

UNIVERSIDADE FEDERAL DE CAMPINA GRANDE
CENTRO DE ENGENHARIA ELÉTRICA E INFORMÁTICA
COORDENAÇÃO DE PÓS-GRADUAÇÃO EM INFORMÁTICA

GeoSen_Tags: Um motor de busca geográfico com suporte a Tags

Ricardo Madeira Fernandes

Cláudio de Souza Baptista
(Orientador)

Campina Grande, Paraíba, Brasil
Agosto de 2010

UNIVERSIDADE FEDERAL DE CAMPINA GRANDE
CENTRO DE ENGENHARIA ELÉTRICA E INFORMÁTICA
COORDENAÇÃO DE PÓS-GRADUAÇÃO EM INFORMÁTICA

GeoSen_Tags: Um motor de busca geográfico com suporte a Tags

Ricardo Madeira Fernandes

Dissertação submetida à Coordenação do Curso de Pós-Graduação em Informática da Universidade Federal de Campina Grande - Campus I, como parte dos requisitos necessários para obtenção do grau de Mestre em Ciência da Computação.

Cláudio de Souza Baptista
(Orientador)

Área de Concentração: Ciência da Computação
Linha de Pesquisa: Banco de Dados e Internet

Campina Grande, Paraíba, Brasil
Agosto de 2010

FICHA CATALOGRÁFICA ELABORADA PELA BIBLIOTECA CENTRAL DA UFCG

F363g Fernandes, Ricardo Madeira.

GeoSen_Tags: um motor de busca geográfico com suporte a Tags /
Ricardo Madeira Fernandes. - Campina Grande, 2010.

107f. : il. col.

Dissertação (Mestrado em Ciência da Computação) – Universidade
Federal de Campina Grande, Centro de Engenharia Elétrica e
Informática.

Referências.

Orientador: Prof. Ph.D. Cláudio de Souza Baptista.

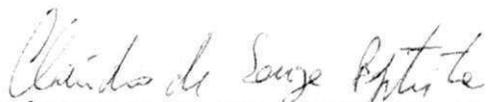
1. Recuperação de Informação. 2. Motor de Busca Geográfico. 3.
Expansão de Consultas. 4. Folksonomia. I. Título.

CDU – 004.738.52(043)

"GEOSEN_TAGS: UM MOTOR DE BUSCA GEOGRÁFICO COM SUPORTE A TAGS"

RICARDO MADEIRA FERNANDES

DISSERTAÇÃO APROVADA EM 31.08.2010


CLÁUDIO DE SOUZA BAPTISTA, Ph.D
Orientador(a)



ULRICH SCHIEL, Dr.
Examinador(a)


ED PORTO BEZERRA, D.Sc
Examinador(a)

CAMPINA GRANDE - PB

Resumo

Os motores de busca atuais possuem capacidade cada vez maior de indexação dos dados disponíveis na Web, tornando ainda mais difícil a tarefa de prover informações úteis para os usuários. As opções avançadas de consultas, como a delimitação de datas, idiomas, regiões geográficas, dentre outras, ajudam a amenizar tal problema. Contudo, mecanismos de busca que empregam técnicas mais elaboradas do que apenas consulta por palavras-chave são necessários. Com o crescimento das redes sociais e anotações de recursos, o usuário torna-se uma entidade fundamental neste processo, de modo que as abordagens cientes de contexto estão presentes nas pesquisas recentes sobre Recuperação da Informação (RI), juntamente com outras já consagradas como a análise sintática e semântica dos termos da consulta. Todavia, as soluções propostas resumem-se, por enquanto, a protótipos acadêmicos focados em apenas uma abordagem. Dessa forma, o presente trabalho tem como objetivo a elaboração de um motor de busca semântico e ciente de contexto, que faz uso em conjunto de algumas técnicas de RI da literatura bem como um novo método para expansão de consultas baseado em anotações. Para a validação do mecanismo proposto, o Sistema de Recuperação de Informação Geográfica, denominado GeoSen, foi estendido com as funcionalidades previstas.

Abstract

Nowadays, search engines have increased their capability of indexing data available on the Web, making the task of providing relevant information to users even more difficult. Advanced search options, such as date, language and geographical region filters, might help to minimize this issue. However, there is still a need to develop search engines that employ more sophisticated techniques than just keyword-based. The growth of social networks and resource tagging turn the user as a key entity in this process, so that context-aware approaches are present in the recent research on Information Retrieval (IR), along with the well established ones, as the syntactic and semantic analysis of query terms. Nonetheless, the proposed solutions are, mainly, academic prototypes focused on only one approach. Hence, this work aims to develop a semantic and context-aware search engine, which combines techniques of IR literature with a new tag-based query expansion method. In order to validate the proposed ideas, the Geographic Information Retrieval System, called GeoSen, was extended with the provided features.

Agradecimentos

À minha mãe querida, pelo incentivo incondicional, pelo amor e estímulo aos estudos desde sempre.

Ao meu pai e irmão, Gilberto e Henrique, pelo apoio e momentos de descontração, alegria, e tristezas, assistindo os jogos do mengão.

À minha avó, *Sabá*, que mesmo estando distante tem torcido por mim.

À minha amada esposa, Luciana, pelo carinho, companheirismo, amor, dedicação e paciência, principalmente na fase final dessa etapa tão importante da minha vida.

Aos meus sogro e sogra, Romeu e Diana, e aos meus cunhados Daniel, Juliana e Mariana, por me acolher em seu lar com tanta atenção e carinho durante todos esses anos.

Ao meu orientador e amigo, professor Dr. Cláudio Baptista, pela orientação, empenho, profissionalismo e conselhos, durante todos esses anos de convivência, que começou ainda na graduação.

A todos os companheiros de trabalho do LSI, pelas experiências compartilhadas, em especial a Hugo, Yuri, Fábio, Damião, Elvis e Tiago.

Às funcionárias da COPIN, Aninha e Vera, pela sua atenção, simpatia e prestatividade.

Aos professores Ulrich Schiel e Ed Porto, por aceitarem o convite para participarem da banca do meu mestrado.

À CAPES pelo apoio financeiro no período em que este foi necessário.

Conteúdo

Conteúdo.....	7
Lista de Figuras	9
Lista de Tabelas	11
Lista de Códigos	12
Lista de Equações	13
Capítulo 1	14
Introdução.....	14
1.1. Objetivos	16
1.1.1. Objetivo Geral	16
1.1.2. Objetivos Específicos	16
1.2. Relevância	17
1.3. Estrutura da Dissertação	17
Capítulo 2	19
Fundamentação Teórica.....	19
2.1. Recuperação da Informação	20
2.2. Recuperação de Informação Geográfica	24
2.3. Expansão de consulta	28
2.4. Ontologia	30
2.5. Folksonomia	33
2.6. Gazetteers	35
2.7. Ciência de Contexto	38
2.8. Considerações Finais	40
Capítulo 3	41
Trabalhos Relacionados.....	41
3.1. Expansão de consultas.....	42
3.2. Folksonomia e Ontologias.....	45
3.1. GeoSen	47
3.2. Considerações Finais.....	50
Capítulo 4	52
GeoSen_Tags.....	52
4.1. Visão Geral do Sistema	52

4.2. Arquitetura.....	53
4.3. Expansão de Consultas Baseada em Anotações.....	58
4.4. Construção da Folksonomia.....	62
4.5. Sugestão de Termos.....	64
4.6. Gazetteers.....	66
4.7. Considerações Finais.....	70
Capítulo 5.....	71
Estudo de Caso.....	71
5.1. Ambiente do Estudo de Caso.....	71
5.2. Estudo de Caso 1: Detecção de Lugares através de Gazetteers.....	72
5.3. Estudo de Caso 2: Sugestão de termos relacionados.....	78
5.4. Estudo de Caso 3: Expansão de Consultas e Folksonomia.....	82
5.5. Considerações Finais.....	87
Capítulo 6.....	88
Conclusão.....	88
6.1. Contribuições.....	89
6.2. Trabalhos Futuros.....	90
Referências Bibliográficas.....	91
Apêndice.....	100
Ontologia directoryOntology.owl.....	100

Lista de Figuras

Figura 2.1 – Componentes de um SRI (extraída de [41]).....	21
Figura 2.2 –Principais Componentes da RIG.	27
Figura 2.3 – Distinção de diferentes vocabulários de acordo com sua expressividade (extraída de [31]).	30
Figura 2.4 – Sistema tradicional VS. Sistema Ciente de Contexto – adaptada de [91].....	40
Figura 3.1 – Exemplo de matriz de correlação tridimensional (extraída de [7]).....	44
Figura 3.2 – Os principais passos do algoritmo empregado no FLOR (extraída de [33]). .	47
Figura 3.3 – Arquitetura do GeoSen (extraída de [71]).....	49
Figura 3.4 – Interface de consulta do GeoSen (extraída de [71]).....	50
Figura 4.1 – Visão geral do GeoSen_Tags.	53
Figura 4.2 – Arquitetura em camadas, segundo o padrão MVC.	54
Figura 4.3 – Arquitetura detalhada do GeoSen_Tags.....	57
Figura 4.4 – Etapas do processo de expansão de consultas.....	58
Figura 4.5 – Análise sintática dos termos, via <i>letter n-grams</i>	59
Figura 4.6 – Análise semântica dos termos.	60
Figura 4.7 – Filtragem espacial dos documentos.	61
Figura 4.8 – Folksonomia de documentos.....	62
Figura 4.9 – Modelo conceitual daFolksonomia.	62
Figura 4.10 – Processo de sugestão de termos.	64
Figura 4.11 – Exemplo de arquivo OWL.	65
Figura 4.12 – Filtragem dos termos sugeridos.	66
Figura 4.13 – Comparativo GeoSen X GeoSen_Tags.....	67
Figura 4.14 – Processo detecção de lugares. (a) Consulta do usuário. (b) Consulta ao GeoNames e Geocoding. (c) Consulta a base de dados geográfica. (d) Lista de lugares.....	68
Figura 5.1 – Interface de consultas do GeoSen_Tags.	73
Figura 5.2 – Tela de desambiguação da consulta espacial.	74
Figura 5.3 – Tela de desambiguação da consulta espacial apresentada pelo GeoSen para o lugar “ <i>parque do povo</i> ”.	75
Figura 5.4 – Consulta por “ <i>turismo</i> ” no “ <i>maracana</i> ”, na tela de consultas do GeoSen_Tags.	76

Figura 5.5 – Tela de desambiguação para a consulta pelo lugar “maracana” no GeoSen_Tags.....	77
Figura 5.6 – Tela de resultados para a consulta por “turismo” no lugar “maracana”	78
Figura 5.7 – Consulta por “verão” no lugar “joao pessoa”	79
Figura 5.8 – Trecho da ontologia que representa as informações da classe Verão.	79
Figura 5.9 – Tela de resultados para a consulta por “verão” no lugar “joao pessoa”	80
Figura 5.10 – Tela de resultados para a consulta “estações do ano”	81
Figura 5.11 – Trecho da ontologia que representa as informações relacionadas a classe Estações do ano.	81
Figura 5.12 – Esquema utilizado no estudo de caso de expansão de consultas e folksonomia.	82
Figura 5.13 – Resultado da consulta 1 executada pelo usuário A.	84
Figura 5.14 – Mapeamento da folksonomia gerada pela consulta 1 do usuário A.....	85
Figura 5.15 – Resultado da consulta 3 do usuário B.	86
Figura 5.16 – Resultado da consulta 4 do usuário C.	87

Lista de Tabelas

Tabela 3.1 – Características importantes de um mecanismo de busca semântico e ciente de contexto.	51
Tabela 3.2 – Associação entre projetos e características importantes de um mecanismo de busca.	51
Tabela 5.1 – Consulta realizada pelo usuário A e sua expansão.	83
Tabela 5.2 – Consulta realizada pelo usuário B e sua expansão.	85
Tabela 5.3 – Consulta realizada pelo usuário B e sua expansão.	86
Tabela 5.4 – Consulta realizada pelo usuário C.	87

Lista de Códigos

Código 4.1 – Algoritmo de busca de termos relacionados.....	60
Código 4.2 – Algoritmo de filtragem geográfica dos documentos.....	61
Código 4.3 – Algoritmo de atualização da Folksonomia.	63
Código 4.4 – Trecho de algoritmo JAVA de busca de lugares, através do GeoNames.	69
Código 4.5 – Trecho de algoritmo JAVASCRIPT de busca de lugares, através do Geocoding.....	70
Código 5.1 – Comando SQL para busca de lugares na base de dados geográfica.	74

Lista de Equações

Equação 2.1 – Cálculo da precisão.....	22
Equação 2.2 – Cálculo da cobertura.....	22
Equação 3.1 – Similaridade entre anotações (extraída [20]).....	43
Equação 3.2 – Similaridade entre recursos (extraída [20]).	43

Capítulo 1

Introdução

O surgimento da Web semântica [1] contribuiu para impulsionar a criação de serviços baseados na personalização e contexto dos usuários. Atualmente, na área de Recuperação da Informação (RI), estes aspectos fazem parte de diversos temas de pesquisa, já que auxiliam na eficácia dos mecanismos de busca, que precisam evoluir para continuar atendendo as necessidades dos usuários, que esperam obter recursos relevantes sempre que realizam consultas.

Uma solução bastante difundida é a de expansão e/ou substituição de consultas, que analisa a sintaxe dos termos, podendo corrigir erros de grafia e efetuar derivações de uma palavra, como o emprego do plural, o uso do seu radical, dentre outros. Além disso, aspectos semânticos também podem ser considerados, por exemplo, através de ontologias de domínio [2]. Neste caso, as palavras informadas pelo usuário são avaliadas de modo que termos relacionados são descobertos e adicionados à consulta original, para que resultados mais precisos sejam obtidos.

Com relação ao usuário, pode-se afirmar que este possui um papel fundamental no cenário de mecanismos de busca e, sendo assim, é justificado o uso crescente de abordagens que consideram parâmetros como o seu contexto e suas preferências. Isto já pode ser visto cada vez mais através do rápido crescimento das redes sociais e anotações de recursos, que tornam a busca por informações na Web uma experiência mais enriquecedora. Por exemplo, os usuários podem atribuir rótulos a suas URLs preferidas, facilitando a organização e disseminação das mesmas.

Diante deste quadro, surgiu um novo termo, denominado Folksonomia, que, de modo geral, significa o ato de atribuir rótulos a objetos armazenados em repositórios

públicos de dados textuais e multimídias, tais como o FLICKR¹ (sistema de compartilhamento de fotos), o DEL.ICIO.US² (sistema de compartilhamento de *bookmarks*) e o YOUTUBE³ (sistema de compartilhamento de vídeos). Dessa forma, as máquinas de busca que antes já avaliavam diversos tipos de informações provenientes dos criadores de sites, como o conteúdo e a estrutura de links, e dos usuários, através de sua navegação, também estão, recentemente, levando em consideração as anotações atribuídas aos recursos disponíveis na Web. Tal informação pode ser de grande valia para melhorar a precisão das máquinas de busca, já que os usuários foram bem treinados utilizando a Internet ao longo dos anos.

Todavia, os motores de busca atuais possuem uma capacidade de indexação tão grande que se torna ainda mais difícil prover informações realmente úteis para os usuários. Uma das soluções mais adotadas é fornecer opções avançadas de consultas como filtros para restringir os resultados e, dentre estes, uma escolha promissora é a de limitar a região geográfica do que se espera obter como resposta.

Com isso, surgiram os motores de busca geográficos, que utilizam a informação de localização dos recursos consultados, uma vez que ela é intrínseca a diversos conteúdos presentes na Web. Por exemplo, se um jovem deseja pesquisar sobre lugares que fornecem aulas de piano, ele provavelmente estará interessado em encontrar informações sobre escolas de música localizadas próximas de onde ele mora, de tal forma que possa futuramente se matricular em alguma.

Enfim, o uso da localização geográfica está presente em diversas pesquisas na área de RI, visto que através dele é possível expressar consultas contendo restrições de lugares de forma mais intuitiva e eficaz do que em sistemas de busca tradicionais.

Tendo em vista os aspectos levantados, o propósito deste trabalho compreende a elaboração de um motor de semântico e ciente de contexto, enriquecido geograficamente.

¹ <http://www.flickr.com/>

² <http://delicious.com/>

³ <http://www.youtube.com/>

1.1. Objetivos

Nesta seção são apresentados os objetivos deste trabalho, divididos em geral e específicos.

1.1.1. Objetivo Geral

O principal objetivo deste trabalho é o desenvolvimento de um motor de busca semântico e ciente de contexto, aplicado a um sistema de recuperação de informação geográfica para Web já existente, denominado GeoSen (Geographic Search Engine).

1.1.2. Objetivos Específicos

Os objetivos específicos do trabalho proposto são:

- extensão do sistema de recuperação de informação geográfica, GeoSen, para desenvolver o mecanismo de busca semântico e ciente de contexto em um motor de busca já existente;
- uso de ontologias para auxiliar na representação semântica da informação presente na Web, bem como nas técnicas de recuperação da informação.
- uso de anotações para prover a construção de uma folksonomia dos documentos indexados pelo motor de busca GeoSen, através de feedback indireto do usuário;
- uso de Gazetteers para melhorar o processo de identificação de lugares geográficos;
- uso do contexto geográfico do usuário para expandir termos da consulta; e
- realização de estudos de caso da solução proposta.

1.2. Relevância

A personalização, as abordagens cientes de contexto e os mecanismos de substituição e/ou expansão dos termos de uma consulta textual ou geográfica estão presentes nas pesquisas recentes sobre RI, com o objetivo principal de fornecer procedimentos eficazes para a busca dos recursos disponíveis na Web.

Entretanto, de modo geral, o funcionamento dos motores de busca tradicionais disponíveis hoje consiste em recuperar recursos que contenham algumas palavras-chave especificadas pelo usuário. As pesquisas sobre a utilização de abordagens sintáticas, semânticas ou centradas no usuário são recentes e os resultados, por enquanto, resumem-se a protótipos acadêmicos focados apenas em uma abordagem.

Diante disto, o trabalho proposto oferece o uso em conjunto de algumas técnicas de RI da literatura bem como um novo método para expansão de consultas baseado em anotações. Dentre as principais contribuições estão:

- Elaboração de um protótipo de motor de busca para validar as idéias propostas.
- Construção de uma folksonomia de recursos da Web através do feedback indireto do usuário, de modo que os termos da folksonomia sejam utilizados na expansão de consultas.
- Uso de ontologias para inferir a semântica dos termos de uma consulta, afim de que termos relacionados possam ser sugeridos para o usuário.
- Novo algoritmo de expansão de consultas baseado em anotações e no contexto geográfico do usuário. A idéia principal é analisar a co-relação entre as anotações através da folksonomia construída.
- Aprimoramento do algoritmo de detecção de lugares do GeoSen com o uso de gazetteers para detectar nomes de parques, prédios, pontos turísticos, dentre outros.

1.3. Estrutura da Dissertação

O restante desta dissertação está organizado da seguinte forma: o Capítulo 2 discute sobre os fundamentos da recuperação de informação e dos motores de busca para

Web, bem como sobre os fundamentos e estado da arte no campo da recuperação de informação geográfica. O Capítulo 3 apresenta as principais contribuições na área e as pesquisas mais importantes que exploram o tema. No Capítulo 4, descrevem-se os métodos e técnicas elaborados neste trabalho; apresenta-se ainda o protótipo desenvolvido para validação das idéias propostas. No Capítulo 5, são mostrados os estudos de caso realizados para avaliação do protótipo desenvolvido. Por fim, discutem-se no Capítulo 6 as limitações observadas e as possibilidades de trabalhos futuros.

