


**UNIVERSIDADE DE LISBOA**  
**FACULDADE DE CIÊNCIAS**  
**DEPARTAMENTO DE INFORMÁTICA**

**UMA METODOLOGIA PARA  
 CONSTRUÇÃO DE GEO-  
 ONTOLOGIAS**

**Marcio Silveira Chaves**  
**Orientadores: Mário J. Silva e Diana Santos**

**Doutoramento em Informática**  
**Especialidade Engenharia Informática**

Lisboa, 16 de outubro de 2009.

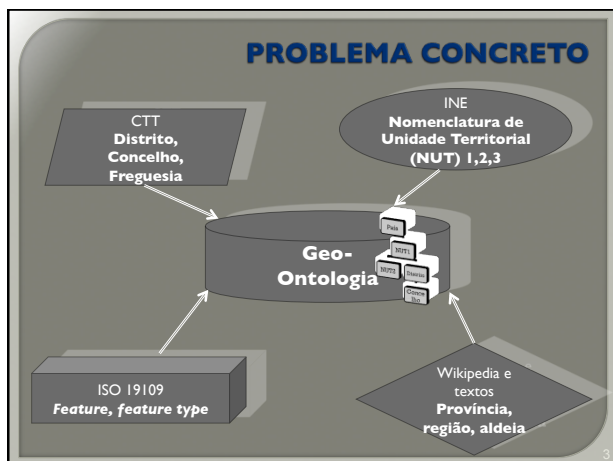
## GEO-ONTOLOGIA

- Geo-ontologia: um conjunto de conceitos geográficos e relacionamentos geográficos definidos formalmente e sem ambiguidade.

Serve para:

- Normalização de terminologia
- Integrar base de dados e textos (85%)

2



## SUMÁRIO

- Desafios
- Solução Proposta
- Estado da Arte
- *Geographic Knowledge Base* (GKB)
- Geograficidade
- Sistema de Extração e Integração de Conhecimento Geográfico (SEI-Geo)
- Metodologia Proposta
- Contribuições

4

## DESAFIOS

- Informação em bases de dados geográficas com baixa qualidade, escondida e sub-utilizada
- Ontologias carecem de ocorrências e, na prática, não estão disponíveis
- Tratamento de informação geográfica em texto
- Ausência de detalhes nas metodologias para construção de ontologias

5

## SOLUÇÃO PROPOSTA

Etapa	Solução Proposta
Encontrar, limpar e integrar informação proveniente de bases de dados geográficas com informações complementares	GKB
Realizar estudos quantitativos	Caracterização da geograficidade presente em textos da web em português
Reconhecer o conhecimento disponível em textos e gerar uma representação formal desse conhecimento	SEI-Geo - Extrator
Integrar a ontologia gerada por um sistema de extração e integração de conhecimento geográfico numa ontologia existente	SEI-Geo - Integrador

6

### ESTADO DA ARTE

#### Representação de conhecimento

	Tipo de relacionamento	Grau de formalidade	Especialista no domínio
Classificação facetada	livre	baixo	sim
Mapa de tópicos	livre	baixo	não
Vocabulário controlado	-	baixo	sim
Taxonomia	É um/parte de	alto	sim
Tesouro	BT/NT SN Use/Use For	alto	sim
Meta-modelo	Livre, porém limitado	baixo	sim
Folksonomia	rasa	baixo	não
<b>Ontologia</b>	<b>livre</b>	<b>alto</b>	<b>sim</b>

7

### ESTADO DA ARTE

#### Sistemas de extração e integração de informação geográfica

	Padrões	Ontologia	Integra conhecimento	Geo	PT
Snowball	✓	X	X	✓	X
Ontolearn	X	✓ (WordNet)	✓	X	X
KnowItAll/KnowItNow	✓	X	X	X	X
OntoSyphon	✓	✓	X	X	X
OnLocus e Endereçamento	✓	✓ (ênfase endereços)	X	✓	✓
<b>SEI-Geo</b>	✓	✓	✓	✓	✓

8

### ESTADO DA ARTE

#### Metodologias para construção de ontologias

	TGN	SPIRIT	Geo-Names	GKB
Limpeza de dados	proposta	não proposta	não proposta	<b>descrita em detalhe</b>
Integração de conhecimento	regras	similaridade	similaridade	<b>regras</b>
Formato	XML	XML, RDF e OWL	XML, RDF e OWL	<b>XML, RDF e OWL</b>
Multi-lingua	sim	não	sim	<b>sim</b>
Versionamento	mês/ano	N/D	variável	<b>variável</b>
Documentação	informal	formal	Informal	<b>formal</b>

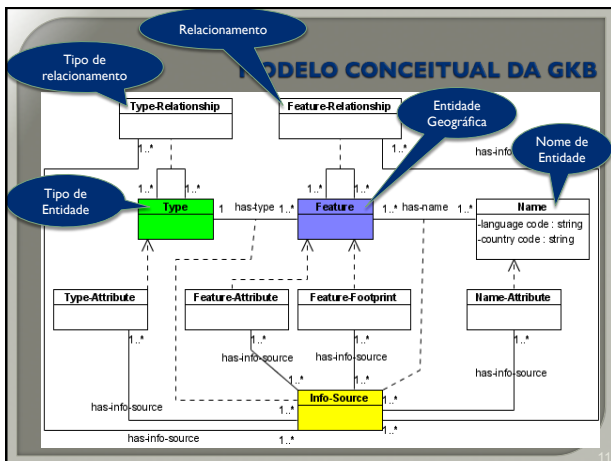
9

### SISTEMA DE GESTÃO DE CONHECIMENTO GEOGRÁFICO



10

### MODELO CONCEITUAL DA GKB



11

### GEO-ONTOLOGIAS PRODUZIDAS

- WGO – World Geographic Ontology
  - Mais de 13 mil entidades geográficas (EG) (mais de 10 mil distintas)
  - Ca. 25.000 relacionamentos
- Geo-Net-PT01
  - Mais de 400 mil EG (34 mil termos distintos)
  - Mais de 75% dos termos formados por 1 ou 3 palavras
  - 45% dos termos formados por uma palavra estão presentes no nome de outras EG formadas por mais de uma palavra

