Criação e expansão de geo-ontologias, dimensionamento de informação geográfica e reconhecimento de locais e seus relacionamentos em textos

Marcirio Silveira Chaves
Pólo XLDB da Linguateca
LaSIGE - Departamento de Informática
Faculdade de Ciências da Universidade de Lisboa
1749-016 Lisboa, Portugal
mchaves@di.fc.ul.pt

1 Introdução

Este artigo resume o trabalho desenvolvido ao longo de mais de quatro anos na Linguateca no âmbito do meu doutorado. Até 2004, a maior parte das fontes de dados geográficos de Portugal encontrava-se distribuída, desintegrada e desconexa. Essas fontes contêm informação complementar, heterogênea e semi-estruturada. Qualquer aplicação que necessitasse utilizá-las tinha que recorrer a diversos bancos de dados, estudar seus esquemas conceituais e traduzir a informação para um formato comum de representação, entre outras tarefas. Além disso, os dados armazenados em banco de dados proprietários são invisíveis para aplicações da Web Semântica.

Nesse contexto havia a necessidade da criação de um modelo genérico suficiente para reunir informação geográfica de diversas fontes, de múltiplos domínios geográficos (e.g. administrativo e físico) e disponibilizá-la de forma integrada e em um formato legível por máquina. Assim, foi criada a GKB (*Geographic Knowledge Base*) (Chaves et al., 2005a,b), um sistema de gerenciamento de conhecimento geográfico, ilustrado na Figura 1 e descrito na próxima seção.

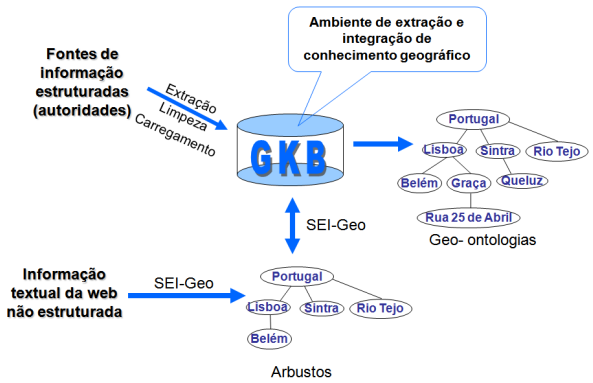

Figura 1: Arquitetura global do sistema de gerenciamento de conhecimento geográfico.

Este artigo está estruturado como segue: a Seção 2 apresenta a GKB. A Seção 3 descreve algumas das aplicações que utilizam as geo-ontologias geradas pela GKB. A Seção 4 introduz os resultados obtidos

com experimentos para dimensionar a geograficidade[1] de textos em português. A Seção 5 descreve o sistema de extração, anotação e integração de conhecimento geográfico (SEI-Geo). A Seção 5.1 apresenta as avaliações realizadas com o SEI-Geo e a Seção 6 conclui o artigo.

2 Geographic Knowledge Base -GKB

A GKB é um dos componentes desenvolvidos no Pólo XLDB da Linguateca em colaboração com o projeto *Geographic Reasoning for Search Engines* (GREASE) (`http://xldb.di.fc.ul.pt/grease`), o qual pesquisa métodos, algoritmos e arquiteturas de software para atribuir âmbitos geográficos para recursos da rede e para recolher documentos usando entidades geográficas.

A GKB é um ambiente de extração e integração de conhecimento geográfico que contém informações provenientes de fontes de dados administrativas semi-estruturadas de autoridades junto com um conjunto de regras para integração de informação. A expansão do conhecimento contido na GKB ocorre com informação proveniente de textos. Esses textos são a entrada de informação para o Sistema de Extração, Anotação e Integração de Conhecimento Geográfico (SEI-Geo), que é o responsável por gerar uma representação estruturada do conhecimento geográfico extraído e integrá-lo no repositório da GKB.

