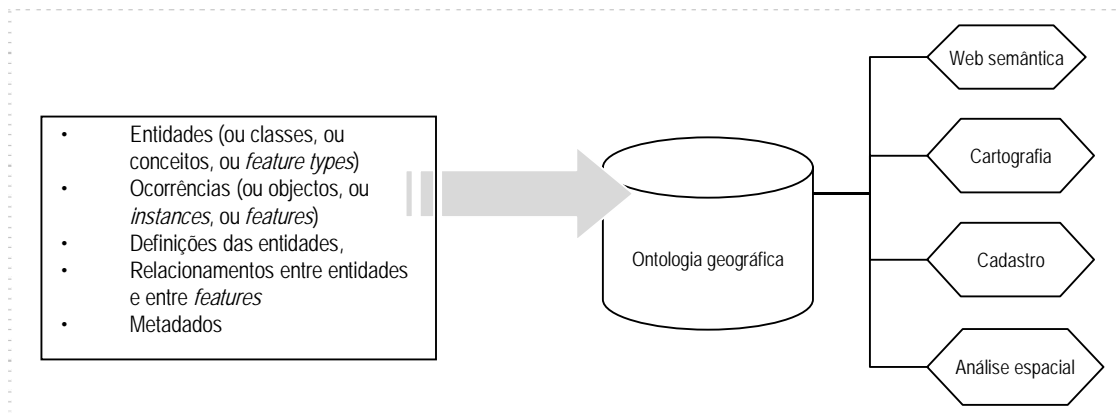


Figura 1: Aplicações das ontologias geográficas

São várias as aplicações que podem tirar partido de uma ontologia geográfica, como se pode ver na Figura 1. Estas aplicações vão desde a produção cartográfica à implementação de projectos de SIG. A representação ontológica da informação geográfica permite a partilha de recursos que podem ser reutilizados na criação de projectos de Sistemas de Informação Geográfica. A reutilização possibilita tanto a redução de tempo e custos de desenvolvimento, quanto o melhoramento da qualidade da informação e análises efectuadas

A aplicação primária da ontologia em desenvolvimento é um motor de pesquisa da web (Silva, 2006). Esta aplicação utiliza recursos da denominada web semântica, uma infra-estrutura que

tem como objectivo a estruturação dos dados publicados na web, de forma a viabilizar a automatização do seu processamento por computadores (Berners-lee, 2001).

A web semântica pretende ser uma evolução da web tradicional, a qual foi desenvolvida para ser entendida apenas pelos utilizadores. A web semântica tem como objectivo ser compreendida pelas máquinas, na forma de agentes computacionais capazes de operar eficientemente sobre as informações, podendo entender os seus significados.

Com a informação presente numa ontologia, podem-se também construir ajudas aos utilizadores, através do fornecimento de terminologia que estes podem não recordar ou desconhecer. Mais ainda, a partir de uma ontologia geográfica pode obter-se conhecimento que permita aos motores fazer inferências sobre conceitos. Por exemplo, uma consulta numa colecção de textos sobre “poluição nas zonas costeiras de Portugal” poderá ser expandida para incluir os termos que designam cada um dos concelhos do litoral português, aumentando o número de documentos relevantes obtidos nessa consulta.

Para poder efectuar estas inferências é necessário que a informação semântica seja incorporada. Isso possibilitará os utilizadores entenderem a informação, assim como também as máquinas poderão entendê-la e executar operações. A web semântica pretende fornecer estruturas e dar significado ao conteúdo das páginas web.

Segundo Berners-Lee (2001) os computadores necessitam ter acesso a colecções estruturadas de informações (dados e metadados) e de conjuntos de regras de inferência que ajudem no processo de dedução automática. Estas regras podem ser especificadas nas ontologias, permitindo explicitar a semântica dos dados. Através dessas ontologias é possível elaborar uma rede enorme de conhecimento humano, complementando o processamento da máquina e melhorando qualitativamente o nível de serviços na web.