12

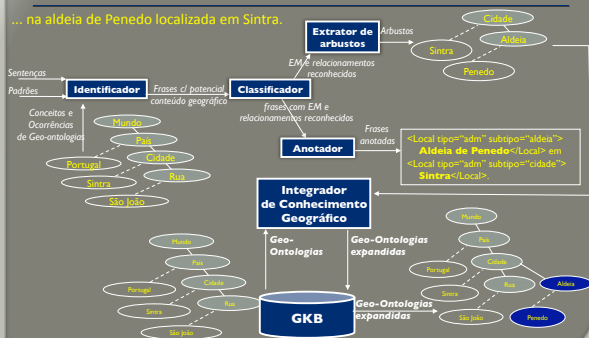
## GEOGRAFICIDADE DA WEB PORTUGUESA

- 78,8% dos termos formados por 1 palavra da Geo-Net-PT estão no WPT 03
  - nomes de locais em nomes de pessoas e organizações
    - amostra aleatória de 32.000 documentos do WPT 03
    - 31% das entidades mencionadas distintas da categoria **pessoa** e
    - mais de 23% das entidades mencionadas distintas da categoria **organização**
- contêm um nome geográfico incluído na Geo-Net-PT**

13

## SEI-Geo

... na aldeia de Penedo localizada em Sintra.



14

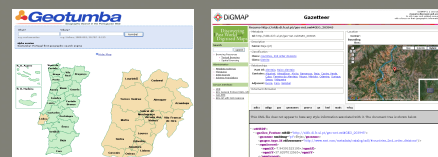
## METODOLOGIA PROPOSTA NESTA TESE

- Concepção de um modelo conceitual
  - Sub-domínios, relacionamentos (intra-inter), variantes, diferenças de opinião, atributos das EG, etc.
- Seleção e limpeza de fontes de informação
  - Autoridade, custo de aquisição, tipo de licenciamento e formato
- Integração de conhecimento
- Exportação de conhecimento e as aplicações

15

## VALIDAÇÃO INDIRETA DA METODOLOGIA

- Sistemas de REM (SEI-Geo e CaGE)
- Módulos do Sistema de Recuperação de Informação Geográfica da UL
- Interfaces



- Sistemas que utilizaram as geo-ontologias produzidas participaram em avaliações internacionais (GeoCLEF e HAREM) com resultados significativos.

16

## CONTRIBUIÇÕES

- Metodologia para construção de geo-ontologias proposta e validada**
  - Sistema de gestão de conhecimento geográfico – GKB
  - Análise da geofricidade
  - Sistema de Extração e Integração de Conhecimento Geográfico – SEI-Geo
  - Geo-ontologias públicas e gratuitas – WGO e



17

## PRINCIPAIS PUBLICAÇÕES

- Chaves, Márcio Silveira. **Geo-ontologias e padrões para reconhecimento de locais e suas relações em textos: a participação do SEI-Geo no Segundo HAREM**. In: Cristina Mota e Diana Santos (eds.). *Desafios na avaliação conjunta do reconhecimento de entidades mencionadas: O Segundo HAREM*. Linguateca, p. 231–245, 2009. ISBN: 978-989-20-1656-6
- Chaves, Márcio Silveira; Rodrigues, Catarina e Silva, Mário J.. **Data Model for Geographic Ontologies Generation**. XATA2007 - XML: Aplicações e Tecnologias Associadas. 15-16 de Fevereiro, Lisboa, Portugal, 2007.
- Santos, Diana e Chaves, Márcio Silveira. **The place of place in geographical IR**. In 3rd Workshop on Geographic Information Retrieval, SIGIR 2006, p. 5-8, August 10th, Seattle, 2006.
- Chaves, Márcio Silveira e Santos, Diana. **What kinds of geographical information are there in the Portuguese Web?**. In Renata Vieira, Paulo Quaresma, Maria da Graça Volpes Nunes, Nuno Mamede, Cláudia Oliveira & Maria Carmelita Dias (eds.), *Encontro para o Processamento Computacional da Língua Portuguesa Escrita e Falada (PROPOR 2006)*, LNAI 3960 - Springer, (Itatiaia, RJ, 13 a 17 de maio), pp. 264-267. ISBN 3-540-34045-9
- Silva, Mário J.; Martins, Bruno; Chaves, Márcio Silveira; Cardoso, Nuno; Afonso, Ana Paula. **Adding Geographic Scopes to Web Resources**. CEUS - Computers, Environment and Urban Systems, Elsevier Science, volume 30, issue 4, July, pages 378-399, 2006.
- Chaves, Márcio Silveira; Silva, Mário J. and Martins, Bruno. **A Geographic Knowledge Base for Semantic Web Applications**. 20th Brazilian Symposium on Databases - SBBD, Uberlândia, Minas Gerais, Brazil, p. 40-54, 3-7 October, 2005.

18