A GKB suporta a definição de relacionamentos ontológicos entre entidades, tais como meronímia, sinonímia e adjacência, entre outros. A GKB também suporta relacionamentos inter-domínios, os quais são associações entre entidades de domínios diferentes. Por exemplo, o âmbito geográfico[2] de uma entidade do domínio de rede é representado como um relacionamento entre um sítio da rede (entidade do domínio da Internet) e uma região geográfica (uma entidade do domínio geográfico).

A informação armazenada no repositório da GKB pode ser exportada com uma ferramenta nomeada GOG (*Geographic Ontology Generator*). A GOG permite selecionar partes da informação armazenada na GKB, uma vez que os repositórios da GKB têm, atualmente, cerca de meio milhão de entidades e o usuário raramente quer receber toda a informação. A GOG exporta a informação no formato OWL, uma representação que estende o RDF[3] e, consequentemente, é uma ocorrência de XML. A geo-ontologia completa de Portugal (Geo-Net-PT) contém mais de 400.000 entidades e é um recurso público disponível em `http://xldb.fc.ul.pt/geonetpt`.

As geo-ontologias exportadas pela GKB têm sido utilizadas por diversas aplicações que incluem: sistemas para reconhecimento de entidades mencionadas (REM), um classificador de documentos de acordo com seu âmbito geográfico, uma interface de recolha de informação para consultas geográficas e uma interface XML para consultas a almanaques geo-temporais, entre outras.

3 Aplicações que Utilizam as Geo-ontologias Geradas a partir da GKB

3.1 Sistemas de REM

CAGE: é um sistema de REM e de atribuição de âmbito geográfico a páginas da rede (Silva et al., 2006; Martins et al., 2007). O Cage utiliza as geo-ontologias geradas a partir da GKB nas fases de identificação e desambiguação de locais (Cardoso et al., 2005). Martins et al. (2007) apresentam a arquitetura do CAGE, bem como a descrição detalhada do uso das geo-ontologias.

Faísca: é um sistema de reconhecimento de locais que faz uso dos conceitos e ocorrências contidos nas geo-ontologias geradas a partir da GKB (Cardoso et al., 2008). Faísca não explora os relacionamentos existentes entre conceitos nas ontologias, mas utiliza os conceitos para desambiguar nomes de locais na fase de REM.

[1] Por geograficidade entende-se a quantidade de informação geográfica presente em textos.

[2] Nesse artigo, entende-se âmbito geográfico como a região geográfica, se ela existe, onde a média das pessoas pensa ser mais relevante para uma página, sítio ou domínio da rede. Por exemplo, o âmbito geográfico do sítio da Câmara de Lisboa (*www.cm-lisboa.pt*) é o concelho de Lisboa.

[3] `http://www.w3.org/TR/REC-rdf-syntax/`

3.2 Módulos de um Sistema de Recolha de Informação Geográfica

As geo-ontologias geradas a partir da GKB têm sido utilizadas por diversos módulos do sistema de recolha de informação geográfica da Universidade de Lisboa no GeoCLEF 2007 (Cardoso et al., 2008).

QueOnde: é um módulo que utiliza as geo-ontologias para dividir o tópico de uma consulta em três partes: 'O que', 'Relacionamento espacial' e 'Onde'. Por exemplo, para o tópico 'tráfego marítimo nas ilhas portuguesas', QueOnde consulta a geo-ontologia e verifica que 'portuguesas' é um adjetivo de Portugal e que 'ilhas' é um conceito geográfico.

QuerCol: é um módulo que utiliza a geo-ontologia para fazer expansão de consulta. QuerCol interpreta uma consulta como duas partes: 'O que' e 'Onde'. A geo-ontologia é usada para expandir o(s) termo(s) da parte 'Onde'. Por exemplo, na consulta 'regiões vinícolas em Portugal', o módulo QuerCol expande o nome Portugal para todas as províncias, distritos, concelhos e freguesias existentes na geo-ontologia e que fazem parte de Portugal.