Capítulo 2

Fundamentação Teórica

Na Web, existe uma vasta quantidade de informações, porém, os usuários, utilizando-se de ferramentas de buscas, muitas vezes não conseguem obter o que eles realmente almejam. O principal motivo para que isso aconteça é que os documentos existentes raramente estão associados a alguma semântica, tornando sua recuperação de forma automática uma tarefa mais difícil.

Com o avanço das redes sociais e a anotação de recursos, esta tarefa tem se tornado mais fácil, uma vez que a Web está cada vez mais deixando de ser uma infraestrutura de somente leitura, com participantes passivos, para uma plataforma de leitura e escrita, com usuários ativos.

Diante disto, neste capítulo, apresenta-se um estudo sobre conceitos importantes para a compreensão do funcionamento de mecanismos de busca para a Web.

O restante do capítulo está organizado como segue: na seção 2.1, são apresentados os fundamentos de RI; na seção 2.2, é relatado o estudo feito sobre os Sistemas de Recuperação de Informação Geográfica; em seguida, na seção 2.3, são mostrados conceitos fundamentais sobre os métodos de expansão de consultas; já na seção 2.4, é tratada a representação de conhecimento através de ontologias; na seção 2.5, é mostrado um estudo feito sobre uma abordagem recente para a atribuição de significados às informações presentes na Web; na seção 2.6, é explicado o que são gazetteers; na seção 2.7, é são apresentados conceitos fundamentais da ciência de contexto; por fim, na seção 2.8, são feitas considerações finais sobre o capítulo.

2.1. Recuperação da Informação

Com o crescimento do volume de informações digitais, ao longo dos anos, foram desenvolvidas técnicas de recuperação da informação para atender às necessidades dos usuários. A estrutura mais importante para auxiliar o processo de recuperação é denominada índice, que é uma coleção de termos que indica o local onde a informação desejada pode ser localizada [40]. Estes termos devem ser organizados de forma a facilitar sua busca.

Atualmente, o volume de dados da Web cresce de forma exponencial. Neste contexto, são estudadas técnicas de armazenamento, processos de busca, recuperação e apresentação de forma amigável das informações.

A crescente complexidade dos objetos armazenados e o grande volume de dados exigem processos de recuperação cada vez mais sofisticados. Diante deste quadro, a Recuperação da Informação apresenta, a cada dia, novos desafios e se configura como uma área de grande significância.

Ela é uma subárea da Ciência da Computação que estuda o armazenamento e recuperação automática de documentos, que são objetos de dados, como textos, vídeos, áudios e imagens. Um Sistema de Recuperação da Informação (SRI) pode ser estruturado conforme a Figura 2.1.

Os componentes do sistema incluem documentos, necessidades do usuário, geração da consulta formulada, e finalmente o processo de recuperação que, a partir das estruturas de dados e da consulta formulada, recupera uma lista de documentos considerados relevantes.

O processo de indexação envolve a criação de estruturas de dados associados à parte textual dos documentos, por exemplo, as estruturas de arranjos de sufixos (*PAT arrays*) e arquivos invertidos, discutidas em [40]. Estas estruturas podem conter dados sobre características dos termos na coleção de documentos, tais como a frequência de cada termo em um documento.

O processo de especificação da consulta geralmente é uma tarefa difícil, devido à distância semântica entre a real necessidade do usuário e o que ele expressa na consulta formulada. Essa distância é gerada pelo limitado conhecimento do usuário sobre o universo de pesquisa e pelo formalismo da linguagem de consulta.

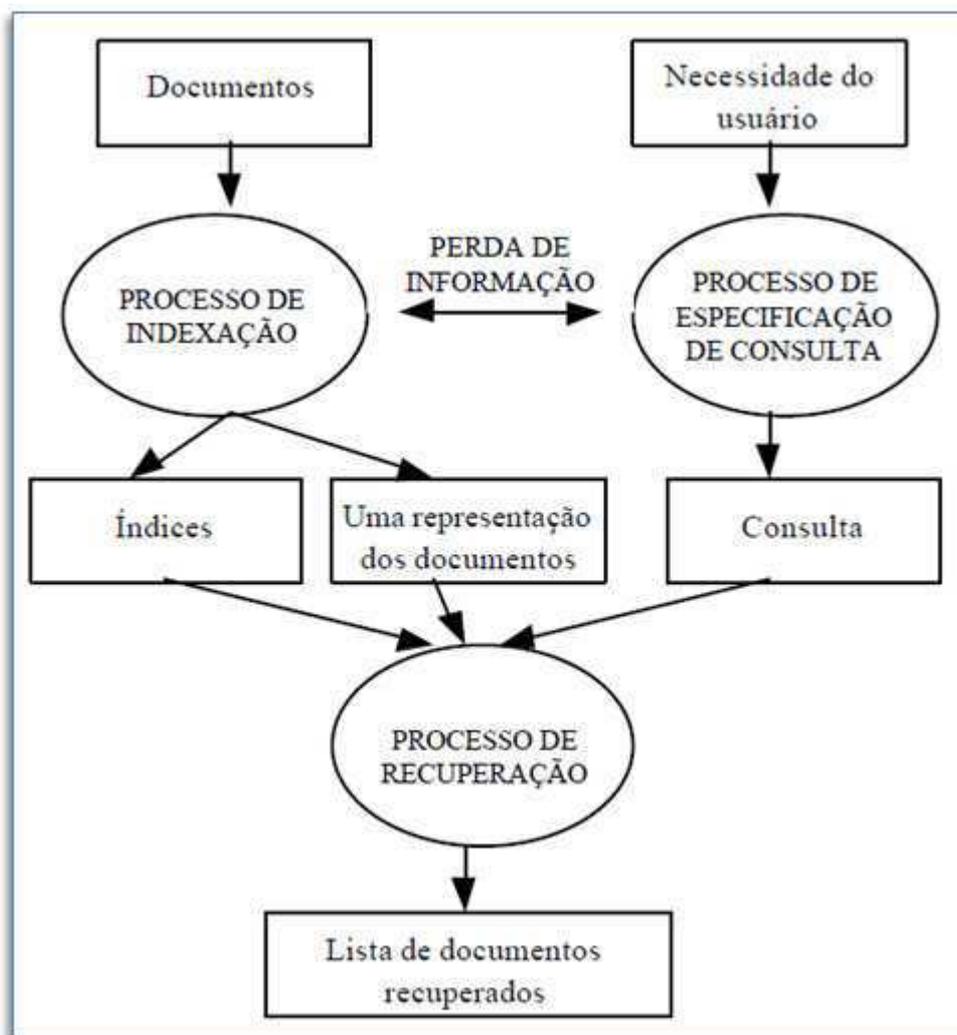


Figura 2.1 – Componentes de um SRI (extraída de [41]).

O processo de recuperação consiste na geração de uma lista de documentos que satisfazem à consulta do usuário. Os índices construídos para a coleção de documentos são usados para acelerar esta tarefa. Além disso, o resultado da consulta é classificado em ordem decrescente de um grau de similaridade entre os documentos e a consulta.

Para avaliar um SRI, existem dois conceitos fundamentais: *precisão* e *cobertura* [11]. Precisão é a fração dos documentos retornados que são relevantes, e cobertura é a fração dos documentos relevantes observada dentre os documentos retornados. Seja N o conjunto de resposta ideal, $|N|$ o número de documentos deste conjunto e R o vetor resultado recuperado pelo SRI. Então:

$$\textit{Precisão} = \frac{|N \cap R|}{|R|}$$

Equação 2.1 – Cálculo da precisão.

$$\textit{Cobertura} = \frac{|N \cap R|}{|N|}$$

Equação 2.2 – Cálculo da cobertura.

Os modelos clássicos, utilizados no processo de recuperação da informação (booleano, vetorial e probabilístico) apresentam estratégias de busca de documentos relevantes para uma consulta. Estes modelos consideram que cada documento é descrito por um conjunto de palavras chaves, chamadas termos de indexação. Associa-se a cada termo de indexação t_i em um documento d_j um peso $w_{ij} \geq 0$, que quantifica a correlação entre os termos e o documento.

Além destes, outros modelos mais avançados de recuperação da informação foram propostos, tais como os baseados em: bases de conhecimento [42], lógica *fuzzy* [43] e redes neurais [44].

No modelo booleano, os documentos recuperados são aqueles que contêm os termos que satisfazem a expressão lógica da consulta. Uma consulta é considerada como uma expressão booleana convencional formada com os conectivos lógicos *AND*, *OR* e *NOT*.

Os principais problemas do modelo booleano são a ausência de ordem na resposta, e a possibilidade destas serem nulas ou muito grandes. As vantagens desse modelo são a facilidade de implementação e a expressividade completa das expressões.

O modelo de espaço vetorial, ou simplesmente modelo vetorial, representa documentos e consultas como vetores de termos. Termos são ocorrências únicas nos documentos. Os documentos devolvidos como resultados para uma consulta são representados similarmente, ou seja, o vetor resultado para uma consulta é montado através de um cálculo de similaridade.

Aos termos das consultas e documentos são atribuídos pesos que especificam o tamanho e a direção de seu vetor de representação. Ao ângulo formado por estes vetores dá-se o nome de q . O $\cos q$ determina a proximidade da ocorrência. O cálculo da similaridade é baseado neste ângulo entre os vetores que representam o documento e a consulta [45].

As principais vantagens do modelo vetorial são a sua simplicidade, a facilidade que ele provê de se computar similaridades com eficiência e o fato de que o modelo se comportar bem com coleções genéricas.

O modelo probabilístico descreve documentos considerando pesos binários que representam a presença ou ausência de termos. O vetor resultado gerado pelo modelo tem como base o cálculo da probabilidade de que um documento seja relevante para uma consulta. A principal ferramenta matemática do modelo probabilístico é o teorema de Bayes [46].

O modelo probabilístico é baseado no princípio probabilístico de ordenação (*Probability Ranking Principle*), que estabelece que este modelo pode ser usado de forma ótima. Este princípio é baseado na hipótese de que a relevância de um documento para uma determinada consulta é independente de outros documentos.

O modelo probabilístico tem como vantagem, além do bom desempenho prático, o princípio probabilístico de ordenação, que uma vez garantido, resulta em um comportamento ótimo do método. Entretanto, a desvantagem é que este comportamento depende da precisão das estimativas de probabilidade. Além disso, o método não explora a frequência do termo no documento e ignora o problema de filtragem de informação.

Existem várias dificuldades para que o usuário transforme suas necessidades em uma consulta devidamente formulada. Geralmente é a má formulação da consulta que prejudica o desempenho dos sistemas. Um método de abordar este problema é considerar uma forma iterativa de construção da consulta, onde o usuário formula uma consulta inicial, examina o resultado diante de suas necessidades e, se necessário, melhora a formulação da consulta.

Uma das estratégias mais populares para reformular consultas é chamada de feedback de relevância [47], cuja idéia principal é como se segue. Após a montagem do vetor resultado baseado na consulta inicial, o usuário seleciona documentos de sua preferência. O sistema então seleciona termos pertencentes aos documentos selecionados e utiliza estes termos para reformular a consulta. Este processo de reformulação pode prosseguir com mais de uma interação.

A principal vantagem do método é que, após a primeira formulação, o usuário interage com o sistema abstraindo-se do processo de formulação, simplesmente identificando documentos como relevantes ou não. Outra vantagem é que o método provê um processo controlado de enfatizar alguns termos e diminuir a importância de outros.

Em recuperação da informação geralmente o usuário tenta identificar qual parte do documento retornado atende sua necessidade. Uma forma de apresentar tal informação é dividir o documento em porções menores denominadas *snippets*. Em alguns casos, aplicar os algoritmos de recuperação a *snippets*, e não a documentos completos, resulta em melhor desempenho do sistema [48].

A divisão dos documentos em *snippets* pode ser feita de três formas. A primeira usa uma abordagem hierárquica, de modo que são consideradas sentenças, parágrafos e seções. A segunda procura por características semânticas do conteúdo de partes do documento, neste sentido os *snippets* agrupam porções do texto que tratam de um determinado assunto. A terceira considera seqüência contígua de palavras. Esse tipo de *snippets* é chamado de janela e o número de palavras na seqüência define o tamanho da janela.

Um problema fundamental em recuperação da informação é que os autores nem sempre usam as mesmas palavras que os usuários para descrever o mesmo conceito [4]. A importância deste problema tende a diminuir com o aumento do tamanho da consulta. Entretanto, em muitas aplicações, as consultas podem possuir uma pequena quantidade de termos. Um caso extremo ocorre no contexto da *Web*, onde as consultas possuem tipicamente duas palavras.

Para expandir uma consulta, é preciso buscar palavras com significados semelhantes aos termos da consulta e acrescentar tais palavras à consulta original, com o objetivo de melhorar o contexto da mesma.

Enfim, o estudo da área de RI é de fundamental importância para a comunidade acadêmica, visto que, com a explosão do número de documentos e usuários na *Web*, modelos para recuperação precisa de informações passam a ser necessários.

2.2. Recuperação de Informação Geográfica

Desde o surgimento dos primeiros Sistemas de Informação Geográfica, na década de 1960 no Canadá, até os dias atuais, a Recuperação de Informações Geográficas (RIG) tem evoluído e ganhado cada vez mais importância.

Com o surgimento da Web e a enorme quantidade de informações não estruturadas ou semi-estruturadas [73] disponíveis atualmente, surge a necessidade do uso de tecnologias semânticas para a obtenção de informações relevantes, visando evitar assim problemas como a sobrecarga de informações [74].

Os algoritmos de recuperação de informação desenvolvidos tradicionalmente para lidar com coleções pequenas de documentos [75] tiveram, portanto, que adaptar-se ao ambiente dinâmico e geograficamente distribuído da Internet, assim como com a sintaxe específica das informações presentes na Web.

Segundo Larson [76], a área de Recuperação de Informação Geográfica (RIG) pode ser vista como um ramo da área de Recuperação de Informação tradicional, incluindo todas suas áreas de pesquisa, mas enfatizando a recuperação e indexação de informações geográficas e espaciais. O objetivo é lidar com todos os problemas da recuperação de informações que contenham algum tipo de consciência espacial (*spatial awareness*), ou seja, que incluam referências geográficas as quais são essenciais para o significado da consulta, por exemplo: “encontre-me hotéis interessantes e baratos próximos a Madrid” [77].

A recuperação de informações na Web dá-se por meio dos mecanismos de busca, que consultam os sites e, através da análise do seu conteúdo, desenvolvem métodos próprios de classificação de propósito geral como o PageRank [78] ou focados em conteúdos específicos, como, por exemplo, o CiteSeer [79].

No entanto, os sistemas de busca tradicionais não apresentam suporte para informações contextuais, não possibilitando a verificação de informações semânticas, como, por exemplo, a localidade referenciada pelos textos, impedindo assim a verificação de informações dentro ou próximo a determinadas regiões geográficas [80].

Uma larga proporção da informação presente na Web pode ser incluída dentro do espaço geográfico e, como consequência, muitos usuários desejariam especificar nomes geográficos de lugares como parte das consultas [81]. Sites que contêm, por exemplo, informações sobre restaurantes, teatros e cinemas são mais interessantes para usuários vizinhos dessas localidades [80]. Informações variadas como as jornalísticas (tempo, condições de tráfego) também são mais úteis se diretamente relacionadas com a região em que ocorrem.

Consultas que incluem pelo menos um termo relacionado à geografia, como nomes de lugar e feições naturais (ex., “praia”, “serra”), são hoje um subconjunto significativo das consultas submetidas aos mecanismos de busca [82].

O caráter semi-estruturado da Web dificulta, no entanto, o acesso a informações geográficas. Jones et al. [83] relacionam as principais dificuldades no uso da web como fonte de informações geográficas:

- o contexto Geográfico é incluído junto das descrições via linguagem natural;
- nomes de lugares são ambíguos e confundidos com nomes de organizações, pessoas, construções e ruas;
- dependência de presença e relação com os termos do texto;
- interpretação das relações espaciais (“próximo”, “ao oeste”, etc); e
- construção de *ranking* específico para definição da relevância geográfica.

Visando solucionar esses problemas e trazer à Web todas as vantagens relacionadas à descrição semântica e geográfica das informações, técnicas têm sido desenvolvidas tanto em ambiente acadêmico como comercial, visando acessar recursos com base em seu contexto geográfico [84].

De acordo com Santos et al. [85], a RIG pressupõe o seguinte:

- possibilidade de associar à coleção de documentos, informações geográficas; e
- existência ou a possibilidade de criação de repositórios semânticos (ontologias geográficas) que permitam a inferência geográfica.

Com isso, segundo Larson [76], a RIG está relacionada com a recuperação de informações determinística (por exemplo, para encontrar todos os documentos relacionados a determinada coordenada geográfica) e com a recuperação de informações probabilística (por exemplo, para encontrar todas as cidades próximas a um determinado rio).

O processo de georreferenciamento de textos, também definido como *geotagging*, possui de forma geral os componentes apresentados conforme a Figura 2.2.

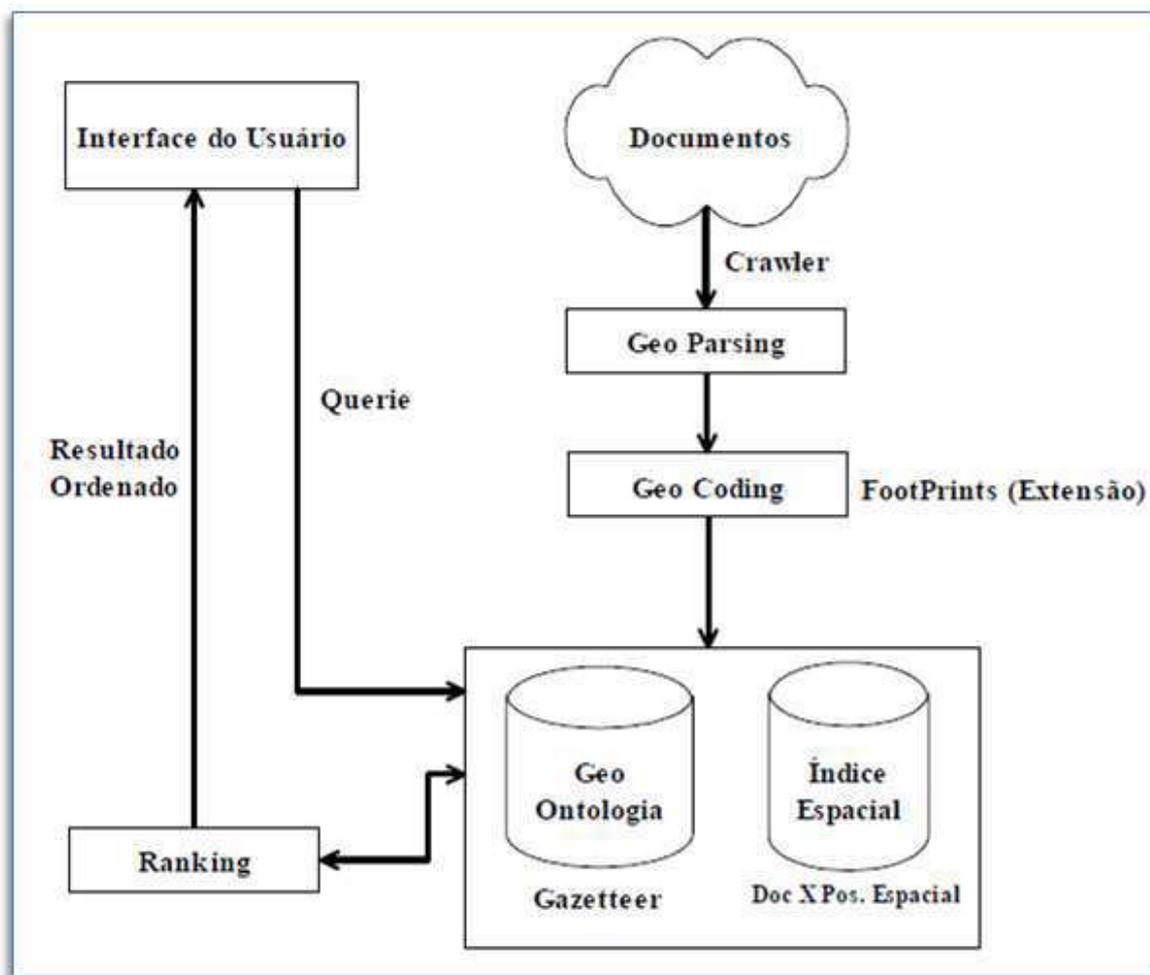


Figura 2.2 –Principais Componentes da RIG.

