

Recuperação de Informação Geográfica

Aluno: Hendrickson Reiter Langbehn

Introdução

- Recuperação de Informação:

- “Ciência que trata da busca de documentos, informação contida em documentos e de metadados sobre documentos, além da busca em Bancos de Dados Relacionais e na Web.”

(Wikipedia)

- “Técnica e processo de buscar, recuperar e interpretar informações provenientes de grandes quantidades de dados armazenados.”

(Answers.com)

Introdução

- Motivação
 - Aproximadamente 14% das consultas feitas à máquinas de busca contêm termos geográficos.
 - Aproximadamente 20% das páginas Web incluem identificadores geográficos.
 - Máquinas de busca ainda não fazem tratamento especial para termos geográficos.
 - Páginas relevantes não recuperadas
 - Páginas não relevantes recuperadas

Trabalhos Relacionados

- Adding Geographic Scopes to Web Resources.
 - Mário J. Silva, Bruno Martins, Marcirio Chaves, Nuno Cardoso, Ana Paula Afonso.

- Semantic Expansion of Geographic Web Queries Based on Natural Language Positioning Expressions.
 - Delboni, T. M., Borges, K. A. V., Laender, A. H. F. Davis Jr., C. A.

Adding Geographic Scopes to Web Resources

Heurísticas

- Utiliza heurísticas para georreferenciar páginas web.
 - Informação textual nas páginas
 - Hyperlinks
 - Ambiente da página
 - Regras de utilização de referências geográficas

Heurísticas

- Informação textual nas páginas
 - Referência geográfica
 - Frequência de entidade geográfica
 - Marcações HTML

Heurísticas

- Hyperlinks
 - Correlação do link com o escopo
 - Agregação de links
 - Anchor text
 - Autoridade das páginas

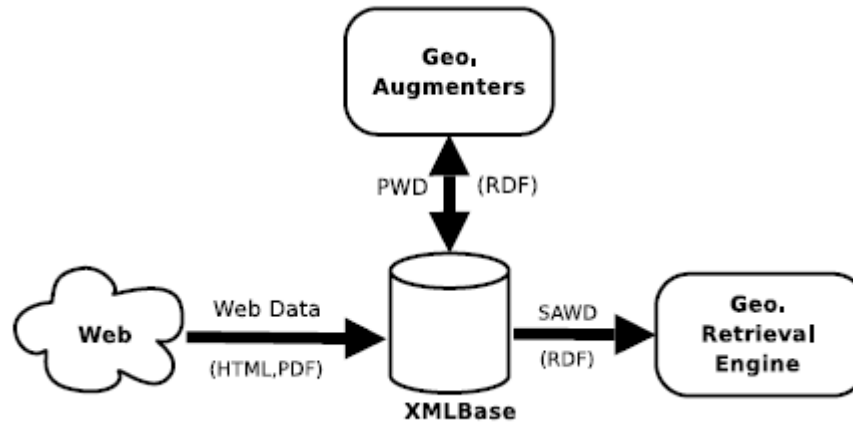
Heurísticas

- Ambiente da página
 - Servidor web da página
 - Site no qual a página está contido

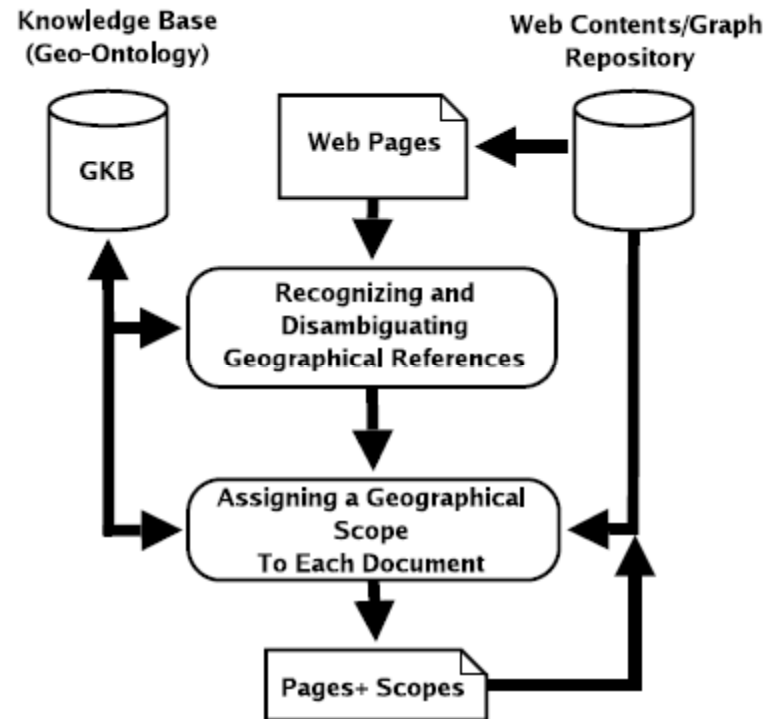
Heurísticas

- Regras de utilização de referências geográficas
 - Escopo único
 - ‘Sub-lugares’
 - Dados demográficos
 - Hierarquia de nomes
 - Referências próprias

Arquitetura Geral



Arquitectura



Geographic Knowledge Base

- Contém dados integrados de diversas fontes externas.
- Gera ontologias de acordo com a Web Ontology Language (OWL).
- Contém diversas informações geográficas:
 - Nomes de lugares
 - Informação de tipo (rua, avenida, cidade)
 - Relacionamentos entre entidades
 - Dados demográficos
 - Códigos postais
 - Coordenadas geográficas
 - Informações referentes a domínios de internet

Semantic Expansion of Geographic Web Queries Based on Natural Language Positioning Expressions

Expressões Posicionais

- As pessoas normalmente se comunicam através de linguagem qualitativa.
 - Frequentemente se orientam através de pontos de referência.
- <lugar, relação espacial, referência>
- Relação espacial usada pelos usuários pode não corresponder ao seu significado topológico real.