Outro módulo do sistema que utilizou as geo-ontologias geográficas é o sistema de reconhecimento de locais Faísca, descrito na seção anterior.

3.3 Interface de Motor de Pesquisa Geográfica

A GKB é usada também na interface do protótipo Geotumba, um sistema para recolha de informação geográfica (ver Figura 2). No campo `Local?` o usuário digita a região, a rua, o código postal ou

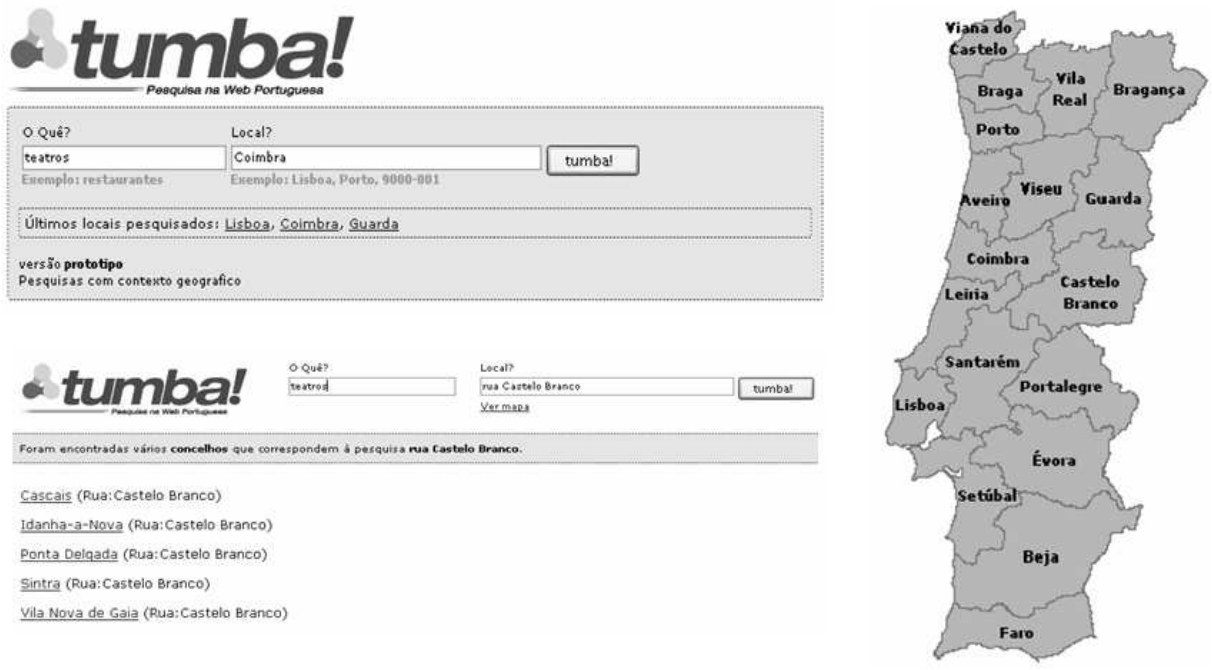

Figura 2: Exemplos de interfaces para recolha de informação geográfica usando a GKB.

outra entidade geográfica para reduzir o âmbito da consulta. Quando um nome geográfico ambíguo é detectado na consulta, Geotumba apresenta as possíveis alternativas para desambiguação da mesma. Por exemplo, o nome "rua Castelo Branco" ocorre em cinco concelhos diferentes na Geo-Net-PT, os quais são apresentados no lado esquerdo inferior da Figura 2. Além da consulta por texto, o usuário pode utilizar os mapas para definir o âmbito da consulta.