Segundo Hendler (2001), a web semântica pode ser considerada como a composição de um grande número de pequenos componentes ontológicos que apontam entre si. É necessário ressaltar que a semântica não está somente relacionada ao conteúdo de um domínio, mas também à forma de como este se relaciona com os demais domínios na web. Portanto, é essencial que os domínios disponibilizados sejam expressivos para que as máquinas ou agentes sejam capazes de processar e entender o significado real do dado.

Este é precisamente o objectivo de investigação do projecto em curso, que tem em desenvolvimento um motor de busca inovador que tira partido do conhecimento geográfico sobre Portugal disponível sob a forma de ontologias para melhorar as pesquisas relacionadas com lugares no território português.

REPRESENTAÇÃO ONTOLÓGICA DO CONHECIMENTO GEOGRÁFICO

O modelo ontológico da geografia que desenvolvemos é representado sobre o meta-modelo do GKB. Pode-se definir um meta-modelo como a representação de um modelo (Schmidt,

2006). Um meta-modelo é, então, uma conceitualização das entidades e relações entre as mesmas presentes num modelo.

Todas as *features* e objectos que constituem o planeta e que se integram no domínio da geografia física têm relações entre si. As relações entre as *features* (objectos) e os seus tipos (classes) são também representadas neste meta-modelo. É com estas duas características das *features* que será possível às máquinas com a web semântica inferir relações e retornar valores para as perguntas tipo “concelhos que tocam Lisboa”.

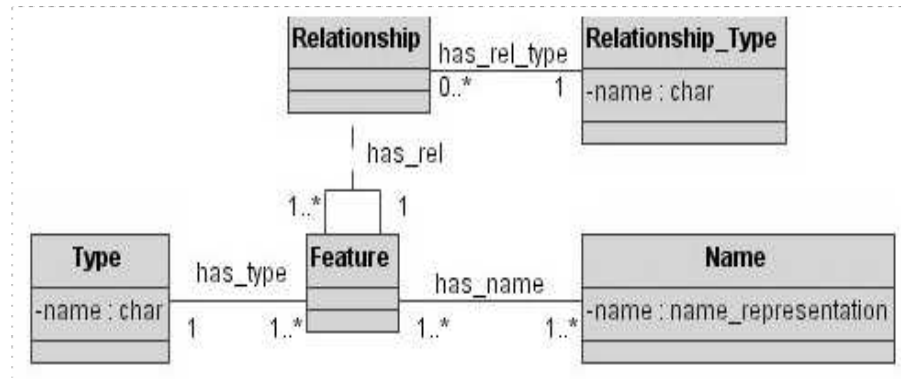


Figura 2: Meta-modelo da Geo-Net-PT01

Na Figura 2 pode ver-se uma representação do meta-modelo definido no GKB. As entidades geográficas são designadas por *features*, tendo associadas tipos (*types*) e nomes (*names*). Permite-se ainda a definição de relações (também estas de vários tipos) entre *features*. Permite-se assim a caracterização dos objectos geográficos tanto em termos dos nomes que lhes são atribuídos como das suas relações com outros objectos. Por exemplo, o Tejo é representado como uma *feature* do tipo Rio e nome Tejo. Por sua vez, o Mar da Palha é também representado como uma *feature* do tipo Rio cujo nome é “Mar da Palha”. Estas duas *features* estão associadas entre si através de um relacionamento do tipo “parte-de”.

A representação das *features* e seus relacionamentos definidas em cada domínio é efectuada através de um modelo que incorpora todas as classes e objectos específicos do domínio da geografia física de Portugal, nomeadamente relações de adjacência (“junto a”) e de pertença (“contido em” e “contido por”).

A Geo-Net-PT01 tem já uma representação do território de Portugal no domínio administrativo muito detalhada. Neste artigo centramo-nos no modelo para representar o domínio da geografia física, assim como os relacionamentos a estabelecer entre objectos destes dois domínios da geografia.