Diferente dos modelos de recuperação de informação tradicionais, como o modelo Vetorial [86], que representa os documentos como índices de termos (posição x frequência) relacionados ao conteúdo parcial ou integral do texto, na RIG, os termos extraídos como índices devem ser relacionados a descrições espaciais, ou seja, a entidades geometricamente definidas e localizadas no espaço [76].

Os processos de reconhecimento do contexto geográfico de textos e a definição de coordenadas espaciais (latitude, longitude, etc) são definidos respectivamente como *geocoding* e *geoparsing* [87]. Os índices dos arquivos podem também ser estruturados conforme uma lista invertida [88], com a diferença que a lista agora apresenta as relações espaciais (documentos X contexto geográfico) em vez de relacionar apenas as palavras com os documentos em que elas ocorrem. O processo de *ranking*, por sua vez, tem o objetivo de ordenar os resultados de acordo com o grau de associação com a localidade espacial exposto na consulta. O componente mais importante é a ontologia geográfica ou

gazetteer que é organizada em uma estrutura hierárquica [89], codificando termos geográficos e relacionamentos semânticos [90] e tornando-se uma das melhores formas de representação do mundo geográfico [89].

2.3. Expansão de consulta

A expansão de consulta é uma técnica na qual se busca aumentar a quantidade de termos que devem ser buscados, sendo que estes devem possuir certo grau de equivalência entre si [62] para aumentar a probabilidade de se encontrarem documentos relevantes.

Atualmente, vários métodos de expansão de consulta são propostos, sendo possível classificá-los em dois principais grupos: métodos iterativos e métodos automáticos [62].

A expansão iterativa de consulta, também conhecida como expansão semi-automática ou feedback de relevância (*relevance feedback*), se caracteriza por ser uma técnica na qual é necessária a iteração entre o usuário e o sistema para que possa ser efetivada a expansão da consulta [63].

As abordagens que fazem uso desta técnica possuem basicamente o seguinte ciclo de vida [64]:

- O usuário insere os atributos-chave da sua consulta e esta é efetuada sem nenhum tipo de expansão;
- Dentre os documentos recuperados na busca, o usuário indica quais contêm informação que ele considera relevante;
- É efetuada então a expansão da consulta, com a adição dos termos mais relevantes dos documentos escolhidos possuindo um alto peso;
- A consulta expandida é efetuada e os seus resultados são apresentados ao usuário.

O processo acima pode ser realizado quantas vezes o usuário desejar, sendo que a cada iteração espera-se que haja uma melhoria do resultado da busca.

A expansão automática de consulta é uma técnica que, diferentemente da expansão iterativa de consulta, não necessita da intervenção do usuário para que a expansão possa ser efetuada, o que a torna uma técnica mais interessante, uma vez que o processo é transparente para o usuário. Esta técnica é também conhecida como pseudo feedback de relevância [72].

Segundo [65], os métodos pertencentes à classe da Expansão Automática de Consulta podem ser subdivididos em dois grupos: os globais e os locais.

Os métodos globais se baseiam nas chamadas hipóteses de associação, que pregam que palavras relacionadas em uma base de dados tendem a aparecer nos documentos desta base.

Um dos primeiros métodos globais da expansão automática de consultas é o do agrupamento de termos (*term clustering*), no qual são formados grupos de documentos com base no aparecimento de termos em documentos da base de dados. A princípio, esta estratégia se mostrou bastante interessante, mas estudos apontaram que esta é pouco eficiente em bases de dados onde há a ocorrência de polissemia nos termos que compõe os documentos da base de dados, já que, com a polissemia, documentos de assuntos distintos, porém com termos em comum, eram agrupados como sendo relacionados.

Os mais recentes métodos globais buscam resolver o problema da polissemia expandindo a consulta como um todo. Uma mostra deste tipo de abordagem pode ser observada no trabalho [63], no qual foi utilizado um thesaurus para que a expansão da consulta fosse efetuada. Nesta abordagem os termos escolhidos para a expansão eram aqueles com maior ocorrência entre os termos da consulta.

Os métodos locais utilizam o resultado da busca com a chave de consulta inserida pelo usuário para fazerem a expansão desta chave.

O método local mais simples entre os propostos é o de feedback local (*local feedback*), proposto por [66], no qual se admite que os primeiros documentos retornados pela consulta com a chave de busca fornecida pelo usuário são relevantes para aquela busca, e utiliza estes documentos para expandir a consulta por um processo similar ao utilizado na expansão iterativa de consultas.

Apesar deste método se mostrar bastante eficiente na expansão de consulta, estudos apontam que este não é um método muito robusto, uma vez que ele possui uma forte dependência de que os primeiros documentos sejam relevantes, fato que este método não busca assegurar.

Em suma, o mecanismo de expansão de consultas é um elemento essencial no processo de recuperação da informação. Com isso, as pesquisas nessa área permanecem na atualidade com abordagens mais voltadas para o usuário, que possui um papel central nesse contexto.

2.4. Ontologia

A Web tornou-se nos últimos anos uma das formas mais importantes de divulgar e recuperar conteúdos digitais. Contudo, a busca por recursos, nos moldes atuais, muitas vezes ainda não é satisfatória para o usuário, visto que ela traz muitas informações irrelevantes e não garante que as mais importantes sejam recuperadas.

A Web Semântica [1] tem como objetivo prover informações mais refinadas sobre o conteúdo dos recursos disponíveis na Web, que possam ser processadas por máquinas. Para isso, são necessárias marcações semânticas, ou seja, anotações que contenham informações sobre o conteúdo dos recursos da Web, por meio de uma linguagem expressiva o bastante para descrever os detalhes relevantes. Além disso, para que os recursos possam ser compartilhados por diferentes sistemas da Web, essa linguagem deve ser adotada como padrão pelas comunidades afins.

A ontologia é uma componente chave para o estabelecimento da Web Semântica [29][30], visto que ela possibilita a definição de novos tipos de informações, bem como relacionamentos entre conceitos, além de regras e axiomas. Na Figura 2.3, é possível verificar a alta expressividade na utilização de ontologias para a representação do conhecimento.

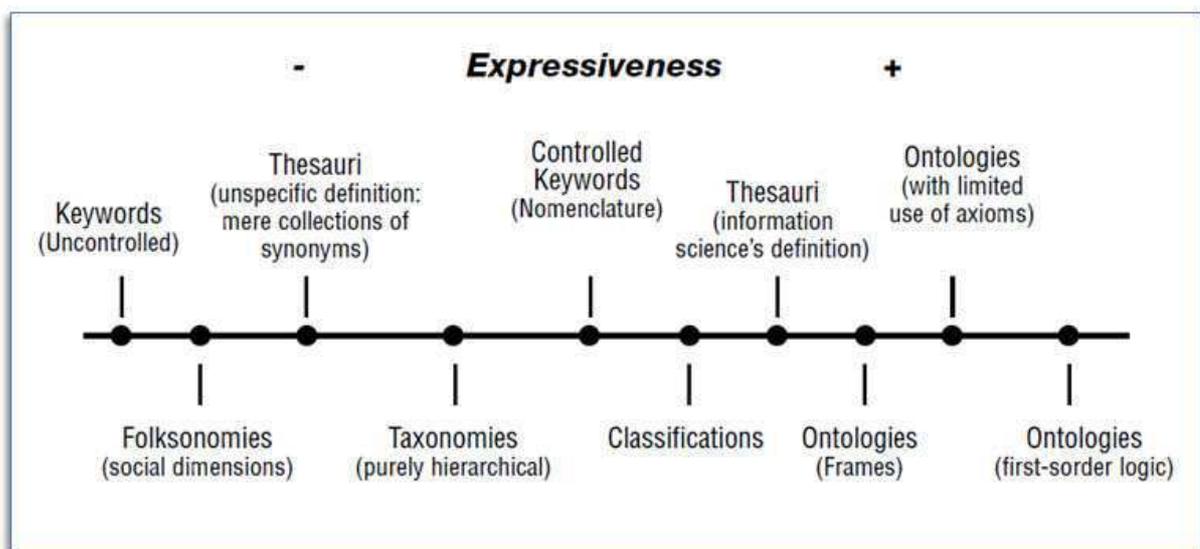


Figura 2.3 – Distinção de diferentes vocabulários de acordo com sua expressividade (extraída de [31]).

Este termo vem do grego “*ontos*”, que significa ser, e é a ciência que estuda o ser e suas propriedades. Sua origem veio dos filósofos gregos, especialmente Aristóteles, que investigaram as propriedades do mundo, dos objetos e como estes são percebidos.

Na computação, pesquisas em ontologias tiveram início nas comunidades de Inteligência Artificial (IA) e de Representação do Conhecimento [49]. Neste contexto, uma ontologia é um formalismo que permite especificar um vocabulário relativo a um determinado domínio. Esse vocabulário define entidades, classes, propriedades, predicados, funções e as relações entre estes componentes no contexto de algum domínio específico [50][51].

Embora o termo ontologia seja hoje em dia frequentemente usado e sua importância reconhecida, não existe um consenso sobre seu exato significado [52]. A comunidade de IA vê ontologias como teorias de lógica formal através das quais não só são definidos termos e relacionamentos, mas também o contexto no qual os termos estão aplicados, e fatos e relacionamentos inferidos. Por outro lado, outras comunidades, como a de bancos de dados, vêem ontologias mais como modelos de objetos, taxonomias e esquemas, não expressando explicitamente restrições importantes.

Ontologias lingüísticas como *WordNet*⁴ ou tesouros expressam vários relacionamentos entre conceitos (ex., sinônimos, antônimos, “é-um”, “parte-de”) mas não descrevem explícita e formalmente o que os conceitos significam [53].

Spaccapietra et al. [49] classificam ontologias em dois tipos: a taxonômica e a descritiva. A ontologia taxonômica abrange a primeira geração de ontologias, ainda muito usada, e está focada na definição de termos e na sua organização em hierarquias de generalização/especialização, enriquecidas com relações semânticas comumente usadas em lingüística, como sinônimo e antônimo. Uma ontologia taxonômica define um vocabulário de referência. Esse tipo de ontologia é relativamente fácil de ser usado. *WordNet* é um exemplo de ontologia taxonômica.

As ontologias descritivas e os esquemas conceituais de bancos de dados compartilham o mesmo esforço de modelar algum domínio ou atividade. No entanto, enquanto as ontologias têm sido tradicionalmente consideradas um meio de descrever o mundo, os esquemas de bancos de dados também auxiliam o gerenciamento dos dados que representam o mundo de interesse.

⁴ <http://wordnet.princeton.edu/>

Como uma ontologia provê um conjunto de conceitos e termos para descrever algum domínio, ela representa a estrutura básica sobre a qual uma base de conhecimento poderá ser construída. A base de conhecimento faz uso desses termos para representar o que é verdadeiro sobre algum mundo real ou hipotético. Ontologias com instâncias são freqüentemente referenciadas como bases de conhecimento [54].

De acordo com o tipo de conhecimento que uma ontologia descreve, ela pode ser classificada como genérica ou dependente do domínio. Uma ontologia genérica especifica conceitos gerais definidos independentemente do domínio da aplicação, podendo ser usada em diferentes domínios. OpenCyc⁵, SUMO⁶ (Suggest Upper Merged Ontology) e *WordNet* são exemplos de ontologias genéricas. Uma ontologia dependente do domínio é dedicada a um domínio particular, capturando o conhecimento de especialistas referente a uma determinada aplicação, como, por exemplo, uma ontologia biomédica [50][53].

Uma ontologia precisa ser especificada em alguma linguagem formal e implementada para que seja entendida, compartilhada e exportada entre diferentes usuários. Existem várias linguagens de especificação de ontologias, com graus de formalismo diferentes. Algumas são baseadas em XML (*eXtensible Markup Language*) como XOL (*XML-Based Ontology Exchange Language*) [55], SHOE (*Simple HTML Ontology Extensions*) [56] e RDF (*Resource Description Framework*) [58], outras em lógica de descrição como KIF (*Knowledge Interchange Format*) [57], existindo ainda aquelas que combinam XML e lógica de descrição, como OIL [59] e DAML+OIL [60].

RDF é um arcabouço desenvolvido pelo W3C (*World Wide Web Consortium*) para a representação de informação na Web utilizando a sintaxe XML. RDF Schema (RDFS) especifica os vocabulários, em RDF, usados para a representação de informação na Web por meio de classes, propriedades, tipos, intervalos e domínios.

RDF/RDFS é um padrão voltado para a Web semântica recomendado pelo W3C para representação de informação. DAML+OIL é uma combinação de DAML (*DARPA Agent Markup Language*) e OIL (*Ontology Inference Layer*) que estende a linguagem RDF/RDFS adicionando mais expressividade. Ela tem uma sintaxe RDF/XML baseada no paradigma de *frames* e descreve a estrutura de um domínio (esquema) em um estilo orientado a objetos.

⁵ <http://www.opencyc.org/>

⁶ <http://www.ontologyportal.org/>

A *Web Ontology Language* (OWL) [61] é uma linguagem de marcação para a descrição de ontologias. Ela foi projetada pelo *Web Ontology Working Group*. Assim como RDF, OWL é um padrão recomendado pelo W3C para a Web Semântica, tanto para especificação como para intercâmbio de ontologias. As principais influências na especificação da linguagem OWL vieram da lógica de descrição, das linguagens RDF/RDFS e DAML+OIL, e do paradigma de *frames*. Nas linguagens baseadas em *frames*, cada conceito é descrito como uma classe e cada classe é descrita em um *frame*. Em mais detalhes, classes são definidas por atributos (*slots*) estruturadas de acordo com o relacionamento com as subclasses dentro de um grafo ou de uma taxonomia.

Há várias propostas de ferramentas para o desenvolvimento de ontologias como, por exemplo, Protégé⁷, KAON⁸, Ontolingua⁹, Ontosauros¹⁰, OntoBroker¹¹, WebOnto¹² e RACER¹³.

2.5. Folksonomia

O termo folksonomia foi cunhado em 2004 por Thomas Vander Wal, e é uma analogia ao termo taxonomia tendo como principal característica a criação de anotações a partir do linguajar das pessoas que a utiliza. Em outras palavras, folksonomia é uma forma relacional de categorizar e classificar informações disponíveis na Web, sejam elas representadas por meio de textos, imagens, áudio, vídeo ou qualquer outro formato.

A finalidade da folksonomia é ordenar o caos existente na Web. Embora a sua característica de liberdade para classificar aponte para a idéia de uma falta de estrutura organizacional, o resultado para quem pesquisa é uma maior facilidade para encontrar termos que as demais linguagens de indexação não conseguem acompanhar em suas tabelas hierárquicas [67].

⁷ <http://protege.stanford.edu/>

⁸ <http://kaon.semanticweb.org/>

⁹ <http://ksl.stanford.edu/software/ontolingua/>

¹⁰ <http://www.isi.edu/isd/ontosaurus.html>

¹¹ <http://ontobroker.semanticweb.org/>

¹² <http://projects.kmi.open.ac.uk/webonto/>

¹³ <http://www.sts.tu-harburg.de/~r.f.moeller/racer/>

Dessa forma, o uso de anotações para classificar recursos na Web é conhecido como *tagging* ou folksonomia, de *Folk* - povo/pessoas e Taxonomia – estudo e classificação sistemática. Assim, o termo folksonomia pode ser traduzido como “classificação feita pelas pessoas”.

A folksonomia se baseia no conceito de Web 2.0, pois, como esta, é realizada de maneira colaborativa, “funciona através da atribuição de anotações, pelos próprios usuários da Internet, a arquivos disponibilizados *online*. Assim, é o usuário que representa e recupera informações através das anotações que ele mesmo cria” [68].

A folksonomia é apresentada como uma classificação que objetiva facilitar a recuperação da informação, isto pelo fato de ser realizada pelo próprio usuário, que assume papel significativo na Web atual.

Como vantagens na adoção da folksonomia na classificação de conteúdos na Web, existem o cunho colaborativo/social e a formação de comunidades. Por outro lado, a falta de controle do vocabulário resultante da liberdade de atribuição de anotações pode gerar falta de precisão nas buscas.

Na folksonomia, após a classificação, as anotações podem ser compartilhadas por um grande conjunto de serviços, relacionando todos os conteúdos, independente de onde eles estejam. As anotações são utilizadas para organizar e facilitar a busca de páginas e recursos na Internet, habilitando os usuários a criarem cabeçalhos de assunto que melhor representem o documento ou objeto que pretendem classificar. Todo esse processo ocorre de maneira clara e objetiva, o usuário acrescenta uma ou mais anotações a seu conteúdo, usando o critério que julgar mais relevante para o momento da busca.

Muitos sites já fazem uso da folksonomia para a classificação e organização de seus conteúdos, tais como: Del.ici.ous, Connotea, Flickr, Wikipédia, YouTube, Last.fm e o Bibsonomy.

Assim, diante do exposto, é possível perceber a grande contribuição do uso da folksonomia para a classificação e organização de conteúdos disponíveis na Internet. O usuário da sociedade atual necessita mais que ter acesso à informação, ele busca interagir com elas, de maneira que este possa sentir-se não só receptor passivo, mas também autor, emissor e editor, atuando de maneira visível e colaborando com outros usuários.

2.6. Gazetteers

No início da década de 90, começaram a surgir as primeiras bibliotecas digitais com informações que estavam dispostas em livros, revistas e enciclopédias. Logo após, passaram a surgir às bibliotecas digitais georeferenciadas, conhecidas como *Geolibraries* [34]. Com a finalidade de incrementar estas bibliotecas surgiram os *Gazetteers*, ferramentas capazes de indexar informações espaciais e não espaciais.