3.4 Interface para Consultas a Almanaques Geo-temporais

A Geo-Net-PT também é utilizada no projeto DIGMAP[4] (*Discovering our Past World with Digitized historical Maps*) (Borbinha et al., 2007), especificamente em uma interface XML para consultas a almanaques geo-temporais. Neste serviço, a Geo-Net-PT é integrada com outros almanaques existentes considerando a dimensão temporal juntamente com o conteúdo geográfico dos almanaques. A Figura 3 apresenta a interface do sistema.

Figura 3: Interface para Consultas a Almanaques Geo-temporais.

Para cada local inserido pelo usuário, o sistema de consultas a almanaques geo-temporais percorre os almanaques e apresenta o nome do local juntamente com seus metadados, relacionamentos e população, entre outras informações subjacentes a cada almanaque. A informação geográfica é apresentada em diversas linguagens (e.g. XML, OWL e KML - *Keyhole Markup Language*), conforme o almanaque disponibiliza.

No exemplo da Figura 3, o sistema apresenta os metadados sobre o 'distrito de Beja', os quais incluem os relacionamentos de `parte-de`, `contém` e `adjacência`. Na parte inferior da figura, estão nove almanaques que contêm informação sobre o 'distrito de Beja'. No canto superior direito, o 'distrito de Beja' é ilustrado no mapa.

Além dessas aplicações, a Geo-Net-PT tem sido requisitada por diversos grupos de pesquisa ao redor do mundo. A Figura 4 apresenta a distribuição geográfica dos pedidos por países. A Geo-Net-PT já foi requisitada por dezenas de investigadores, na sua maioria de Portugal e do Brasil, evidenciando o interesse da comunidade em estruturas de representação de conhecimento geográfico.

Por fim todo o conteúdo das geo-ontologias geradas pela GKB pode ser visualizado com a interface Geobase, apresentada na Figura 5.

As aplicações que utilizam as geo-ontologias geradas pela GKB necessitam de informação geográfica além daquela proveniente de fontes de informação estruturadas e semi-estruturadas. Nomes históricos e alternativos de locais, por exemplo, ainda não estão na GKB, mas podem ser encontrados em textos.

[4] `http://gaz.digmap.eu/`

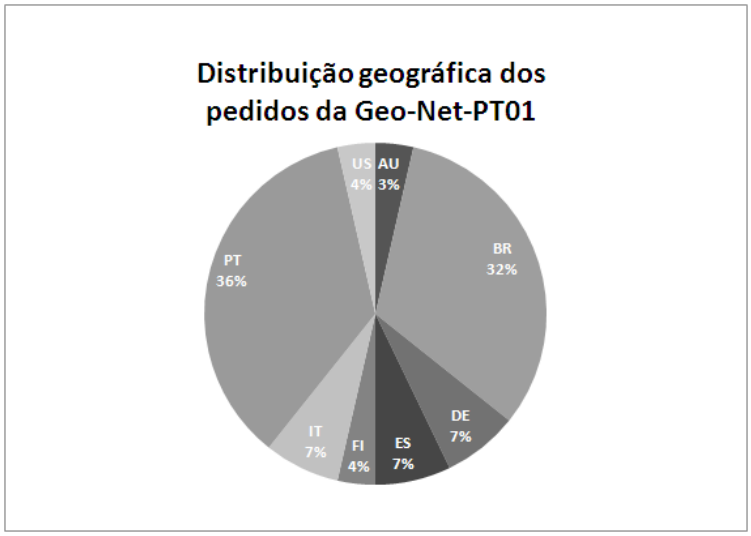

Figura 4: Distribuição geográfica dos pedidos da Geo-Net-PT por países.

Programas foram implementados para dimensionar a geograficidade de textos em português e para conhecer a sobreposição da informação armazenada na GKB com a informação geográfica em textos.