Para a definição deste modelo procurámos analisar exhaustivamente as entidades possíveis em Portugal no domínio da geografia física. Estas entidades contemplam domínios da geografia, tais como a hidrografia ou rede viária e o uso e ocupação do solo.

A nova versão da ontologia de Portugal, Geo-Net-PT02, conterá informação necessária para responder de forma satisfatória a consultas como quais os "parques empresariais na zona de Lisboa" ou as "áreas protegidas em Portugal".

Para concretizar este objectivo é necessário definir o que são as áreas industriais, assim como o que se considera zona de Lisboa. Nos casos em apreço, a nossa decisão foi considerar como zona industrial as áreas delimitadas num plano de ordenamento, e classificadas como tal, infraestruturadas ou não previamente. Como zona de Lisboa considerámos toda a área metropolitana de Lisboa.

Uma outra questão relativa à implementação do modelo ontológico da geografia física tem a ver com a definição de "ilha". Dentro do conceito "ilhas portuguesas" deverão incluir-se todas as ilhas existentes em Portugal, desde a Ilha de Tavira e da Armona, às ilhas dos Arquipélagos da Madeira e Açores. No modelo desenvolvido considerámos os dois arquipélagos portugueses como adjacentes ao Oceano e não contidos neste. Esta forma de conceptualização resulta de na definição de ilha já estar implicado que estas são uma porção de terra rodeada por mar, o que também se aplica aos continentes.

METODOLOGIA

A integração do vocabulário e geração da ontologia será realizada com recurso a uma aplicação já existente que foi criada para a Geo-Net-PT01, a GKB.

A GKB é uma base de conhecimento desenvolvida no Grupo XLB, que tem como objectivo a integração de dados geográficos provenientes de fontes de informação diversas com vista à posterior geração de ontologias com o conhecimento geográfico organizado. A base de conhecimento é alimentada por múltiplas fontes de informação heterogéneas, compreendendo desde bases de dados geo-referenciados a textos de páginas web com nomes de entidades geográficas e seus relacionamentos.

A Geo-Net-PT02 será, tal como a Geo-Net-PT01, construída com recurso a múltiplas fontes de informação, como se ilustra na Figura 3.

A geração da ontologia pela GKB implica o fornecimento da informação recolhida em múltiplas fontes, assim como o modelo ontológico e respectivo meta-modelo. Com esta informação a GKB produz a ontologia com conhecimento geográfico, no caso específico com conhecimento geográfico dos domínios administrativo e da geografia física no formato OWL padronizado para a web semântica

Como se pode ver na Figura 3, toda a informação relativa ao meta-modelo, modelo ontológico assim como relativa às *features* e relações entre estas é fornecida à GKB que transforma esta informação numa ontologia.