Vários tipos de informação possuem referência a locais específicos da superfície da terra. Relatórios sobre o meio ambiente, fotografias de um determinado lugar, imagens da terra obtidas por satélites, estatísticas econômicas e demográficas, guias de cidades e planos municipais, são exemplos de informações que podem ser georeferenciadas. A forma que tem prevalecido para se fazer a geo-referência é através do nome, associado às coordenadas geográficas.

Os usuários que pretendem localizar informações relevantes sobre um determinado assunto necessitam, muitas vezes, fazê-lo através da referência a uma localização específica, usualmente descrita por um nome geográfico. Por exemplo, “Encontrar todas as informações relacionadas à floresta amazônica”. Os usuários têm interesse em encontrar todos os itens relevantes que contenham as palavras *floresta amazônica* e também informações como fotografias aéreas, imagens de satélites, vídeos e áudios, que se referem à área geográfica da floresta amazônica. Essa forma de referência espacial, feita através do nome do lugar geográfico, é conhecida como referência indireta e é suportada pelo uso de Gazetteers.

Um Gazetteer é considerado como um dicionário de nomes geográficos, provendo nome, localização e as características dos dados geográficos [10]. O termo ‘nomes geográficos’ é utilizado para determinar um “lugar” ou “tipo” geográfico, tal como, cidade (tipo) de Campina Grande (lugar). A localização geográfica é normalmente indicada pela coordenada de latitude e longitude relativas ao lugar geográfico.

Em um Gazetteer, a referência indireta é feita através da entrada do nome do lugar geográfico para recuperar informações, tais como, as coordenadas, imagens espaciais, tipo ou categoria e outras informações sobre o lugar geográfico de interesse. Dentre alguns

exemplos de Gazetteers estão: *Columbia Gazetteer of the World* [35]; *Getty Thesaurus of Geographic Names*¹⁴; *GeoNet Names Server*¹⁵; e *Geographic Names Information System*¹⁶.

Nos Gazetteers, um lugar geográfico é caracterizado por um conjunto de atributos. Alguns destes atributos são indispensáveis na definição de um lugar geográfico [37]:

- **Nome:** o nome determina um lugar geográfico. Porém, o atributo nome não é único, para um mesmo lugar geográfico podem existir nomes diferentes, por exemplo: *Pequim* e *Beijing* na China. Da mesma forma, vários lugares geográficos podem possuir o mesmo nome, por exemplo, nos Estados Unidos existem 18 cidades chamadas *Springfield*.
- **Localização:** a localização determina as coordenadas (latitude e longitude) que podem ser expressas através de um ponto, um retângulo ou um polígono. Por exemplo, a localização do Brasil pode ser feita com o esboço de um retângulo, sobre um mapa geográfico, utilizando dois pares das coordenadas de latitude e longitude, que representam os dois pontos extremos do retângulo, tais como: 74°.20'O, -33°.66'S e -34°.78'L, 05°.20'N;
- **Tipo:** o tipo determina a característica de um lugar geográfico. Por exemplo, o tipo 'país' contém lugares geográficos como Brasil, Itália e Argentina.

O atributo "tipo" pode incluir diferentes categorias como: áreas políticas (cidades, estados, países, etc); áreas turísticas (parques, museus, monumentos, etc) características naturais (montanhas, vulcões, florestas, oceanos, ilhas, etc); estruturas feitas pelo homem (pontes, canais, torres, etc); e ainda áreas imprecisas (sudeste da Paraíba, sul do Brasil, etc) [36].

Uma categoria pode ser definida de forma mais específica, como lagos, ou mais genérica, como características hidrográficas, permitindo assim uma maior flexibilidade no processo de consultas ao Gazetteer.

Atualmente, existem Gazetteers com diferentes domínios [38]. Por exemplo, o *New York Times Atlas*¹⁷ possui uma seção que pode ser usada para observar um nome geográfico e encontrar páginas (URLs) com informações do lugar pesquisado.

¹⁴ http://www.getty.edu/research/conducting_research/vocabularies/tgn

¹⁵ <http://www.nima.mil/gns/html>

¹⁶ <http://geonames.usgs.gov/>

¹⁷ <http://geogdata.csun.edu/NYpage1.html>

Alguns Gazetteers possuem informações sobre lugares e diversos tipos de características, como por exemplo, a história do local, dados da população, dados físicos como altitude e mudanças geopolíticas que ocorreram em um período de tempo.

Existem ainda Gazetteers onde as informações sobre os lugares geográficos estão disponíveis em termos de hierarquias através de thesaurus, designados para recuperação de informações. Como exemplo de tais sistemas, pode-se citar os sistemas do governo americano: *U.S. Library of Congress*¹⁸ e o *GeoRef Thesaurus*¹⁹ desenvolvido pela *American Geological Institute*.

O governo americano também possui agências que mantêm Gazetteers que contêm nomes oficiais de lugares geográficos (cidades, estados, países). Como exemplo, tem-se: *National Imagery and Mapping Agency's Geographic Names Processing System*²⁰; e *U.S. Board of Geographic Names*²¹.

Com relação às consultas que podem ser efetuadas em um Gazetteer, as principais são: a textual, através do nome do lugar geográfico; e a forma espacial, através da seleção de uma região geográfica sobre um mapa. Seguem alguns exemplos de consultas a Gazetteers.

- O usuário entra com o “nome” de um lugar geográfico e encontra a localização espacial disposta em uma imagem, com coordenadas geográficas e outras informações;
- O usuário entra com um “tipo” ou categoria, tal como, áreas políticas, e encontra um conjunto de lugares geográficos deste “tipo”, com informações específicas de cada lugar;
- O usuário esboça sobre o mapa uma região geográfica e encontra nomes e características daquela região. Esse tipo de consulta pode utilizar o filtro por categorias ou “tipos”, por exemplo: ‘encontrar todas as características hidrográficas (lagos, pântanos e rios) de uma determinada região’. A área de pesquisa é selecionada sobre o mapa através de técnicas de seleção espacial, tais como: MBR (menor polígono possível que cobre a área de um lugar geográfico), Centróide e Fecho Convexo.

¹⁸ <http://www.loc.gov>

¹⁹ <http://www.agiweb.org/georef/>

²⁰ <http://www.nima.mil>

²¹ <http://mapping.usgs.gov/www/gnis/bgn.html>

Enfim, os Gazetteers representam uma forma inteligente de recuperação da informação. A grande capacidade que eles possuem, em relacionar representações gráficas de um objeto com nomes de lugares e localização geográfica, tem incentivado diversas pesquisas em torno da criação de técnicas de seleção espaciais mais refinadas e eficientes [39].

2.7. Ciência de Contexto

A área da computação ubíqua foi uma das pioneiras no estudo e utilização do conceito de contexto e, com isso, demonstrou o potencial da aplicação desse aspecto nos sistemas computacionais. Atualmente, o conceito de contexto tem sido objeto de estudo em diversas áreas, como a Psicologia Cognitiva e a Linguística, além da própria Computação. Sendo que, nesta última, a motivação dos pesquisadores abrange, além da computação ubíqua, áreas como: Inteligência Artificial, Interação Humano-Computador, Sistemas Colaborativos, Hiperídia Adaptativa, entre outras, que buscam estudar tal conceito e pesquisar como este pode ser formalizado e empregado pelos sistemas computacionais [91].

No início da década de 90, Weiser [92] vislumbrou ambientes físicos habitados com dispositivos computacionais que auxiliariam nas tarefas diárias, fornecendo aos usuários informações e serviços de forma contínua e onipresente, o mais transparente possível.

A presença da computação em meios ubíquos dá aos usuários a expectativa de que é possível acessar as informações e serviços em qualquer lugar. Além disso, a mobilidade fornecida pela ubiquidade faz o contexto do usuário tornar-se mais dinâmico. Aliado a isso, as pessoas estão acostumadas, nas ações realizadas em sua vida cotidiana, a fazer uso de informações situacionais implícitas para tomar decisões.

Sendo assim, entender o contexto em que ocorre uma determinada interação é fundamental para que os indivíduos possam responder de maneira apropriada a uma determinada situação [93]. Em linhas gerais, pode-se então dizer que as pessoas têm a capacidade de usar informações situacionais implícitas, ou informações contextuais, ou simplesmente contexto, para melhorar a qualidade da conversação e compreender certas circunstâncias, ações ou eventos através de sua caracterização.

Contudo, esta habilidade de transmitir ideias não é feita de forma transparente na interação do homem-máquina sem que haja interferência. Desta forma, ao criar mecanismos de acesso de contexto nos computadores, a comunicação na interação homem-máquina torna-se mais rica e possibilita que sejam oferecidos serviços mais convenientes.

De modo geral, as informações contextuais podem ser identificadas a partir de seis dimensões básicas que são bastante referenciadas na literatura e conhecidas como 5Ws + 1H [94][95]: *when* (tempo), *where* (localização), *who* (identificação), *what* (ação, atividade), *why* (motivação por trás das ações) e *how* (meio, que identifica como os elementos de contexto são coletados).

O conceito de contexto possui diversos significados na literatura, sendo que muitos deles são complementares e outros limitados à área da computação. Bazire e Brézillon [96] coletaram um conjunto de aproximadamente 150 definições de contexto, originadas de diferentes domínios e chegaram a duas principais conclusões: (i) o contexto atua como um conjunto de restrições que influenciam o comportamento de um sistema embutido em uma dada tarefa; e (ii) a definição de contexto depende da área de conhecimento à qual pertence.

No entanto, há uma definição genérica e clássica de contexto fornecida por Dey [97], que é bastante referenciada e aceita na comunidade acadêmica:

“Contexto é qualquer informação que pode ser utilizada para caracterizar a situação de uma entidade. Uma entidade, por sua vez, pode ser uma pessoa, lugar ou objeto que é considerado relevante para a interação entre um usuário e uma aplicação, incluindo o usuário e a aplicação em si.”

Diante da definição acima, o contexto só existe quando relacionado a uma entidade, ou seja, havendo uma entidade, qualquer informação útil que a circunda em uma determinada situação é considerada contexto. Por exemplo, um telefone celular sempre irá vibrar e nunca emitir um alerta sonoro em um concerto, caso o sistema reconheça a localização do telefone celular e o horário do concerto [94].

Na Figura 2.4, é possível verificar que os sistemas tradicionais consideram apenas as informações fornecidas pelos usuários na forma de dados de entrada, diferentemente do que ocorre com um sistema ciente de contexto, que considera as informações explícitas fornecidas pelos usuários, aquelas armazenadas em uma base de conhecimento contextual, aquelas inferidas por meio de raciocínio e/ou ainda aquelas percebidas a partir do ambiente. Com base nessas informações contextuais, o sistema tem condições de

identificar e otimizar fluxos de interação, determinar ações e adaptações e fornecer assistências às tarefas que estão sendo realizadas [91].

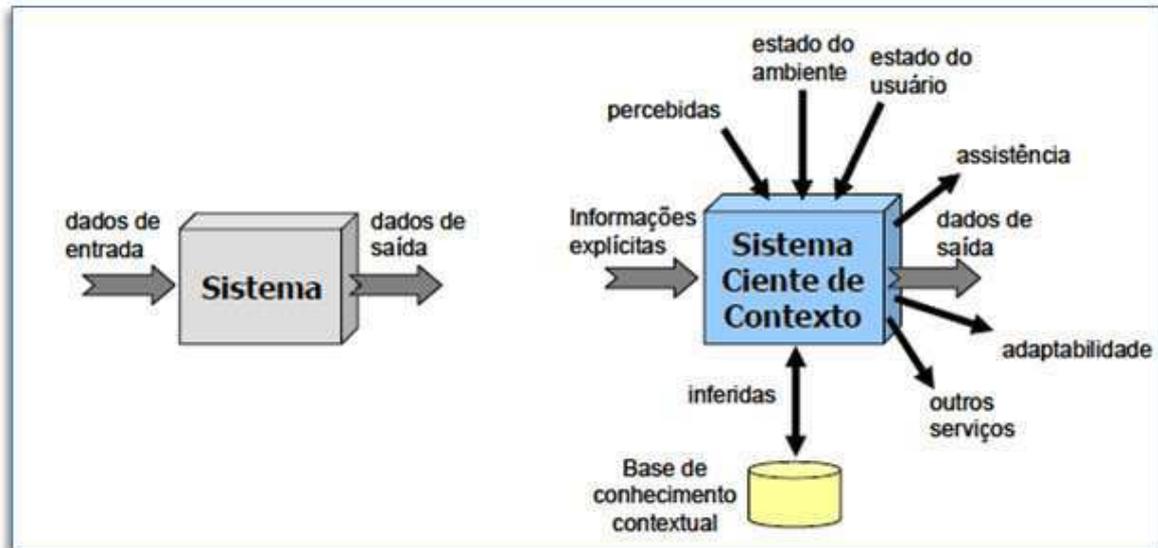


Figura 2.4 – Sistema tradicional VS. Sistema Ciente de Contexto – adaptada de [91].

Diversas técnicas vêm sendo utilizadas na literatura por diferentes sistemas para representar o contexto, como par chave-valor, linguagem de marcação, orientação a objetos, mapas de tópicos e ontologias. Segundo Vieira et. [91], não há uma abordagem ideal para todos os sistemas cientes de contexto uma vez que diferentes sistemas impõem diferentes restrições.

Vale ressaltar que, neste trabalho, o conceito de contexto está associado apenas à localização geográfica usuário.

2.8. Considerações Finais

Neste capítulo, foram apresentados alguns dos principais conceitos relacionados ao entendimento do presente trabalho, tais como: Recuperação da Informação, Expansão de Consultas, Folksonomias, dentre outros. No capítulo seguinte, serão apresentados alguns trabalhos recentes da literatura que lidam com o aprimoramento do funcionamento dos mecanismos de busca da Web.

Capítulo 3

Trabalhos Relacionados

Na área de RI, existe diferença entre dado e informação. Conforme [11], a recuperação de um dado consiste, basicamente, em determinar os recursos que possuem as palavras-chave de uma determinada consulta. Por outro lado, a informação é um dado relevante para o usuário.

Desta forma, é preciso caracterizar a necessidade do usuário para que apenas recursos relevantes sejam apresentados como resultado de uma consulta. Porém, este é um problema complexo e, por isso, novas abordagens vêm sendo desenvolvidas ao longo dos anos na área de RI, tais como: a análise sintática e semântica dos termos, o uso de personalização e contexto, dentre outras.

Neste capítulo, serão relatados alguns trabalhos recentes na área de RI que tratam da melhoria dos mecanismos de busca através de anotações, folksonomias e ontologias.

O restante do capítulo está organizado da seguinte forma. Na seção 3.1, são apresentados trabalhos sobre expansão de consultas e o uso de anotações no auxílio da mesma. Na seção seguinte, são mostradas pesquisas sobre o uso de ontologias e folksonomias para melhorar os mecanismos de busca. Na seção 3.3, é apresentado o GeoSen, o motor de busca geográfico que foi estendido com as idéias propostas neste trabalho. Por fim, na seção 3.4, são mostradas as considerações finais sobre o levantamento bibliográfico feito para o desenvolvimento do GeoSen_Tags, bem como suas principais contribuições.

3.1. Expansão de consultas

A expansão de consultas é um processo de adição de termos que parte do pressuposto que a consulta inicial fornecida pelo usuário, geralmente, não representa adequadamente suas necessidades [4][13][14][15].

A busca por termos similares tem sido objeto de pesquisa ao longo dos anos. As abordagens estatísticas têm provado serem úteis [4]. Por outro lado, a utilização de thesaurus apresenta problemas de ambigüidade, que acaba por se propagar entre os sinônimos [5].

As abordagens empregadas podem ser classificadas em duas grandes categorias: as que se baseiam no resultado das consultas e as que fazem uso de alguma base de conhecimento. A primeira classe depende do processo de busca e utiliza feedback para a geração de novos termos [15][16][17]. Já a segunda, não depende do processo de busca e para encontrar termos similares para a expansão, ela navega em alguma estrutura semântica, tais como ontologias [2][13][18].

Estudos recentes apontam para a elaboração de técnicas que consideram parâmetros como o contexto e a preferência dos usuários, visto que estes possuem um papel fundamental no cenário dos mecanismos de busca. Além disso, com o avanço das redes sociais e a anotação de recursos, a colaboração entre os usuários também é um aspecto que está sendo considerado [6][7].

Carman et al. [19] descrevem os resultados de experimentos realizados para explorar diferentes soluções para expansão de consultas. Eles concluem que adicionar informações extraídas de perfis de usuários pode ajudar no ranking de relevância de documentos.

Chirita et al. [22] propõem um método para aprimorar buscas com poucas palavras-chave. Para isso, a consulta é expandida com termos coletados de um repositório de informações pessoais dos usuários. O processo de geração de novos termos é dividido em cinco etapas, de tal forma que são analisados os dados dos usuários em diferentes granularidades, partindo de termos individuais para composições de palavras, com estatísticas de co-ocorrência globais, além do uso de thesaurus.

Cucerzan e White [23] descrevem uma abordagem para sugestão de termos através da seleção de refinamentos de consultas. Estes são feitos a partir da combinação de

diversos padrões de navegação dos usuários e dos logs de consultas de motores de busca. Eles concluem que é eficaz utilizar ambas as informações para sugerir novas palavras.

No trabalho [20], os autores propõem um método, denominado *Social Ranking*, que tem como objetivo melhorar a busca por recursos no cenário da Web 2.0. A técnica explora o conceito de similaridade entre usuários e entre anotações.

Na similaridade entre os usuários, a idéia é que quanto maior for a similaridade entre o usuário que está fazendo a consulta e o usuário que anotou o recurso, maior será a chance de o recurso ser relevante para o usuário que fez a consulta.

Na similaridade entre anotações, quanto maior for o número de recursos que duas anotações compartilham, maior será sua similaridade.

As fórmulas para o cálculo da similaridade entre usuários e anotações, respectivamente, são:

$$\text{sim}(u_i, u_j) = \text{cos}(v_i, v_j) = \frac{v_i \cdot v_j}{\|v_i\| * \|v_j\|}$$

Equação 3.1 – Similaridade entre anotações (extraída [20]).

Onde u é um usuário e v o vetor que contém o número de vezes que uma anotação é associada a um usuário.

$$\text{sim}(t_i, t_j) = \text{cos}(w_i, w_j) = \frac{w_i \cdot w_j}{\|w_i\| * \|w_j\|}$$

Equação 3.2 – Similaridade entre recursos (extraída [20]).

Onde t é uma anotação e w o vetor que contém o número de vezes que uma anotação é associada a um recurso.

O Gossple [21] é uma solução, centrada no usuário, para melhorar a navegação na Internet. Dessa forma, o processo de expansão de consultas é feito considerando anotações que são “próximas” de usuários “próximos”.

Para desenvolver a técnica proposta, os autores mostram como é criada, automaticamente, uma rede de relacionamentos, cujas conexões representam a estimativa de distância entre os usuários que atribuem anotações a recursos. Tal estimativa é calculada através da similaridade entre as anotações e os recursos, pela fórmula do cosseno.

Na pesquisa de Biancalana e Micarelli [7], é apresentada uma máquina de busca desenvolvida com o intuito de explorar o potencial do método de expansão de consultas via uma abordagem social. De modo geral, os autores sugerem que o comportamento dos usuários deve ser armazenado e interpretado para auxiliar no fornecimento de resultados de consultas que satisfaçam as necessidades reais dos usuários. Dessa forma, sempre que um usuário faz uma consulta e visualiza um resultado, tais ações são gravadas no seu perfil.