4 Geograficidade de Textos

Para verificar a geograficidade em textos da rede, foi utilizado o corpus WPT 03, uma coleção da rede portuguesa de 2003, com 12 Gbytes e 3.7 milhões de páginas e 1.6 bilhões de palavras (`http://linguateca.di.fc.ul.pt/WPT03/`) (Cardoso et al., 2007). Aproximadamente 68.6% dessas páginas estão em português e mais de 1.5 milhão de paginas distintas. O sistema de REM utilizado foi o SIEMÊS (Sarmento, 2006), que na avaliação conjunta do Primeiro HAREM (Santos et al., 2006) alcançou 70% de precisão e 75% de abrangência para a categoria local. Entretanto, a versão utilizada em nossos experimentos é uma versão com melhoramentos sobre aquela utilizada no Primeiro HAREM.

A caracterização da informação geográfica em textos passa pela verificação da presença de nomes de locais em nomes de pessoas e organizações. Considerando uma amostra aleatória de 32.000 documentos da rede, os resultados evidenciam que 31% das entidades mencionadas distintas da categoria pessoa e 23,43% das entidades mencionadas distintas da categoria organização contêm um nome geográfico incluído na Geo-Net-PT.

Para investigar se o tipo de local ocorrendo em textos da rede portuguesa tinha diferentes propriedades (granularidade, geografia física (rios, montanhas, etc.)), foram verificados os tipos das entidades mencionadas da categoria local que o SIEMÊS encontrou após ser executado sobre a mesma amostra de 32.000 documentos. O resultado mostrou que 85% dos tipos de locais reconhecidos pelo SIEMÊS estão concentrados em apenas três (povoamento, endereço completo e sociedade/cultura) dos tipos de locais definidos no Primeiro HAREM. Estatísticas mais detalhadas sobre a geograficidade em textos estão em (Chaves e Santos, 2006).

Quanto aos tipos de arruamentos, os predominantes na geografia administrativa de Portugal são ruas e travessas. Somente ruas representam mais de 60% dos tipos de arruamentos do país. Rua também é o tipo de arruamento mais freqüente no WPT 03, após o tipo ambíguo acesso. Por outro lado, as travessas ocorrem com bem menos freqüência no WPT 03, sendo apenas o 28º tipo de arruamento mais freqüente.

Cera de 60% dos nomes[5] presentes na Geo-Net-PT estão presentes no WPT 03. Aqueles compostos por quatro palavras são os menos freqüentes, ao passo que os nomes formados por uma palavra atingem

[5]Nomes na Geo-Net-PT acima do nível hierárquico de arruamento, ou seja, todos os nomes da Geo-Net-PT exceto nomes de arruamentos e códigos-postais.

Figura 5: Geobase: interface de visualização da Geo-Net-PT.

quase 80% de presença nesse corpus da rede. Outros resultados sobre a presença de informação geográfica de ontologias em textos e sobre a ambigüidade existente entre nomes de uma ontologia são descritos em (Santos e Chaves, 2006).

Após verificar que existe informação geográfica presente em textos suficiente para expandir geo-ontologias, foi desenvolvido o SEI-Geo.

5 Sistema de Extração, Anotação e Integração de conhecimento Geográfico - SEI-Geo

O SEI-Geo foi desenvolvido no Pólo XLDB da Linguateca no âmbito do projeto GREASE e tem como objetivo reconhecer o conhecimento geográfico disponível em textos, gerar uma representação estruturada desse conhecimento e integrá-lo em geo-ontologias. O sistema é composto por dois módulos principais: o de Extração de Informação Geográfica (EIG) e o de Integração de Conhecimento Geográfico (ICG).

O EIG recebe como entrada um conjunto de textos que são segmentados em frases. O EIG contém uma quantidade abrangente de regras que indicam a presença de conceitos e relacionamentos nas frases. Tais frases, juntamente com conceitos de geo-ontologias, são a entrada de uma função que extrai frases com potencial conteúdo geográfico. Essas frases são a entrada de dois sub-módulos: o extrator de arbustos[6] e o anotador. O extrator de arbustos detecta ocorrências geográficas e relacionamentos semânticos e tem uma função de filtro, na qual o conteúdo geográfico, duplicado ou sobreposto, é eliminado. O resultado desse processo é um conjunto de arbustos que são utilizados como entrada no ICG. O anotador insere etiquetas com nomes de categoria semântica, tipo e subtipo. O anotador também possui a capacidade de reconhecer e anotar relacionamentos entre locais.