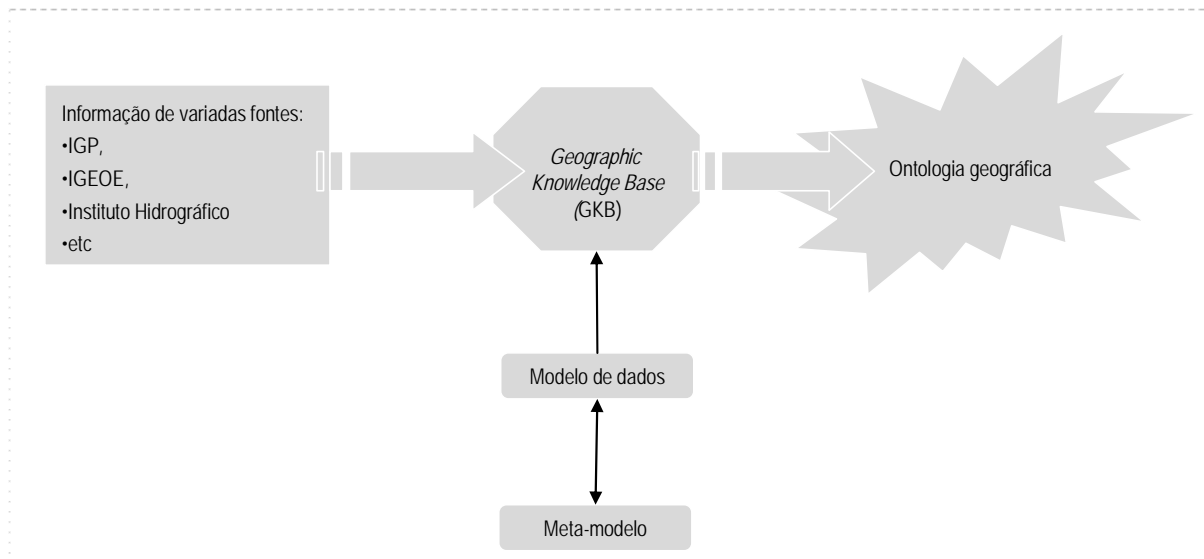


Figura 3: Fases da geração da ontologia geográfica

Pode-se comparar a função da GKB com um SIG onde são fornecidos vários *layers* para se obter novo *layer* com a informação requerida pelo utilizador, através do cruzamento dos dados dos *layers* de entrada.

A informação que constituirá a ontologia, como já foi referido anteriormente, será recolhida de diversas fontes, como tal antes da recolha da informação é necessário efectuar o mapeamento das fontes com a autoridade sobre o domínio, atingindo assim um objectivo importante para este e futuros trabalhos.

A falta deste mapeamento é um obstáculo para que a informação possa assumir um grau de interoperabilidade satisfatório, pois não há consenso nas definições. Como tal cada autoridade lhe atribui o seu significado. Neste projecto apenas criamos definições quando estas não existam. Para cada uma das *features* e objectos constituintes dessas *features* será procurada uma definição na respectiva fonte com autoridade sobre o domínio.

Assim, pretende-se que toda a informação relativa ao meio marinho seja a que o Instituto Hidrográfico utiliza, para as utilizações deste meio as entidades e definições da Secretaria de Estado das Pescas e Ministério do Ambiente.

Na Figura 4 pode ver-se o esquema do processo de construção da ontologia: mapeamento das bases de dados e fontes de informação com dados do domínio da geografia física, recolha desses dados, recolha de informação do domínio da rede, integração de toda a informação na GKB e representação ontológica de todos os dados que foram integrados na GKB na forma da ontologia Geo-Net-PT02.