Na expansão da consulta, todos os significados dos termos são levados em consideração, de tal forma que caso haja termos ambíguos, recursos relacionados a todos esses termos farão parte do resultado da consulta, sendo diferenciados por temas. Por fim, para modelar o usuário é proposta uma matriz de correlação tridimensional (Figura 3.1), onde cada termo da matriz é ligado inicialmente a um nível intermediário, que contém classes de categorização, para depois se relacionar com os outros termos presentes na matriz.

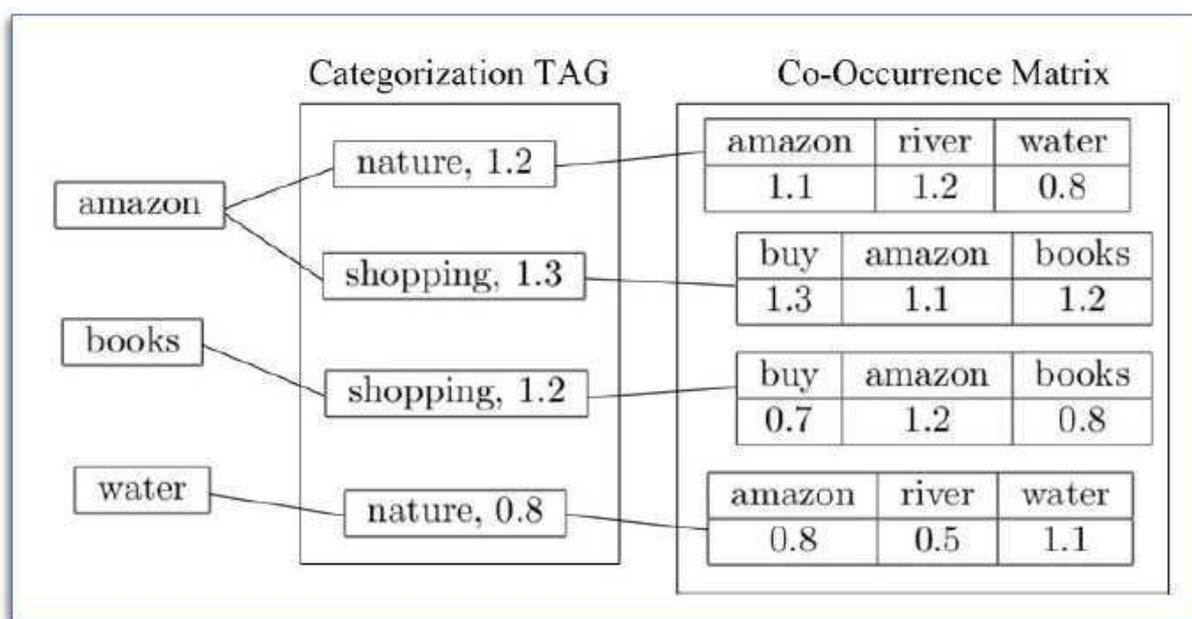


Figura 3.1 – Exemplo de matriz de correlação tridimensional (extraída de [7]).

Wang e Davison [26] exploram o uso de anotações no processo de expansão e sugestão de consultas.

Primeiro, eles desenvolveram um módulo para anotar automaticamente recursos da Web. Este módulo foi implementado por um conjunto de classificadores binários, onde cada um é focado em uma anotação, já que eles partem do pressuposto de que as anotações são independentes.

No processo de classificação foi adotada a técnica SVM (Support Vector Machine) e os dados para treinamento foram coletados do *del.icio.us*.

Tendo esse módulo de anotação automática disponível, o módulo de sugestão de consultas funciona da seguinte maneira: após o resultado da busca, são analisados os dez primeiros documentos para se ter uma idéia do significado dos termos da consulta inicial, de modo a poder escolher os termos mais adequados para sugestão. Essa análise é auxiliada pela base de anotações dos recursos previamente gerada.

Delboni et al. [100] exploram o uso do escopo geográfico de consultas, visto que 20% delas possuem referências geográficas. Para isso, eles analisam expressões naturais da linguagem, denominadas expressões posicionais, de modo que informações que dizem respeito a lugares possam ser detectadas. Como resultado, é proposta uma técnica de expansão de consultas para obtenção de recursos mais relevantes considerando os termos geográficos inferidos da consulta inicial do usuário.

3.2. Folksonomia e Ontologias

Estudos recentes na área da Web Semântica e da Web 2.0 têm levado a um grande interesse por métodos de representação do conhecimento. De um lado, as ontologias estão sendo elaboradas para realizar integrações avançadas de informações. Do outro, usuários estão atribuindo anotações descritivas a uma grande coleção de documentos. Ainda assim, ontologias e folksonomias não devem ser vistas como sistemas rivais [25].

De acordo com Vander Wal [26], duas grandes categorias de folksonomias podem ser distinguidas: as que permitem diferentes anotações de diversas pessoas para um mesmo documento (por exemplo, *del.icio.us*) e as que limitam o número de anotações por documento para apenas uma, pertencente ao seu autor (Flickr). Devido à natureza da primeira categoria, é possível gerar as nuvens de anotações, que representam a popularidade das anotações de um determinado recurso ou de toda a folksonomia.

Xu et al. [27] propõem um framework de busca personalizado, que emprega o conceito de folksonomia. Para isso, eles exploram três propriedades desse conceito:

categorização, palavra-chave e estrutura. Diante disto, é sugerido um algoritmo de ranking que considera não apenas o casamento dos termos da consulta com os recursos, mas também o casamento dos interesses do usuário com os tópicos dos documentos.

Di Matteo e Peroni [28] introduzem o FolksEngine, uma arquitetura de motor de busca paramétrica para folksonomias, que permite a especificação de algoritmos de agrupamento das anotações através de três passos:

- expansão dos termos da consulta pela aplicação de regras semânticas;
- execução da consulta expandida em uma folksonomia; e
- ordenação dos resultados de acordo com regras semânticas associadas às anotações da folksonomia.

As ontologias desempenham um papel fundamental no campo da Web Semântica, visto que elas auxiliam na elaboração de significados para recursos da Web, servindo como uma fonte precisa de termos para este tipo de metadado [8][9]. Isto faz com que os sistemas de busca da Web ampliem suas capacidades de recuperação da informação.

Weller e Peters [32] relatam métodos para representar o conhecimento através da análise dos relacionamentos descritos nas ontologias. Dentre os principais estão:

- Controle do vocabulário via a captura de sinônimos; e
- Classificação de conceitos para dispô-los em uma hierarquia;

Diante do que foi discutido nesta seção, é possível perceber dois grandes problemas, um relacionado à folksonomia e outro a ontologia. Na folksonomia, existe uma falta de controle na utilização do vocabulário. Já na ontologia, o custo de criação das mesmas é muito alto, por ser complexa e muitas vezes específica de um determinado domínio.

Angeletou et al. [33] apresentam o FLOR, um projeto que tem como intuito melhorar as buscas em folksonomias através do enriquecimento semântico do “espaço de anotações”. Para isso, é explorado o uso de fontes heterogêneas de conhecimento de grande escala disponíveis tais como ontologias e o WordNet.

O algoritmo empregado no FLOR, (Figura 3.2), transforma, de modo automático, o “espaço de anotações” em uma estrutura mais rica em semântica, via associação de cada anotação com um conceito que representa seu significado e pelo estabelecimento de relacionamentos formais entre os conceitos.

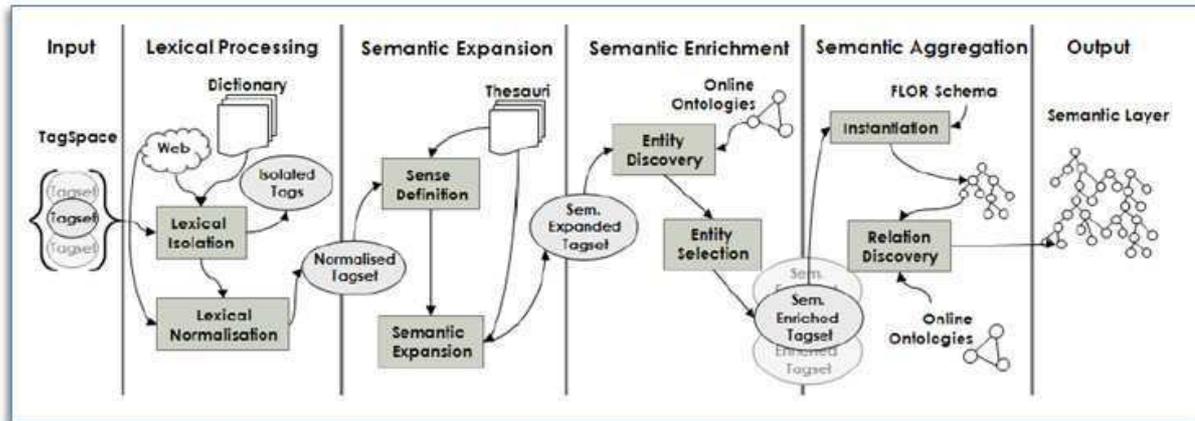


Figura 3.2 – Os principais passos do algoritmo empregado no FLOR (extraída de [33]).

Borges et al. [99] apresentam uma abordagem para descoberta de lugares geográficos através do reconhecimento, extração e decodificação de endereços urbanos. Para isso, são empregadas ontologias de domínios com duas granularidades: cidades e locais dentro da cidade.

3.1. GeoSen

Muito dos documentos disponíveis na Web possui algum tipo de contextualização geográfica. Além disso, para muitas buscas realizadas pelos usuários, os resultados mais relevantes estão relacionados a uma determinada região geográfica.

Contudo, para lidar com esse tipo de informação é preciso efetuar um tratamento especial nos dados, o que não é feito por motores de busca tradicionais. Diante disto, o GeoSen [69][70] foi desenvolvido com o intuito de realizar a recuperação da informação considerando o âmbito textual e geográfico.

O GeoSen, é um protótipo de motor de busca para Web, com enfoque geográfico, que implementa os modelos e técnicas definidos para a construção de um sistema Recuperação da Informação Geográfica. Ele tem como princípio a utilização de tecnologias livres, que são adotadas, por exemplo, na plataforma operacional, nas ferramentas de programação e nos servidores de aplicação e de gerenciamento de bancos de dados.

A arquitetura do GeoSen é orientada a plugins, uma vez que o sistema estende o motor de busca Nutch²². Conforme Figura 3.3, os pontos de extensão adotados pelo GeoSen são:

- *Parser* – realiza a leitura dos documentos capturados e extrai os dados a serem indexados. Nesta etapa, o GeoSen detecta termos geográficos, que são utilizados para análise do escopo geográfico dos documentos;
- *Indexing Filter* – efetua a indexação das informações dos documentos que passam pelo Parser. O GeoSen adiciona nos índices informações sobre o escopo geográfico dos documentos;
- *Query Filter* – efetua consultas nos índices gerados pelo passo de indexação. O GeoSen adiciona a capacidade de interpretar as especificações de lugares nas consultas dos usuários;
- *URL Filter* – analisa as URLs para determinar se estas devem ser acessadas ou ignoradas. Neste passo, o GeoSen também efetua seu algoritmo de detecção de lugares;

²² <http://nutch.apache.org/>

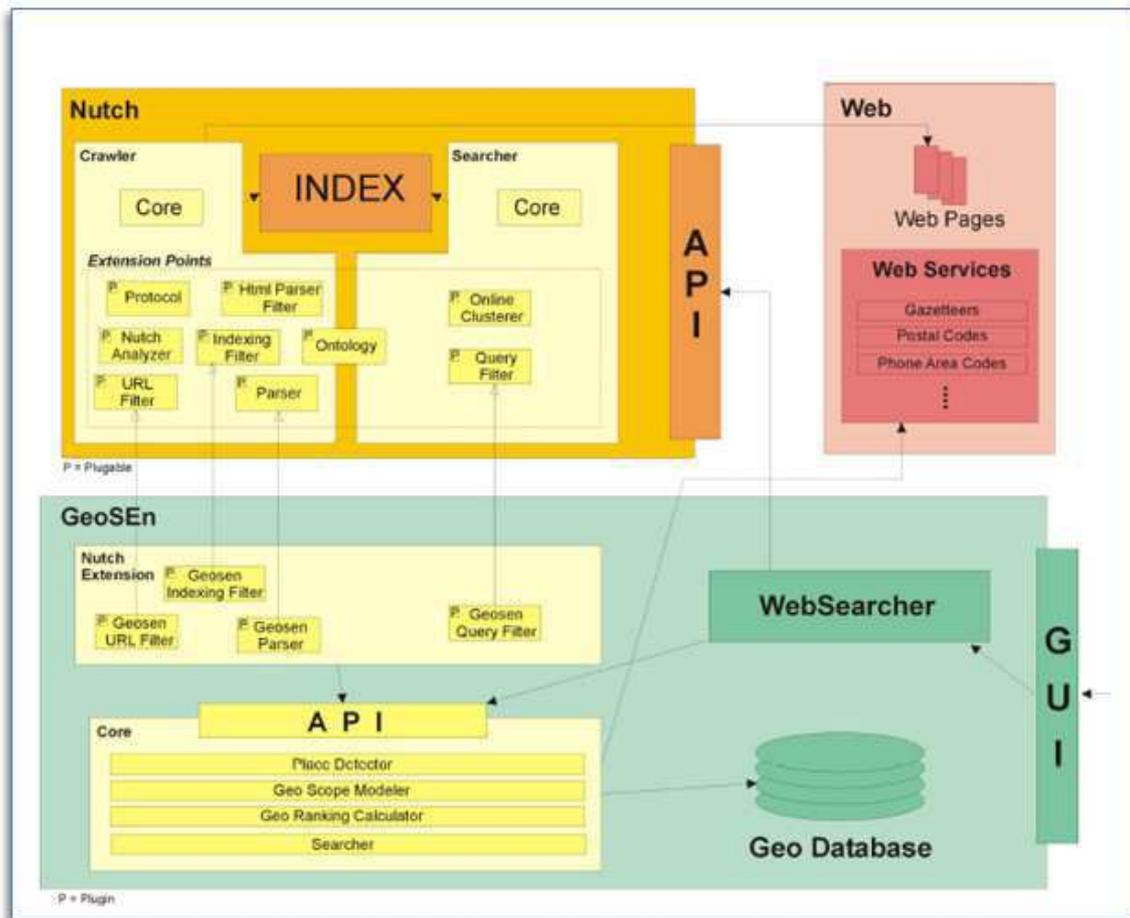


Figura 3.3 – Arquitetura do GeoSen (extraída de [71]).

Por fim, o GeoSen fornece uma interface multi-modo de consulta (Figura 3.4) onde os usuários podem interagir de modo textual e/ou espacial. O modo textual é similar aos encontrados em sistemas de busca tradicionais, com a entrada de dados sendo feita através de alguns campos de formulário. Já no modo espacial, o usuário dispõe de um mapa interativo, onde pode realizar operações básicas de visualização como *pan*, *zoom* e controle de camadas, além da seleção visual das localidades a serem utilizadas na consulta.



Figura 3.4 – Interface de consulta do GeoSen (extraída de [71]).

3.2. Considerações Finais

Neste capítulo, foram apresentados alguns trabalhos recentes na área de RI, que tratam do aprimoramento dos métodos de busca através de algoritmos de expansão de consultas. Estes empregam abordagens atuais que, além de tratar aspectos sintáticos e semânticos, também levam em consideração o usuário, que possui um papel fundamental neste cenário. Todavia, um parâmetro importante, como a localização geográfica, não é ponderado no uso do contexto do usuário.

Diante dos trabalhos citados, foram escolhidos alguns para comparação com o GeoSen_Tags, de modo que foram selecionadas nove características importantes de um

mecanismo de busca semântico e ciente de contexto, que estão sumarizadas na Tabela 3.1. Em seguida, na Tabela 3.2, relacionam-se estas características aos principais trabalhos descritos, através da marcação (•). A ausência desta marca significa que a característica não está presente ou que esta informação não foi encontrada na literatura.

Característica	Descrição
C1	Análise sintática dos termos.
C2	Análise semântica dos termos.
C3	Expansão dos termos da consulta.
C4	Uso de anotações.
C5	Construção de folksonomias.
C6	Uso de gazetteers.
C7	Uso de ontologias.
C8	Uso do contexto geográfico.
C9	Uso de tesauros.

Tabela 3.1 – Características importantes de um mecanismo de busca semântico e ciente de contexto.

Característica	FLOR	Gossple	FolksEngine	GeoSen	Borges et al.	Chirita et al.	GeoSen_Tags
C1	•		•			•	•
C2	•	•	•		•	•	•
C3	•	•	•		•	•	•
C4	•	•	•				•
C5		•					•
C6					•		•
C7	•		•		•		•
C8				•	•		•
C9						•	

Tabela 3.2 – Associação entre projetos e características importantes de um mecanismo de busca.

Tendo em vista os trabalhos relacionados e suas características mais relevantes, no próximo capítulo, serão mostradas, em detalhes, as soluções empregadas na presente pesquisa, bem como os aspectos arquiteturais e comportamentais do protótipo desenvolvido para sua validação.

Capítulo 4

GeoSen_Tags

Neste capítulo, apresenta-se um novo método de expansão de consultas baseado em anotações. A principal característica da técnica sugerida é o baixo custo, durante o processo de expansão, de análise dos termos de consultas feitas por usuários. Além disso, é mostrado um protótipo de motor de busca geográfico para Web, denominado GeoSen_Tags, que estende o GeoSen, com o objetivo de melhorá-lo através do uso de ontologias e gazetteers, juntamente com o novo método de expansão de consultas proposto.

O restante deste capítulo está organizado como segue. Na seção 4.1, é exposta uma visão geral do sistema. Em seguida, na seção 4.2, é apresentada a arquitetura do protótipo. Na seção 4.3, o algoritmo de expansão de consultas é descrito em detalhes. Na seção seguinte, é mostrado o processo de construção da folksonomia de documentos utilizada no GeoSen_Tags. Na seção 4.5, mostra-se como são utilizadas ontologias no auxílio à recuperação da informação. Já na seção 4.6, é explicado o uso de gazetteers pelo GeoSen_Tags. Por fim, na seção 4.7, são feitas as considerações finais do capítulo.

4.1. Visão Geral do Sistema

O GeoSen_Tags foi desenvolvido com o intuito de prover um motor de busca semântico e ciente de contexto, enriquecido geograficamente. Para isso, foi utilizado um motor de busca já existente: o GeoSen. A escolha pelo GeoSen, para ser estendido com as novas funcionalidades, deu-se pela idéia promissora que o sistema propõe da utilização do contexto geográfico, apoiada pelo uso de softwares livres, tais como:

ao sistema, já que existe uma separação entre as camadas, possibilitando que estas sejam modificadas de forma independente.

A primeira camada é a de apresentação ou de interface, na qual o usuário tem acesso as funcionalidades presentes no sistema. A segunda camada é a de controle, que tem como função receber as requisições da interface do usuário e repassá-las para um objeto específico da camada de negócio, além de receber a resposta e repassar para a camada de interface. Na terceira camada, é onde se encontra toda a lógica utilizada no sistema. Por fim, a camada de persistência é responsável por armazenar e fornecer os dados usados pelo protótipo.