Expressões Posicionais

- Experimentos
 - Listagem de 214 pontos de referência de Belo Horizonte.
 - Listagem de 29 relações espaciais da língua portuguesa.
 - Geração da coleção de documentos.
 - Expansão dos pontos de referência.
 - Extração de expressões posicionais.

Relações Espaciais

Spatial Relation	Occurrence	
próximo a (near)	978	22.4%
em frente a (in front of)	736	16.8%
a N km de (N km from)	648	14.8%
ao lado de (beside)	382	8.7%
perto de (close to)	336	7.7%
a N minutos de (N minutes from)	229	5.2%
a N quilômetros de (N kilometers from)	216	4.9%
dentro de (inside)	216	4.9%
atrás de (behind)	96	2.2%
nas proximidades de (in the vicinity of)	90	2.1%
	3927	89.8%

Relações Espaciais

- Equivalência de Relações Espaciais
 - Necessidade de verificar se existe equivalência entre relações espaciais.
 - Medição da distância média entre local e ponto de referência.
 - Agrupamento de relações espaciais equivalentes.

Relações Espaciais

Spatial Relation	Average Distance (m)
dentro de (inside of)	0.0
ao lado de (beside)	142.0
em frente a (in front of)	183.4
atrás de (behind)	326.6
perto de (close to)	527.0
a 1 minuto de (1 minute from)	530.9
próximo a (near to)	634.5
nas proximidades de (in the vicinity of)	782.0
a 1 km de (1 km from)	1,000.0
a 1 quilômetro de (1 kilometer from)	1,000.0

Expansão de Consultas

- Consultas formadas de duas partes: local e ponto de referência.
- Consulta é expandida de forma a levar em consideração todas as relações espaciais do conjunto em questão.
- Consulta expandida é submetida a uma máquina de busca convencional.

Expansão de Consultas

- Exemplo de consulta: hotel; Minas Shopping
- Consulta expandida: “hotel AND ('dentro de Minas Shopping' OR 'ao lado de Minas Shopping' OR 'em frente a Minas Shopping' OR ... OR 'nas proximidades de 'Minas Shopping'')”

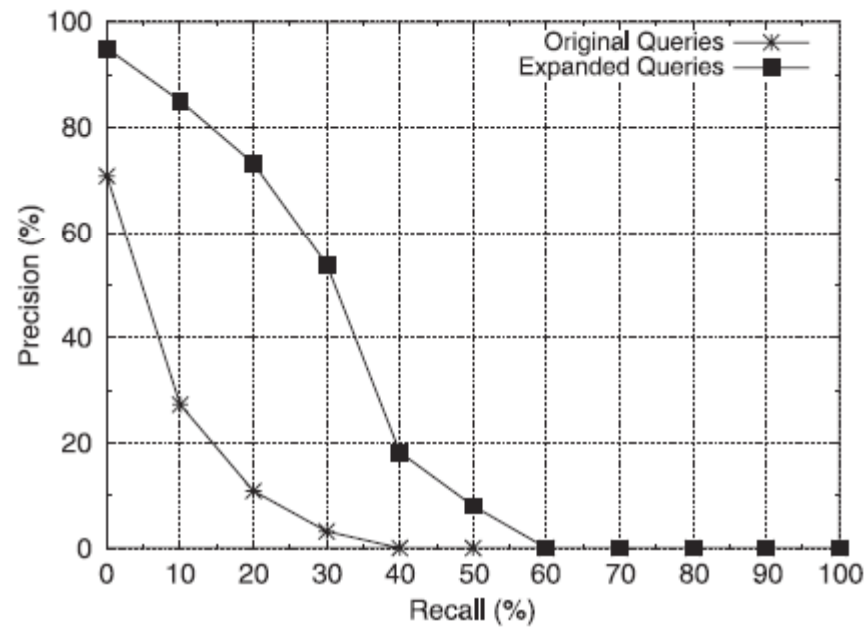
Problemas

- Máquina de busca deve permitir o uso de operadores booleanos nas consultas.
- Máquina de busca utilizada (Google) limita a quantidade de termos de uma consulta a 32 termos.
- Utilização de um conjunto reduzido de relações espaciais.

Experimentos

- Elaboração de 10 consultas.
- Utilização de 15 relações espaciais.
- Formação de um conjunto de 340 páginas por consulta.
- Verificação manual de respostas relevantes.

Resultados



Conclusões

- Área ainda não muito explorada por máquinas de busca.
- Necessidade do serviço por parte do usuário.
- Abordagens estudadas apresentam melhoria significativa nas respostas.