O ICG recebe os arbustos extraídos e o conhecimento armazenado na GKB, faz a integração e retorna para a GKB o conhecimento geográfico expandido. A integração de conhecimento textual em geo-ontologias concentra-se em encontrar informação geográfica complementar àquela existente nas geo-ontologias e integrar essa informação no nível de granularidade mais adequado nas geo-ontologias. A integração de conhecimento geográfico com o SEI-Geo ocorre quando novos fatos geográficos são descobertos em texto.

[6]Um arbusto é composto por pelo menos duas entidades geográficas candidatas a locais e um relacionamento. Esse conjunto forma uma tripla. Não há número máximo de entidades e relacionamentos pré-definido.

5.1 Avaliação do SEI-Geo

O SEI-Geo tem sido avaliado na sua capacidade de extrair, anotar e integrar conhecimento geográfico. O SEI-Geo participou no Segundo HAREM e conseguiu atingir resultados satisfatórios no cenário seletivo de identificação e reconhecimento de locais. Considerando somente a medida F, o SEI-Geo foi o segundo melhor sistema nesse cenário com 0,5953, enquanto o melhor sistema atingiu 0,6246 na tarefa de classificação semântica. A participação do SEI-Geo no Segundo HAREM é descrita em (Chaves, 2008).

Além da tarefa de anotação de textos, o SEI-Geo foi avaliado, através de testes de mutilação, na sua capacidade de extrair locais e recompor uma geo-ontologia existente. Testes de mutilação consistem na destruição de parte de um objeto de estudo e na sua reconstrução. Especificamente quando se trata de estruturas de representação de conhecimento tal como ontologias, um (ou vários) nível da hierarquia de conceitos e ocorrências é destruído e a partir de informação textual tenta-se reconstruir a informação retirada inicialmente. Para implementar esse teste, foram retiradas todas as ocorrências do tipo de entidade `ISO-3166-1` (que corresponde a países e territórios) da ontologia WGO (Martins et al., 2006). Todos os arbustos extraídos pelo SEI-Geo que contêm o tipo de entidade `ISO-3166-1` foram enviados à geo-ontologia com o objetivo de encontrar um identificador para cada entidade geográfica reconhecida. A Tabela 1 apresenta os resultados dos testes de mutilação nas quatro corpos jornalísticos, duas do Público e duas da Folha de São Paulo (Santos e Rocha, 2004).

Tabela 1: Resultado do teste de mutilação para países e territórios nos corpora jornalísticos.

	Público 1994	Público 1995	FSP 1994	FSP 1995
SEI-Geo mutilado	148 (70,47%)	161 (76,30%)	117 (62,56%)	109 (60,55%)
ISO-3166-1 na coleção	210	211	187	180

Conforme a Tabela 1, o Público é uma fonte mais rica em informação geográfica ao nível de países e territórios do que a Folha de São Paulo. Dos 211 países e territórios existentes no corpus do Público do ano de 1995, 161 (70,30%) foram reconhecidos e representados em triplas no formato de arbusto. Das 238 ocorrências do tipo de entidade `ISO-3166-1` da WGO, 211 ocorrem nesse corpus. Um dos fatores que levam o Público a conter mais locais da WGO é os nomes de locais estarem na sua maioria descritos no português de Portugal. Exemplos desses casos encontrados no Público e ausentes na Folha de São Paulo são: 'Coreia do Sul', 'Eslovénia' e 'Ilhas Caimão'. Os resultados dos testes de mutilação indicam que o SEI-Geo é capaz reconstituir uma geo-ontologia recebendo como entrada conceitos sem ocorrências.