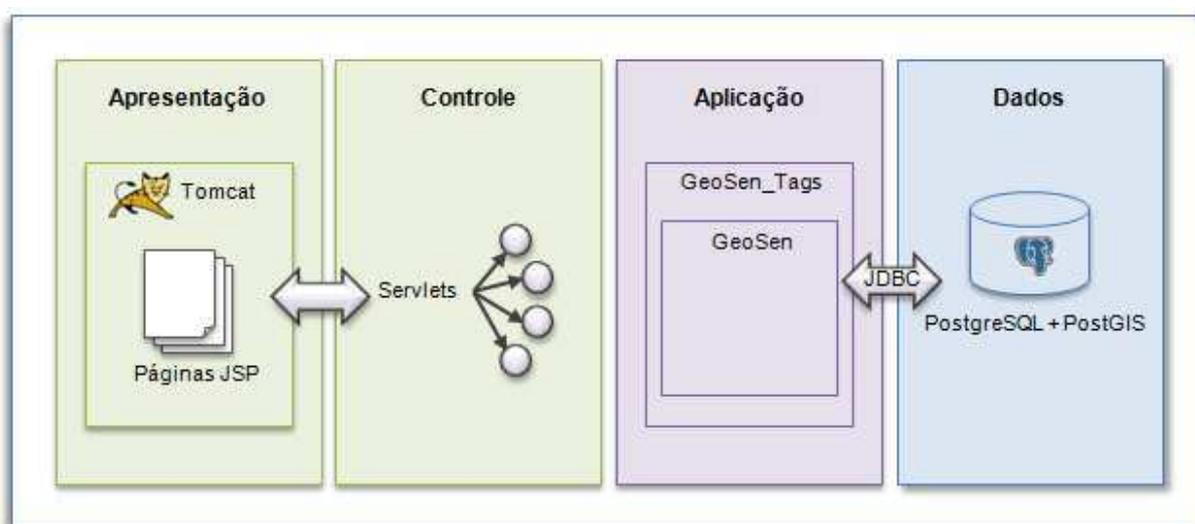


Figura 4.2 – Arquitetura em camadas, segundo o padrão MVC.

Com relação às tecnologias empregadas, o sistema foi desenvolvido através da linguagem de programação JAVA²³, devido as suas diversas vantagens como portabilidade, robustez e orientação a objetos, além de ser a linguagem usada pelo motor de busca GeoSen, que foi estendido com as novas funcionalidades propostas pelo GeoSen_Tags.

A interface do sistema foi implementada através de páginas JSP²⁴ (Java Server Pages), para os conteúdos dinâmicos, e HTML²⁵ (HyperText Markup Language), para os conteúdos estáticos, e são hospedadas no servidor de aplicações Apache Tomcat²⁶.

²³ <http://www.java.com>

²⁴ <http://java.sun.com/products/jsp>

²⁵ <http://www.w3schools.com/html/>

Na camada de controle foi utilizada a biblioteca Servlet²⁷, que facilita a implantação do padrão MVC.

Já na camada de dados, decidiu-se fazer uso da biblioteca JDBC²⁸ (Java Database Connectivity) e do SGBD (Sistema Gerenciador de Banco de Dados) PostgreSQL²⁹ por ser um software livre, possuir um certo grau de maturidade e robustez no tratamento de informações geográficas através da sua extensão PostGIS³⁰, que obedece ao padrão OpenGIS³¹ da OGC (Open Geospatial Consortium), bem como ser usado pelo GeoSen.

Na Figura 4.3, é exibida a arquitetura detalhada do GeoSen_Tags. O sistema foi desenvolvido sobre o GeoSen, um motor de busca com enfoque geográfico.

Inicialmente, o usuário submete uma consulta através da interface gráfica Web, que se comunica com o módulo chamado *Buscador Web*. Este é implementado via Servlets e tem como função receber todas as requisições do usuário e repassá-las para o núcleo do sistema processá-las. Feito isso, o *Buscador Web* recebe o resultado do processamento e envia a resposta para a interface do usuário.

O acesso ao núcleo do sistema é feito através de uma API (Application Programming Interface) disponibilizada pelo GeoSen. Dentre as principais operações dessa API estão: *indexação* espaço-textual de páginas da Web, computação de *ranking* de relevância espaço-textual, *consulta* aos documentos indexados e *deteção de lugares* geográficos nos termos informados pelo usuário.

A indexação é realizada periodicamente, independente das requisições do usuário. Neste processo, robôs capturam páginas da Web que passam por um *parser* que detecta dados com potencial de serem convertidos em informações geográficas. Por fim, as informações encontradas, juntamente com seu escopo geográfico, são armazenadas em arquivos de índices.

O *ranking* de relevância ordena os documentos recuperados pelo módulo de busca de acordo com uma ponderação atribuída a importância dos documentos em relação aos argumentos textuais e espaciais da busca submetida ao sistema.

²⁶ <http://tomcat.apache.org>

²⁷ <http://java.sun.com/products/servlet>

²⁸ <http://www.oracle.com/technetwork/java/javase/tech/index-jsp-136101.html>

²⁹ <http://www.postgresql.org>

³⁰ <http://postgis.refrations.net>

³¹ <http://www.opengeospatial.org>

Os dados contidos na *Base de Dados Geográfica* são coletados de diversas fontes, tais como IBGE³² (Instituto Brasileiro de Geografia e Estatística), Wikipedia³³ e Correios³⁴.

³² <http://www.ibge.gov.br>

³³ <http://pt.wikipedia.org>

³⁴ <http://www.correios.com.br/>

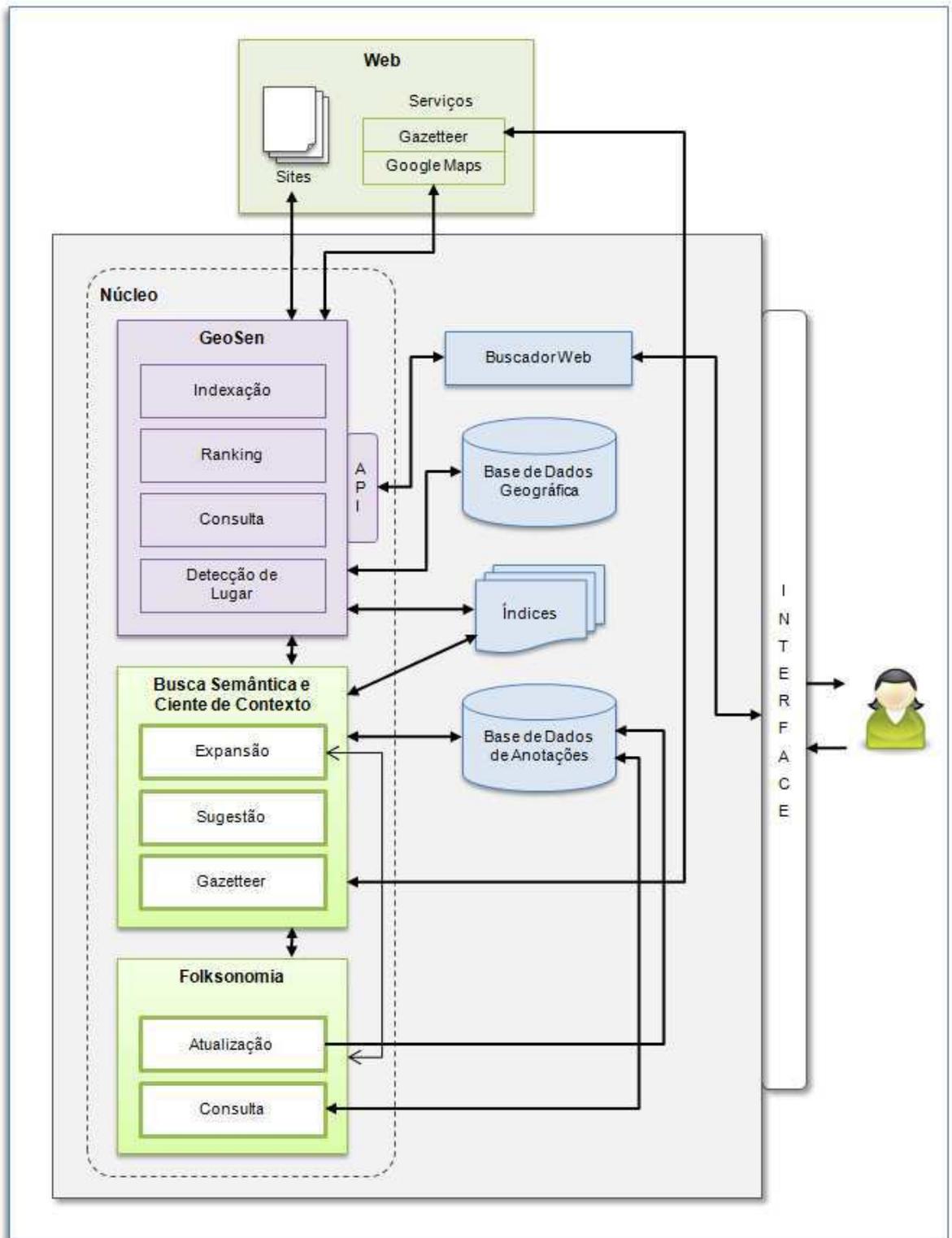


Figura 4.3 – Arquitetura detalhada do GeoSen_Tags.

Os módulos de *Consulta* e *Detecção de Lugar* foram aprimorados através do mecanismo de Busca Semântica e Ciente de Contexto, que possui os módulos: *expansão*, *sugestão* e *gazetteer*.

A expansão dos termos da consulta, por sua vez, faz uso do mecanismo de *folksonomia*, com o intuito de avaliar que outros termos podem ser utilizados para melhorar a precisão do resultado da busca. Esta *folksonomia* é armazenada na *base de dados de anotações* e é atualizada sempre que um usuário visualiza uma página que resultou de uma consulta.

Por fim, a sugestão de termos é feita com o uso de ontologias e a detecção de lugares através de gazetteers.

Nas próximas seções deste capítulo, é explicado com maiores detalhes o funcionamento dos principais módulos do protótipo desenvolvido. Dentre eles estão: expansão de consultas, sugestão de termos, utilização de gazetteers e construção da *folksonomia*.

4.3. Expansão de Consultas Baseada em Anotações

De modo geral, se um documento não contém as palavras informadas em uma consulta, então ele não irá ser encontrado pelo motor de busca. Sendo assim, o objetivo principal da expansão de consulta é reduzir a incompatibilidade dos termos informados pelo usuário com o documento, através da obtenção de palavras com significado similar e/ou correção de erros de grafia.

O algoritmo de expansão de consultas aqui proposto é composto de três etapas, conforme Figura 4.4.

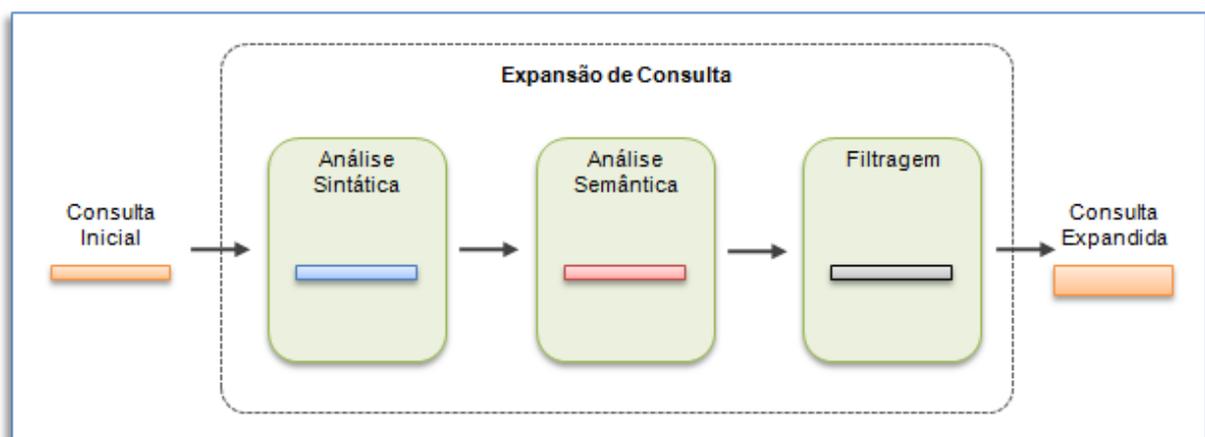


Figura 4.4 – Etapas do processo de expansão de consultas.

Na primeira, os termos da consulta são analisados sintaticamente, passando por um processo de correção de erros ortográficos e variação linguística, que gera radicais, plural, etc. Já na segunda etapa, é feita uma análise semântica das palavras resultantes do primeiro passo, de modo a se obter termos relacionados. Por fim, é feita uma filtragem da lista de palavras geradas pelo segundo passo, tendo como resultado, os termos que serão utilizados na consulta.

A correção de erros ortográficos é feita pela biblioteca SpellChecker³⁵ do Lucene³⁶, que usa a abordagem *letter n-grams* (Figura 4.5). Nesta, uma palavra pode ser dividida em uma seqüência de n letras. Por exemplo, a palavra *lucene* pode ser representada através de quatro *3-grams*: “luc”, “uce”, “cen”, e “ene”.

A idéia básica é que erros de grafia tipicamente afetam apenas alguns dos *n-grams*. Com isso, o termo correto pode ser localizado buscando-se termos de grafia corretas que compartilham uma alta proporção de *n-grams* com o termo incorreto.

Para computar essa similaridade é usado um índice invertido dos *n-grams* com as palavras. Essa técnica também pode sugerir variações da palavra original como o seu radical, plural, etc.

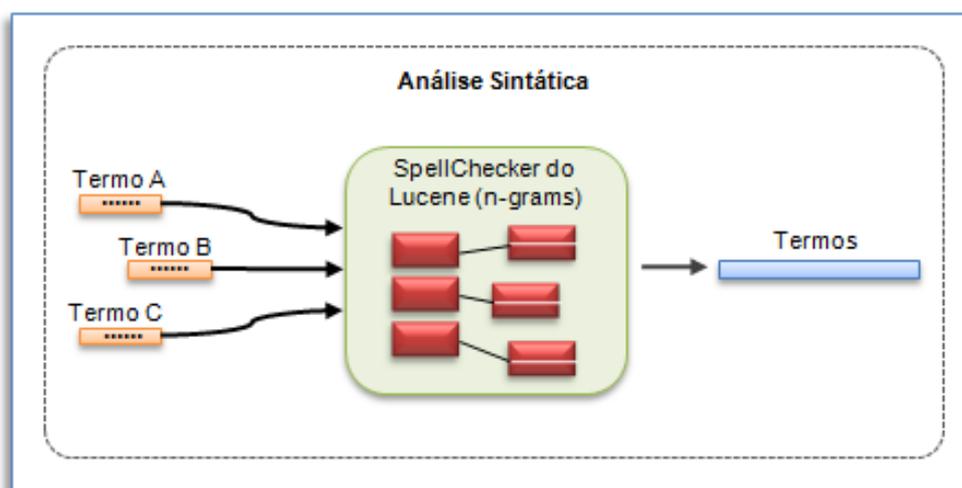


Figura 4.5 – Análise sintática dos termos, via *letter n-grams*.

Na análise semântica (Figura 4.6), a ideia básica utilizada é procurar anotações de documentos da Web que sejam similares aos termos que estão sendo analisados, além de

³⁵ <http://wiki.apache.org/lucene-java/SpellChecker>

³⁶ <http://lucene.apache.org/>

considerar o contexto geográfico do usuário. Esta procura é feita em uma folksonomia, cujo processo de construção será descrito com maiores detalhes na seção 4.4.

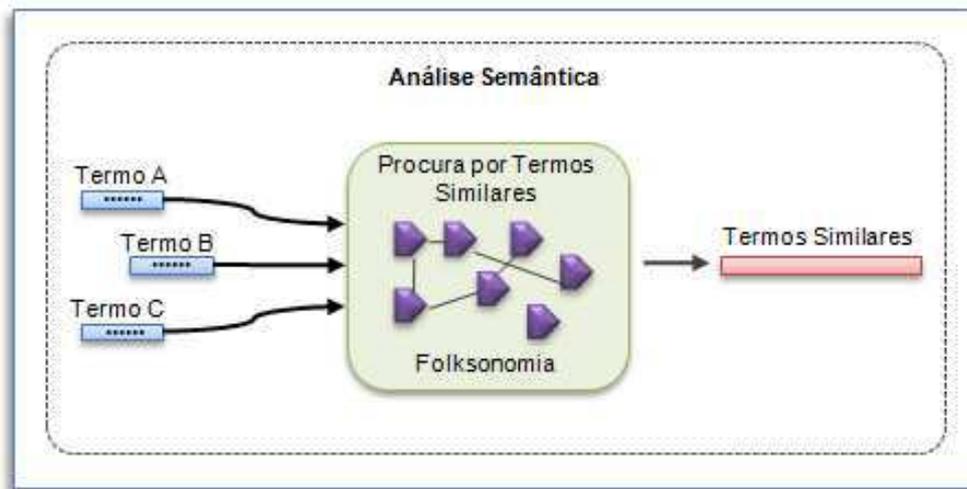


Figura 4.6 – Análise semântica dos termos.

No Código 4.1, encontra-se o algoritmo de busca de termos relacionados.

```

1  Função buscaTermosRelacionados (Lista Termos, Geometria ContextoUsuário) {
2      Lista Recursos;
3      para cada elemento I de Termos
4          Recursos.adiciona( buscaRecursos(I)
5      se ContextoUsuário for diferente nulo
6          Recursos ← filtragemGeográfica(Recursos, ContextoUsuário);
7      Lista Anotações;
8      para cada elemento J de Recursos
9          Anotações.adiciona(buscaAnotações(J));
10     retorna Anotações;
11 }
```

Código 4.1 – Algoritmo de busca de termos relacionados.

Inicialmente, é gerada uma lista com os recursos de cada termo. Diante dessa lista, são filtrados os documentos de acordo com o contexto geográfico do usuário, a não ser que

o mesmo não seja considerado. Por fim, é gerada uma lista com todas as anotações de cada documento da lista de recursos.

O algoritmo de filtragem geográfica dos documentos é mostrado no Código 4.2.

```
1  Função filtragemGeográfica (Lista Recursos, Geometria ContextoUsuário) {  
2      Lista RecursosFiltrados;  
3      para cada elemento I de Recursos  
4          se ContextoUsuário.contém(I.geometria)  
5              RecursosFiltrados.adiciona(I);  
6      retorna RecursosFiltrados;  
7  }
```

Código 4.2 – Algoritmo de filtragem geográfica dos documentos.

Cada recurso Web possui uma identidade geográfica, que é sua coordenada de latitude e longitude. Diante desta informação, uma consulta espacial é feita para determinar quais documentos fazem parte do contexto geográfico do usuário (Figura 4.7).