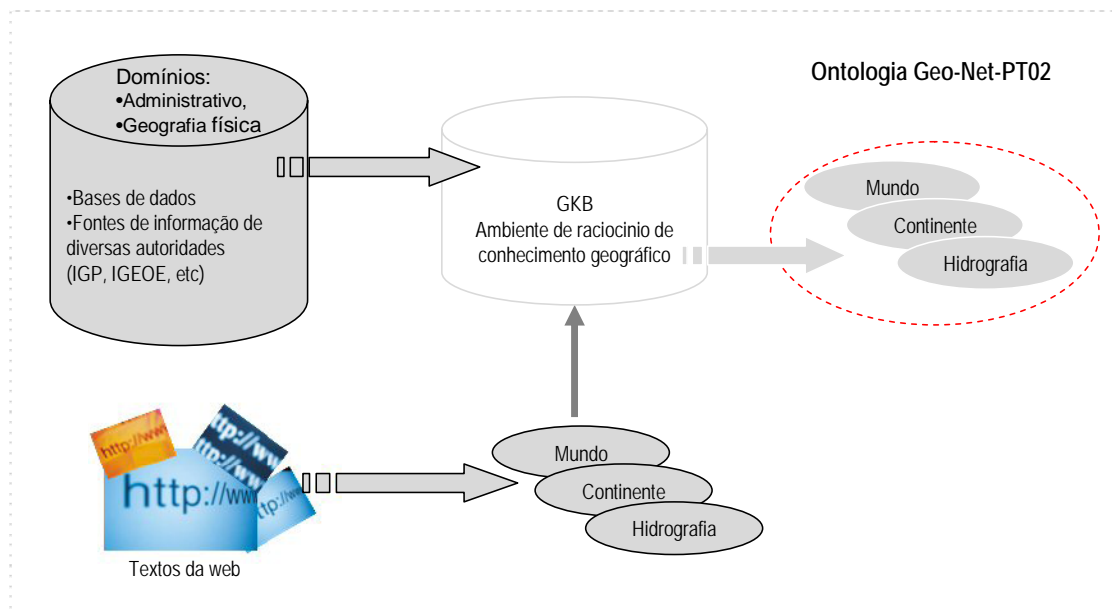


Figura 4: Esquema do processo de construção de ontologias

Assim pode afirmar-se que a ontologia constitui apenas uma representação da informação obtida. Como tal a qualidade da ontologia depende directamente da qualidade e quantidade de informação fornecida.

CONCLUSÕES

Este trabalho representa um esforço para convergir as diversas visões sobre a representação da geografia física num modelo comum e disponibilizar esse conhecimento integrado de modo processável. Um mapeamento da informação relativa ao domínio da geografia física de Portugal também está em desenvolvimento.

Outro objectivo deste trabalho é estender a actual ontologia de Portugal Geo-Net-PT01, produzindo uma nova versão que integre o domínio da geografia física; o resultado será a publicação da Geo-Net-PT02 que, tal como a versão anterior, será disponibilizada gratuitamente à comunidade científica.

Por fim, representações de conhecimento geográfico na forma de ontologias estão disponíveis na literatura, como por exemplo (Manov et al., 2003). Em nosso trabalho, tentaremos reutilizar tais representações sempre que possível, principalmente os conceitos em nível mais abstrato, os quais tendem a ser mais consensuais.

REFERÊNCIAS

- Berners. T.B.; Hendler; J., Lassila, O. The Semantic Web. Scientific American, Maio 2001. Disponível em: <http://www.scientificamerican.com/2001/0501issueberners-lee.html>> Acesso em: 20 Outubro 2006.
- Chaves, Marcirio Silveira; Silva, Mário J. e Martins, Bruno. A Geographic Knowledge Base for Semantic Web Applications. 20º Simpósio Brasileiro em Banco de Dados- SBBD, Uberlândia, Minas Gerais, Brazil, pp. 40-54, 3-7 October, 2005.
- Chaves, Marcirio Silveira; Martins, Bruno; Silva, Mário J.. GKB - Geographic Knowledge Base. Department of Informatics, University of Lisbon, DI/FCUL TR, 05--12, 2005.
- Freitas, Sérgio; Afonso, Ana Paula e Silva, Mário J. Mobile Geotumba: Geographic Information Retrieval System for Mobile Devices. Proc. of the 4th MiNEMA Workshop, July 2006.
- Gruber, T. A Translation Approach to Portable Ontology Specifications *Knowledge Acquisition*, 5 pp.199-220. 1993.
- Hendler, J. Agents and the semantic web. IEEE Intelligent Systems, Mar./Abr. 2001.
- Manov, Dimitar; Kiryakov, Atanas; Popov, Borislav; Bontcheva, Kalin; Maynard, Diana; Cunningham, Hamish. Experiments with geographic knowledge for information extraction}. Proceedings of the HLT-NAACL 2003 workshop on Analysis of geographic references Association for Computational Linguistics, Morristown, NJ, USA pp. 1—9, 2003.
- Schmidt, Douglas. Model-driven Engineering, IEEE Computer, Fevereiro de 2006, pp 25-31.
- Silva, Mário J.; Martins, Bruno; Chaves, Marcirio Silveira; Cardoso, Nuno e Afonso, Ana Paula. Adding Geographic Scopes to Web Resources. CEUS - Computers, Environment and Urban Systems, Elsevier Science, Volume 30, Issue 4, pp. 378-399, July 2006.