Quanto à expansão de geo-ontologias, o SEI-Geo recebe como entrada um corpus e geo-ontologias e devolve como resultado um conjunto de arbustos com as geo-ontologias enriquecidas com novos locais e relacionamentos reconhecidos no corpus. Esses locais podem ou não estar presentes na geo-ontologia. Se o SEI-Geo encontra uma ocorrência de um conceito e essa ocorrência já está na geo-ontologia, o resultado permite validar a ocorrência e a geo-ontologia não é expandida.

A primeira avaliação foi realizada no corpus do Público do ano de 1995. De um total de 50.495 arbustos, foi selecionada aleatoriamente uma amostra de 100 arbustos compostos por 143 triplas. Cada tripla dessa amostra foi avaliada manualmente de acordo com os seguintes critérios:

Integrável (I): quando as duas entidades geográficas da tripla forem realmente locais e a relação entre elas estiver correta.

Integrável com Assistência (IA): quando duas entidades geográficas forem corretas e não existir relacionamento explícito no texto ou o algoritmo não conseguiu identificar. Nesse caso o avaliador deve inserir o relacionamento correto.

Existente (E): quando as entidades geográficas e o relacionamento reconhecido entre essas entidades geográficas já está em pelo menos uma das ontologias.

Falso (F): quando no máximo uma entidade geográfica da tripla é um local ou as duas entidades geográficas não possuem relacionamento no mundo real.

Tabela 2: Resultado da avaliação das triplas dos arbustos para expansão de geo-ontologias.

I	IA	E	F
2	60	19	67

A Tabela 2 apresenta os resultados alcançados, os quais indicam que a maior parte das triplas integráveis são integráveis com assistência. Ainda resta um número elevado de triplas falsas, mas esses valores já eram esperados dados os resultados da participação do SEI-Geo no Segundo HAREM.

6 Considerações Finais

Este artigo resumiu meu trabalho no âmbito da Linguateca ao longo dos últimos anos. A base de conhecimento geográfico armazena o conteúdo exportado como geo-ontologias que estão disponíveis publicamente. Esse conteúdo geográfico é expandido com informação textual extraída pelo SEI-Geo. O SEI-Geo foi avaliado no Segundo HAREM no que diz respeito à sua capacidade de anotação de locais e também apresentou resultados encorajadores nos testes de mutilação e expansão de geo-ontologias.

Após as geo-ontologias terem sido utilizadas por várias aplicações, torna-se iminente a criação de uma geo-ontologia mundial com nomes de locais em português, abrangendo as variantes da língua de Portugal e do Brasil. Essa nova geo-ontologia pode ser criada reutilizando o modelo-base no qual a GKB foi concebida.

Referências

José Luis Borbinha, Gilberto Pedrosa, Diogo Reis, João Luzio, Bruno Martins, João Gil e Nuno Freire. DIGMAP - Discovering Our Past World with Digitised Maps. Em *ECDL*, 2007. p. 563–566.

Nuno Cardoso, Bruno Martins, Marcirio Silveira Chaves, Leonardo Andrade e Mário J. Silva. The XLDB Group at GeoCLEF 2005. Em Carol Peters, Fredric C. Gey, Julio Gonzalo, Henning Müller, Gareth J. F. Jones, Michael Kluck, Bernardo Magnini e Maarten de Rijke, editores, *Accessing Multilingual Information Repositories, 6th Workshop of the Cross-Language Evalution Forum, CLEF 2005, Vienna, Austria, 21-23 September, 2005, Revised Selected Papers*, 2005. p. 997–1006. ISBN 3-540-45697-X.

Nuno Cardoso, Bruno Martins, Daniel Gomes e Mário J. Silva. *WPT 03: Recolha da Web Portuguesa*, IST Press, 2007.