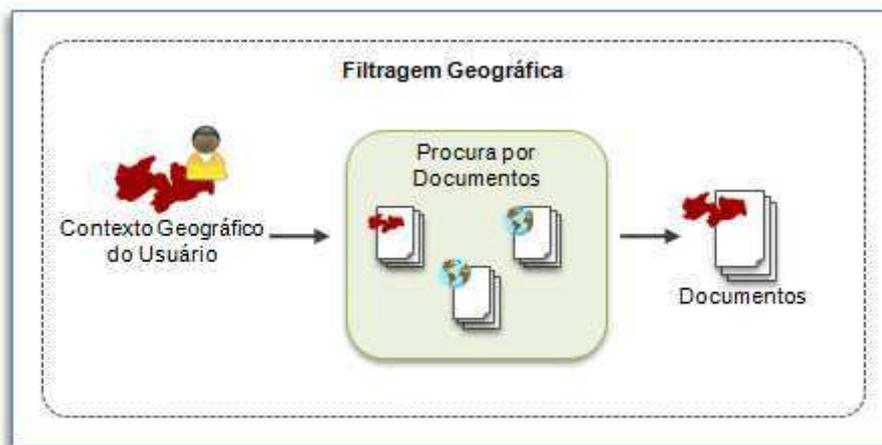


Figura 4.7 – Filtragem espacial dos documentos.

A última etapa do algoritmo de expansão compreende a escolha das anotações mais populares, ou seja, aquelas que são mais utilizadas para descrever os recursos. Mesmo assim, ainda é preciso definir um limite, visto que o número de termos gerados pode ser grande, podendo piorar o resultado da consulta. Desta forma, o número máximo de termos a serem expandidos é um parâmetro configurável no sistema.

4.4. Construção da Folksonomia

Folksonomia é a atribuição de rótulos a objetos, representando então um sistema de classificação (Figura 4.8). Uma grande vantagem de se usar a informação proveniente da folksonomia é que ela pode aumentar a precisão de consultas, uma vez que várias pessoas podem atribuir diversas anotações a um mesmo recurso.

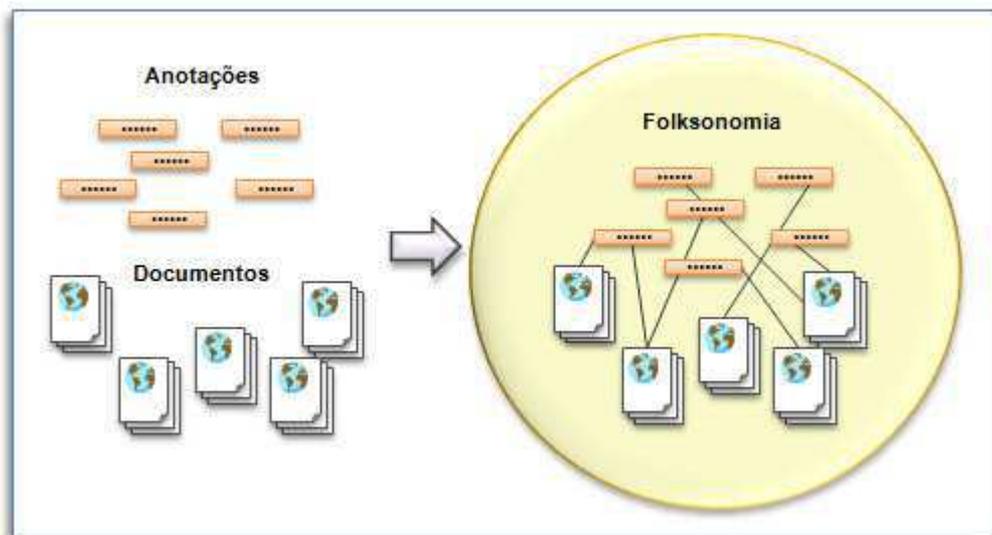


Figura 4.8 – Folksonomia de documentos.

Na Figura 4.9, é mostrado o modelo conceitual da folksonomia adotada no GeoSenTags.

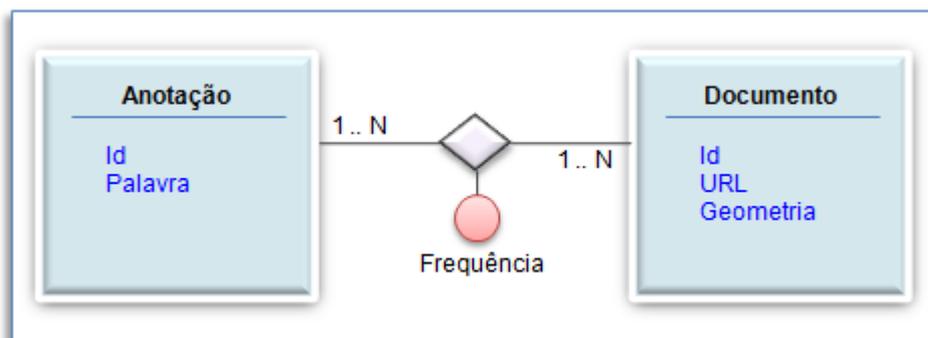


Figura 4.9 – Modelo conceitual da Folksonomia.

Existem duas entidades: a anotação e o documento. Estas se relacionam de modo que uma anotação pode ser feita para vários documentos e um documento pode possuir diversas anotações.

Para armazenar as informações foi criada uma base de anotações, via o SGBD PostgreSQL. A atualização desta base é feita automaticamente, através do feedback indireto do usuário, e o seu algoritmo pode ser visto no Código 4.3.

```

1  Procedimento atualizaFolksonomia (Lista Anotações_Recursos) {
2      Anotação A;
3      Recurso R;
4      Frequência F;
5      para cada elemento I de Anotações_Recursos
6           $A \leftarrow I.$ anotação;
7           $R \leftarrow I.$ recurso;
8          se não existe(A)
9              insereAnotação(A);
10         se não existe(R)
11             insereRecurso(R);
12         senão
13             atualizaRecurso(R);
14         se não existe(I)
15              $F \leftarrow 1$ ;
16             insereAnotaçãoRecurso(I,F);
17         senão
18              $F \leftarrow I.$ frequência + 1;
19             insereAnotaçãoRecurso(I,F);
20     }
```

Código 4.3 – Algoritmo de atualização da Folksonomia.

Sempre que um usuário realiza uma consulta e visualiza um documento resultante da mesma, é verificado se os termos informados por ele já existem na tabela de anotações. Caso não existam, uma nova entrada na tabela é criada.

O passo seguinte é verificar se o documento visualizado já se encontra na tabela de documentos. Caso ele já esteja inserido é feita apenas uma atualização no seu contexto geográfico, senão, é criada uma nova entrada na tabela de documentos.

Por fim, no terceiro passo, é verificada a existência do relacionamento entre a anotação e o documento. Se relacionamento já existir a frequência do mesmo é incrementada de 1. Caso contrário, é criada uma nova entrada com frequência igual a 1.

4.5. Sugestão de Termos

No GeoSen_Tags, é feito uso de ontologias para refinar as consultas dos usuários, sugerindo termos caso o resultado da consulta inicial não seja satisfatório. Na Figura 4.10, é descrito o processo de sugestão de termos, através de ontologias, no GeoSen_Tags.

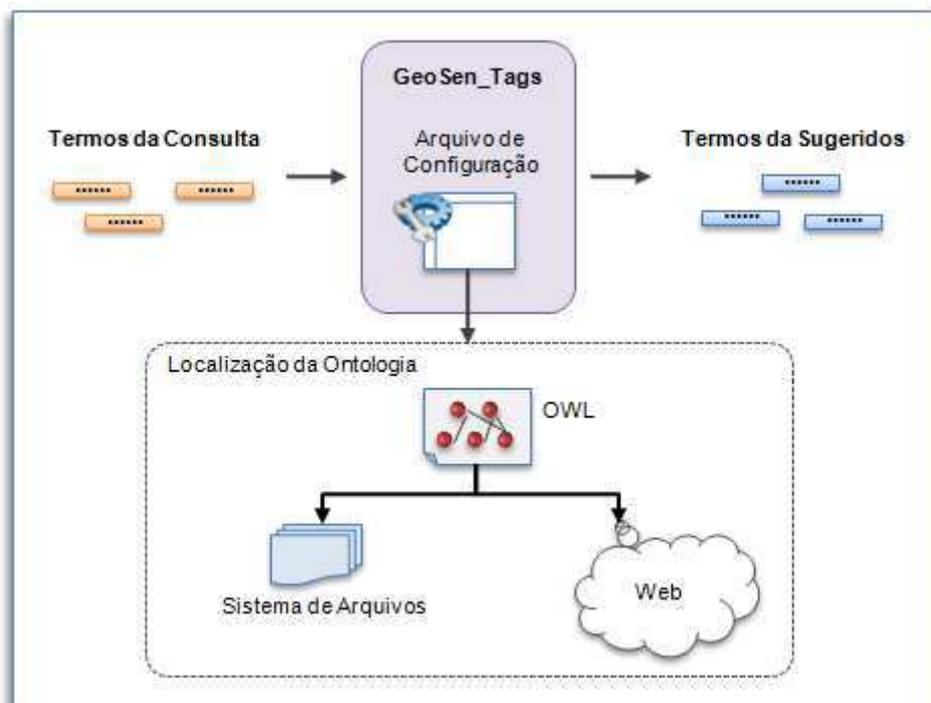


Figura 4.10 – Processo de sugestão de termos.

Assim que o sistema é inicializado, arquivos no formato OWL (Web Ontology Language) são lidos, processados e seus conteúdos ficam disponíveis para serem consultados. Estes arquivos podem conter descrições de classes e suas respectivas

propriedades e seus relacionamentos. Na Figura 4.11, é apresentado um exemplo de arquivo OWL.

```
<?xml version="1.0" encoding="UTF-8"?>
<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:owl="http://www.w3.org/2002/07/owl#"
  xmlns:xsd="http://www.w3.org/2001/XMLSchema#"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#">
  <owl:Ontology rdf:about="">
    <owl:versionInfo>
      1.0</owl:versionInfo>
    </owl:Ontology>
    <owl:Class rdf:about="http://www.w3.org/2002/07/owl#Thing"/>
    <owl:Class rdf:ID="EntidadeTemporal"/>
    <owl:Class rdf:ID="ObjetoTemporal">
      <rdfs:subClassof rdf:resource="#EntidadeTemporal"/>
    </owl:Class>
    <owl:Class rdf:ID="Duracao">
      <rdfs:subClassof rdf:resource="#ObjetoTemporal"/>
    </owl:Class>
    <owl:Class rdf:ID="Estacao">
      <rdfs:subClassof rdf:resource="#Duracao"/>
    </owl:Class>
    <owl:Class rdf:ID="Primavera">
      <rdfs:subClassof rdf:resource="#Estacao"/>
    </owl:Class>
    <owl:Class rdf:ID="Verao">
      <rdfs:subClassof rdf:resource="#Estacao"/>
    </owl:Class>
    <owl:Class rdf:ID="Outono">
      <rdfs:subClassof rdf:resource="#Estacao"/>
    </owl:Class>
    <owl:Class rdf:ID="Inverno">
      <rdfs:subClassof rdf:resource="#Estacao"/>
    </owl:Class>
  </rdf:RDF>
```

Figura 4.11 – Exemplo de arquivo OWL.

Para que o sistema possa localizar um arquivo OWL, que pode estar em um repositório de arquivos ou na Web, é preciso informar seu endereço de acesso em um arquivo de configuração.

Com o arquivo OWL disponível no sistema, é possível navegar na estrutura da ontologia. Dessa forma, sempre que o usuário faz uma consulta, os termos informados por ele são pesquisados na ontologia para que termos relacionados (superclasses, subclasses,

classes equivalentes, sinônimos, etc) sejam encontrados. Essa pesquisa é feita através de uma biblioteca denominada JENA³⁷, que é um framework JAVA para o desenvolvimento de aplicações da Web Semântica.

Por fim, alguns dos termos encontrados são exibidos no resultado da consulta como sugestão para novas buscas. A escolha desses termos é feita por um filtro, conforme Figura 4.12.

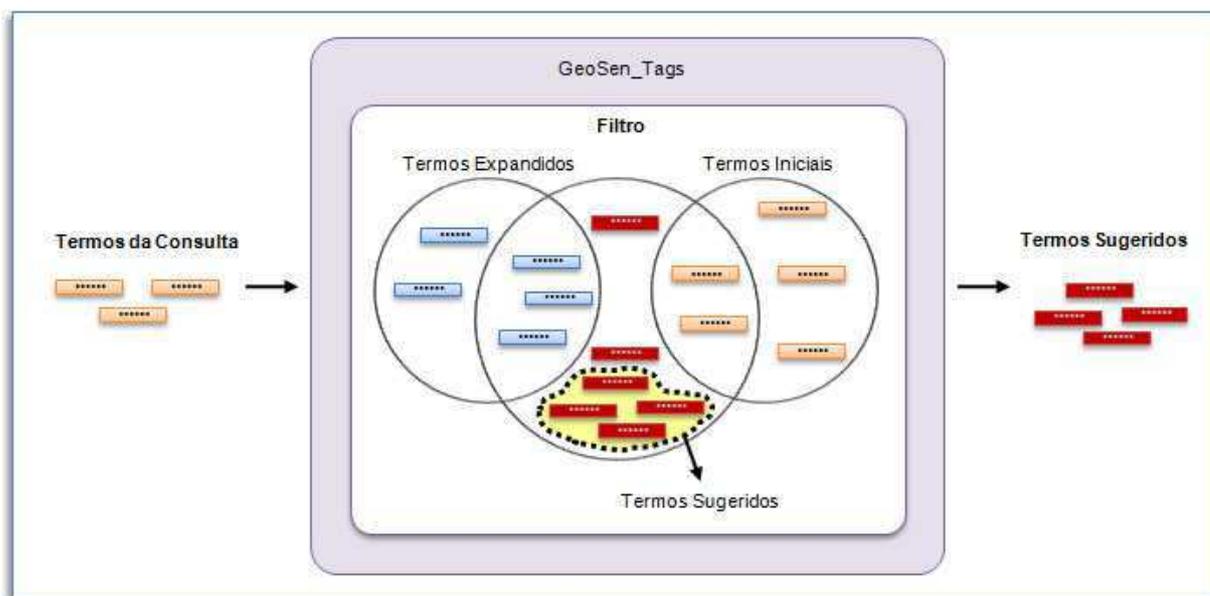


Figura 4.12 – Filtragem dos termos sugeridos.

Dentre os termos encontrados na ontologia, N são selecionados, onde N é um parâmetro configurável.

4.6. Gazetteers

No GeoSen_Tags, é feito o uso de gazetteers para aprimorar o algoritmo de detecção de lugares do GeoSen, no momento da consulta, já que este não localiza parques, prédios, pontos turísticos, etc.

O GeoNames³⁸ é um exemplo de gazetteer bastante utilizado, visto que pode ser acessado via serviços Web, de modo que a localização dos lugares podem ser encontradas

³⁷ <http://jena.sourceforge.net>

³⁸ <http://www.geonames.org>

não só através de seus nomes mas também por códigos postais, proximidades, dentre outras. Além deste, também existe o Google Geocoding API³⁹, que recebe como entrada um endereço e retorna como saída uma coordenada geográfica.

Conforme a Figura 4.13, é possível perceber que se o usuário deseja fazer uma consulta por “atividade física” que esteja relacionada ao lugar “Parque Ibirapuera”, o GeoSen não consegue detectar a palavra “Parque Ibirapuera” como sendo um lugar geográfico. Todavia, com o auxílio de gazetteers, o GeoSen_Tags é capaz de identificar que a palavra “Parque Ibirapuera” é um nome de lugar e que tal lugar pertence a cidade de São Paulo, que por sua vez pertence ao estado de São Paulo.



Figura 4.13 – Comparativo GeoSen X GeoSen_Tags.

³⁹ <http://code.google.com/intl/pt-BR/apis/maps/documentation/geocoding>

Dentre os gazetteers pesquisados, a escolha pelo GeoNames se deu pelo fato deste possuir uma vasta base de dados de informações geográficas, além de disponibilizar uma biblioteca JAVA para acesso as informações.

Contudo, o Google Geocoding também está sendo empregado no trabalho, visto que, ao realizar algumas buscas por informações sobre lugares localizados no Brasil, o GeoNames não obteve sucesso.

Desta forma, sempre que o GeoNames não encontra o lugar informado pelo usuário, é feita uma nova tentativa utilizando o Google Geocoding.

Na Figura 4.14, é apresentado o processo de detecção de lugares.

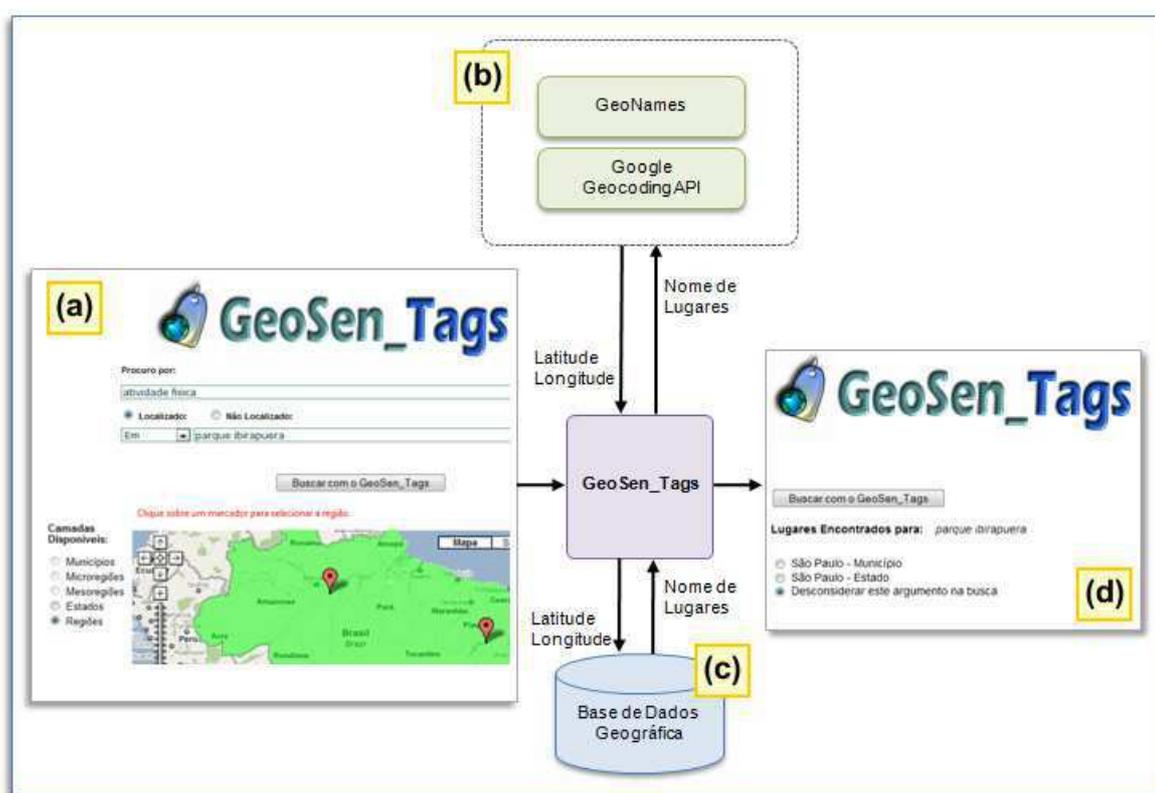


Figura 4.14 – Processo detecção de lugares. (a) Consulta do usuário. (b) Consulta ao GeoNames e Geocoding. (c) Consulta a base de dados geográfica. (d) Lista de lugares.