Nuno Cardoso, David Cruz, Marcirio Silveira Chaves e Mário J. Silva. Using Geographic Signatures as Query and Document Scopes in Geographic IR. Em *Advances in Multilingual and Multimodal Information Retrieval: 8th Workshop of the Cross-Language Evaluation Forum, CLEF 2007*, 2008. p. 802–810. Revised Selected papers.

Marcirio Silveira Chaves. Geo-ontologias para reconhecimento de relações entre locais: a participação do SEI-Geo no Segundo HAREM. Em Cristina Mota e Diana Santos, editor, *Desafios na avaliação conjunta do reconhecimento de entidades mencionadas: Actas do Encontro do Segundo HAREM.*, 7 de Setembro de 2008.

Marcirio Silveira Chaves e Diana Santos. What kinds of geographical information are there in the Portuguese Web? Em Vieira et al. (2006). ISBN 3-540-34045-9.

Marcirio Silveira Chaves, Mário J. Silva e Bruno Martins. A Geographic Knowledge Base for Semantic Web Applications. Em Carlos Alberto Heuser, editor, *Proc. of the 20th Brazilian Symposium on Databases*, October, 3–7 de 2005. p. 40–54.

Marcirio Silveira Chaves, Mário J. Silva e Bruno Martins. GKB - Geographic Knowledge Base. DI/FCUL TR 05–12, Departamento de Informática da Faculdade de Ciências da Universidade de Lisboa, Julho de 2005.

Bruno Martins, Nuno Cardoso, Marcirio Silveira Chaves, Leonardo Andrade e Mário J. Silva. The University of Lisbon at GeoCLEF 2006. Em Carol Peters, Paul Clough, Fredric C. Gey, Jussi Karlgren, Bernardo Magnini, Douglas W. Oard, Maarten de Rijke e Maximilian Stempfhuber, editores, *CLEF*, 2006. p. 986–994. ISBN 978-3-540-74998-1.

Bruno Martins, Mário J. Silva e Marcirio Silveira Chaves. *O Sistema CaGE no HAREM - Reconhecimento de Entidades Geográficas em Textos da Língua Portuguesa*, Linguateca, 2007. ISBN: 978-989-20-0731-1.

Diana Santos e Marcirio Silveira Chaves. The place of place in geographical IR. Em *Proc. of the 3rd Workshop on Geographic Information Retrieval, SIGIR'06*, August 10th de 2006. p. 5–8.

Diana Santos e Paulo Rocha. The Key to the First CLEF with Portuguese: Topics, Questions and Answers in CHAVE. Em Carol Peters, Paul Clough, Julio Gonzalo, Gareth J. F. Jones, Michael Kluck e Bernardo Magnini, editores, *CLEF*, 2004. p. 821–832. ISBN 3-540-27420-0.

Diana Santos, Nuno Seco, Nuno Cardoso e Rui Vilela. HAREM: an Advanced NER Evaluation Contest for Portuguese. Em *Proceedings of LREC'2006*, 22-28 May de 2006. p. 1986–1991.

Luis Sarmento. SIEMÊS - a named entity recognizer for Portuguese relying on similarity rules. Em Vieira et al. (2006). ISBN 3-540-34045-9.

Mário J. Silva, Bruno Martins, Marcirio Silveira Chaves, Nuno Cardoso e Ana Paula Afonso. Adding Geographic Scopes to Web Resources. *CEUS - Computers, Environment and Urban Systems - Elsevier Science*, 30(4):378–399, Julho de 2006.

Renata Vieira, Paulo Quaresma, Maria das Graças Volpe Nunes, Nuno J. Mamede, Claudia Oliveira e Maria Carmelita Dias, editores. *Computational Processing of the Portuguese Language, 7th International Workshop, PROPOR 2006, Itatiaia, Brazil, May 13-17, 2006, Proceedings*, volume 3960 de *Lecture Notes in Computer Science*, 2006. Springer. ISBN 3-540-34045-9.