Sempre que o usuário faz uma consulta informando um nome de lugar para filtrar os resultados (Figura 4.14 (a)), o GeoNames, em conjunto com o Google Geocoding, é utilizado para encontrar a coordenada geográfica do respectivo lugar (Figura 4.14 (b)).

De posse da localização geográfica, o sistema procura na base de dados geográfica que cidades, regiões, países, contém a coordenada do lugar pesquisado (Figura 4.14 (c)).

Por fim, é sugerido para o usuário escolher, dentre os encontrados, qual lugar ele deseja obter informações a respeito (Figura 4.14 (d)).

Um trecho do algoritmo, na linguagem JAVA, para busca de lugares geográficos, através do GeoNames, pode ser visto no Código 4.4.

```
1    ...
2    List results = new ArrayList();
3    ToponymSearchCriteria searchCriteria = new ToponymSearchCriteria();
3    searchCriteria.setQ("Parque Ibirapuera");
4    ToponymSearchResult searchResult = WebService.search(searchCriteria);
5    for (Toponym toponym : searchResult.getToponyms()) {
6        results.add(toponym.getLatitude()+" "+ toponym.getLongitude());
7    }
8    ...
```

Código 4.4 – Trecho de algoritmo JAVA de busca de lugares, através do GeoNames.

Na linha 2, é instanciado um objeto que realiza buscas em uma estrutura topológica, de modo que na linha seguinte é passado o termo da consulta que se deseja fazer. Na linha 4, um objeto de resultados é inicializado com a resposta de um serviço Web que procura o lugar geográfico na base de dados do GeoNames. Por fim, nas linhas de 5 a 7, é feita uma varredura nos resultados de modo que apenas as coordenadas de latitude e longitude são recuperadas.

No Código 4.5, é apresentado o algoritmo, na linguagem JAVASCRIPT, para busca de locais através do Geocoding.

```
1    ...
2    function getGeocodingLatLong(placeName) {
3        geocoder.getLatLng( placeName, function(point) {
4            if (!point) {
5                document.mainForm.placeLatitude.value = point.lat();
6                document.mainForm.placeLongitude.value = point.lng();
7            } else
8                document.mainForm.placeLatitude.value = " ";
9                document.mainForm.placeLongitude.value = " ";
```

```
10         }));  
11     }  
12     ...
```

Código 4.5 – Trecho de algoritmo JAVASCRIPT de busca de lugares, através do Geocoding.

Na Linha 3, é utilizado o objeto *geocoder*, fornecido pela API do Geocoding, para buscar as coordenadas de latitude e longitude de um nome de lugar passado como parâmetro para a função. Como resposta, é retornado um objeto *point*, que também é fornecido pela API do Geocoding, caso o lugar pesquisado seja encontrado, senão é retornado um objeto vazio. Sendo assim, de posse do objeto retornado, é verificado se ele é um ponto ou não, para que valores de latitude e longitude sejam repassados para a lógica do sistema.

4.7. Considerações Finais

Neste capítulo foi apresentado o GeoSen_Tags, um motor de busca geográfico com suporte a anotações. Ele estende outra máquina de busca existente, denominada GeoSen, e foi desenvolvido com o intuito de validar as pesquisas realizadas neste trabalho.

Inicialmente sua arquitetura foi descrita, de maneira geral e detalhada, juntamente com as tecnologias empregadas no desenvolvimento do mesmo.

Por fim, foi explicado, em detalhes, o funcionamento dos módulos do mecanismo de busca semântico e ciente de contexto e da folksonomia de documentos utilizada.

O capítulo a seguir relata um estudo de caso realizado com o sistema apresentado.

Capítulo 5

Estudo de Caso

Neste capítulo, serão apresentados três estudos de caso, com o intuito de validar a aplicabilidade e as contribuições do mecanismo de busca aqui proposto. Todas as funcionalidades implementadas no protótipo serão comparadas com as do GeoSen, o motor de busca escolhido para ser estendido com as novas abordagens descritas neste trabalho.

Inicialmente, são apresentadas algumas observações sobre a preparação do protótipo para a realização dos estudos de caso. Em seguida, o primeiro estudo trata da melhoria na detecção de lugares através do uso de gazetteers. No segundo, é apresentada a sugestão de termos relacionados via ontologias de domínio. No terceiro estudo de caso, é analisado o método de expansão de consultas baseado em anotações juntamente com a folksonomia de documentos gerada. Por fim, são feitas considerações sobre os estudos de caso relatados.

5.1. Ambiente do Estudo de Caso

Para a execução dos estudos de caso foram selecionados, manualmente, diversos sites da Web. A seleção dos mesmos baseou-se em seus respectivos conteúdos, para que fossem obtidas informações variadas sobre diferentes áreas do saber.

As URLs dos sites escolhidos serviram como semente para o *Crawler* varrer a Internet em busca de novas páginas. Ao final deste processo, foram indexados 44.351 documentos, pertencentes a classes de notícias, turismo, governos, universidades, arte e blogs diversos.

Para validar o método de sugestão de termos relacionados, foi criada a ontologia `directoryOntology.owl`, cuja estrutura de classes se baseia nos diretórios do *Yahoo Directory*⁴⁰. Nesta, é possível encontrar diversos temas, como por exemplo, universidades, viagens, eleições e estações do ano. A ontologia completa é apresentada no Apêndice.

5.2. Estudo de Caso 1: Detecção de Lugares através de Gazetteers

A interface de busca do GeoSen disponibiliza um campo textual onde o usuário pode informar qual lugar ele deseja obter informações a respeito. Contudo, tal sistema não consegue detectar locais tais como museus, parques, estádios de futebol, ginásios poliesportivos, praças, dentre outros.

Como solução, o GeoSen_Tags utiliza gazetteers para procurar a coordenada geográfica do lugar informado pelo usuário, e, de posse dela, realizar uma busca na base de dados geográfica do sistema para determinar a lista de lugares a ser sugerida para o usuário escolher para efetuar a desambiguação da consulta geográfica.

Suponha que o usuário deseja obter informações sobre o *São João* da cidade de *Campina Grande*, que tem como palco principal de atrações o *Parque do Povo*. Para isso, o GeoSen_Tags disponibiliza uma interface que permite ao usuário fazer consultas do tipo “*o que/onde*”, conforme Figura 5.1.

⁴⁰ <http://dir.yahoo.com/>



Figura 5.1 – Interface de consultas do GeoSen_Tags.

Inicialmente, ele deve preencher o campo “**Procuo por**” com os termos “*sao joao*”, marcar a opção “**Localizado**”, escolher o operador espacial, que no exemplo acima foi “**Em**”, informar o lugar como sendo “*parque do povo*” e clicar no botão “**Buscar com o GeoSen_Tags**”.

Feito isso, o GeoSen_Tags recebe o nome “*parque do povo*” e tenta recuperar as coordenadas deste lugar, utilizando primeiramente o GeoNames, e se, não obtiver sucesso, o Google Geocoding.

No caso do lugar “parque do povo”, a procura no GeoNames falha, uma vez que esta informação não está presente na base de dados deste Gazetteer. Logo, a busca parte para o Geocoding, que consegue inferir as coordenadas de latitude e longitude do parque do povo como sendo, -7.2242 e -35.887624, respectivamente.

De posse dessa informação, é feita uma consulta espacial na base de dados geográfica, para determinar a lista lugares para desambiguação. No Código 5.1, é exibido o comando SQL executado para busca dos lugares.

```
1 select name
2 from geoscopecontroltb
3 where Contains(geom, GeomFromText('POINT(<longitude> <latitude>')));
```

Código 5.1 – Comando SQL para busca de lugares na base de dados geográfica.

Logo após, é exibida uma tela (Figura 5.2) com uma lista de lugares relacionados ao nome “parque do povo”.



Figura 5.2 – Tela de desambiguação da consulta espacial.

Dentre alguns dos nomes sugeridos estão: *Paraíba*, *Campina Grande*, *Nordeste*, *Agreste Paraibano*, etc. Com isso, é possível perceber que as localidades informadas para a desambiguação possuem uma relação espacial com o nome “parque do povo”, já que este se encontra na cidade de *Campina Grande*, que pertence ao estado da *Paraíba*, que por sua vez está inserido na região *Nordeste* do Brasil.

Caso esta mesma consulta seja submetida ao GeoSen, nenhum lugar é encontrado, como pode ser visto na Figura 5.3.



Figura 5.3 – Tela de desambiguação da consulta espacial apresentada pelo GeoSen para o lugar “*parque do povo*”.

Suponha agora que o usuário deseja obter informações sobre *turismo* no estádio do *Maracanã*. Para isso, ele preenche o campo “**Procuo por**” com o termo “*turismo*”, marca a opção “**Localizado**”, escolhe o operador espacial “**Em**” e informa o lugar como sendo “*maracana*”, conforme Figura 5.4.



Figura 5.4 – Consulta por “turismo” no “maracana”, na tela de consultas do GeoSen_Tags.

Após clicar no botão “Buscar com o GeoSen_Tags”, a lista com as localidades sugeridas (Figura 5.5) apresenta nomes como: município do *Rio de Janeiro*, estado do *Rio de Janeiro*, região *Sudeste*, município do *Maracanã*, estado do *Pará*, região *Norte*, etc.



Figura 5.5 – Tela de desambiguação para a consulta pelo lugar “maracana” no GeoSen_Tags.

Sendo assim, percebe-se que como o usuário não especificou o lugar como sendo *estádio* do *Maracanã*, então também foi sugerido um município do estado do *Pará*, que também tem o nome de *Maracanã*, o que justifica a importância desse processo de desambiguação.

Na Figura 5.6, é mostrada a tela de resultados para a consulta “*turismo*” no lugar “*maracanã*” e com a escolha do município do *Rio de Janeiro*, na desambiguação da consulta espacial.



Figura 5.6 – Tela de resultados para a consulta por “turismo” no lugar “maracana”.

5.3. Estudo de Caso 2: Sugestão de termos relacionados

O GeoSen_Tags pode ser configurado para utilizar a estrutura semântica das ontologias para sugerir termos relacionados a uma determinada consulta. Neste estudo de caso, a ontologia `directoryOntology.owl` foi configurada para ser utilizada como base semântica para inferência de termos relacionados. Esta foi modelada para representar alguns dos temas e subtemas abordados no Yahoo Directory, de modo que uma classe da ontologia representa um tema, uma subclasse representa um subtema e a propriedade `rdfs:seeAlso` é utilizada para indicar classes relacionadas.

No GeoSen_Tags, quando o usuário fizer a consulta “verão” em “joao pessoa” (Figura 5.7), o termo “verão” é submetido a pesquisa de termos relacionados na ontologia `directoryOntology.owl`. Nesta, são buscadas as subclasses e as propriedades `rdfs:seeAlso`

relacionadas ao termo “*verão*”. Como resultado, obtém-se: “*estacoes do ano*”, “*inverno*”, “*outono*”, “*primavera*”, “*clima*” e “*temperatura*”, conforme Figura 5.8.

Na lista final de sugestões de pesquisas relacionadas, não são utilizados todos os termos recuperados da ontologia. Uma filtragem é feita, de modo que, os termos que também fazem parte da consulta inicial do usuário e os resultantes da expansão da consulta, não são considerados. Além disso, é usado um parâmetro configurável, que determina o limite de termos a serem sugeridos. Para o estudo de caso tal parâmetro foi definido como sendo 4.

Na Figura 5.9, é exibida a tela de resultados da consulta, onde no final da mesma são listados os termos relacionados a palavra “*verão*”. Dentre eles estão: “*estações do ano*”, “*inverno*”, “*outono*” e “*primavera*”.



Figura 5.7 – Consulta por “*verão*” no lugar “*joao pessoa*”.

```
<owl:Class rdf:about= "#Verao ">
  <rdfs:seeAlso rdf:resource= "#Estacoes_do_ano "/>
  <rdfs:seeAlso rdf:resource= "#Inverno "/>
  <rdfs:seeAlso rdf:resource= "#Outono "/>
  <rdfs:seeAlso rdf:resource= "#Primavera "/>
  <rdfs:seeAlso rdf:resource= "#Clima "/>
  <rdfs:seeAlso rdf:resource= "#Temperatura "/>
  <rdfs:subClassOf rdf:resource= "#Estacoes_do_ano "/>
</owl:Class>
```

Figura 5.8 – Trecho da ontologia que representa as informações da classe Verão.



The screenshot shows the GeoSen_Tags search interface. At the top left is a logo featuring a globe and a tag. The main title 'GeoSen_Tags' is displayed in a large, blue, stylized font. Below the title, a green bar indicates the search results: 'Resultados da Pesquisa: verao (Foram encontrados aproximadamente 4 resultados)'. The search results are listed below, each with a title, a snippet, a URL, and links for 'Explain', 'Geo Explain', and 'Cache'. The results include: 'Fest Verão 2010 João Pessoa/Paraíba', 'João Pessoa (PB) lança programação do verão 2010', 'Estação Nordeste - Prefeitura Municipal de João Pessoa', and 'Programação Cultural Verão 2010'. At the bottom, a green bar labeled 'Pesquisas Relacionadas:' lists related search terms: 'Inverno', 'Outono', 'Primavera', and 'Estações do ano'.

Figura 5.9 – Tela de resultados para a consulta por “*verão*” no lugar “*joao pessoa*”.

Caso o usuário escolha a pesquisa relacionada “*estações do ano*”, será exibida a página de resultados com as seguintes pesquisas relacionadas: “*clima*”, “*temperatura*”, “*inverno*” e “*primavera*” (Figura 5.10). O trecho da ontologia que justifica esta escolha é mostrado na Figura 5.11.



GeoSen_Tags

Resultados da Pesquisa: Estações do ano (Foram encontrados aproximadamente 5 resultados).

Festival de Flores de Holambra começa nesta quinta
 ... plantas que florescem em todas as **estações do ano**...
<http://www.joaopessoa.pb.gov.br/noticias/?n=14499>
[Explain](#) [Geo Explain](#) [Cache](#)

Paraibanews.com » Estação Cabo Branco inaugura 1ª etapa do Museu
 ... Um experimento que possui ajuste azimutal para latitude e **estações do ano**...
<http://www.paraibanews.com/2010/07/02/estacao-cabo-branco-inaugura-1%C2%AA-etapa-do-museu-da-ciencia/>
[Explain](#) [Geo Explain](#) [Cache](#)

Funes e APAPB promovem Ano Internacional da Astronomia em João Pessoa
 ... Palestra – **Estações do ano**, eclipses e fases da Lua...
<http://www.portalcorreo.com.br/entretenimento/matier.asp?newsId=66212>
[Explain](#) [Geo Explain](#) [Cache](#)

Festival de flores de Holambra começa nesta quinta - PS Online
 ... enfatizar o inverno, serão comercializadas plantas que florescem em todas as **estações do ano**...
http://www.psonlinebr.com/portal/?p=noticia_interna&id=85171
[Explain](#) [Geo Explain](#) [Cache](#)

João Pessoa, Brasil Tempo : Weather Underground
 ... Média do tempo nas **estações do ano**...
<http://portuguese.wunderground.com/global/stations/82798.html>
[Explain](#) [Geo Explain](#) [Cache](#)

Pesquisas Relacionadas:

- Clima
- Temperatura
- Inverno
- Primavera

Figura 5.10 – Tela de resultados para a consulta “estações do ano”.

```

<owl:Class rdf:ID= "Estacoes_do_ano ">
  <rdfs:seeAlso>
    <owl:Class rdf:ID= "Temperatura "/>
  </rdfs:seeAlso>
  <rdfs:seeAlso>
    <owl:Class rdf:ID= "Clima "/>
  </rdfs:seeAlso>
</owl:Class>
...
<owl:Class rdf:ID= "Outono ">
  <rdfs:subClassOf rdf:resource= "#Estacoes_do_ano "/>
  ...
</owl:Class>
<owl:Class rdf:ID= "Inverno ">
  <rdfs:subClassOf rdf:resource= "#Estacoes_do_ano "/>
  ...
</owl:Class>
<owl:Class rdf:ID= "Primavera ">
  <rdfs:subClassOf rdf:resource= "#Estacoes_do_ano "/>
  ...
</owl:Class>
<owl:Class rdf:ID= "Verao ">
  <rdfs:subClassOf rdf:resource= "#Estacoes_do_ano "/>
  ...
</owl:Class>

```

Figura 5.11 – Trecho da ontologia que representa as informações relacionadas a classe Estações do ano.

5.4. Estudo de Caso 3: Expansão de Consultas e Folksonomia

O usuário, muitas vezes, não consegue expressar corretamente o que ele deseja consultar, e acaba obtendo resultados com muitos documentos que não atendem aos seus anseios. No GeoSen, não é feita expansão de consultas nem análise sintática ou semântica dos termos. Por outro lado, o GeoSen_Tags emprega tais métodos.

O método semântico utiliza uma folksonomia de documentos, criada automaticamente, a partir das consultas realizadas pelos usuários e leva em consideração o contexto geográfico deles, com o intuito de fornecer aos mesmos resultados de consultas com recursos mais relevantes.

Sendo assim, neste estudo de caso, será mostrada a criação da folksonomia de documentos e sua utilização na expansão de consultas (método semântico), juntamente com a análise sintática. Para isso, serão analisadas quatro consultas submetidas por três usuários diferentes, de modo que dois deles possuem o mesmo contexto geográfico, conforme Figura 5.12.

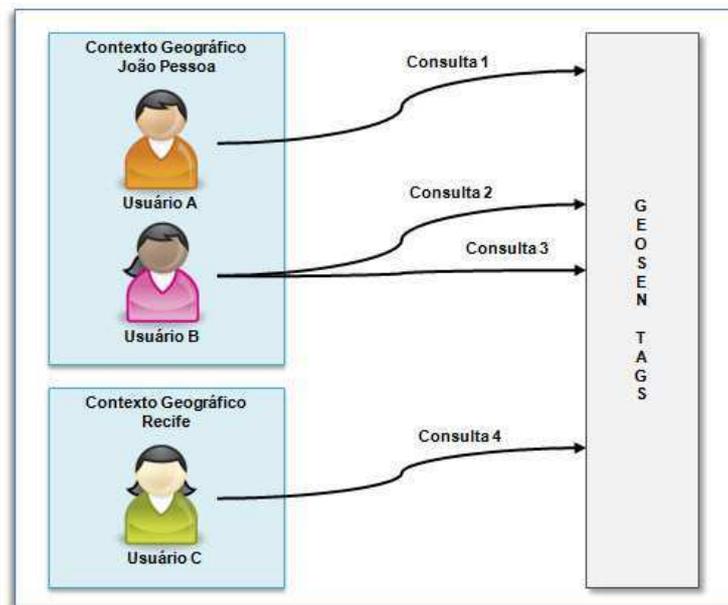


Figura 5.12 – Esquema utilizado no estudo de caso de expansão de consultas e folksonomia.

Inicialmente, o sistema não possui nenhuma anotação associada a recursos, visto que a construção da folksonomia é feita através do feedback indireto do usuário.

Desta forma, na primeira consulta submetida ao sistema, no método de expansão de consultas é realizada apenas a análise sintática, que consiste na correção de grafia e variação dos termos, como determinação de plural e radical.

A **consulta 1** executada pelo **usuário A** é mostrada na Tabela 5.1. Observe que, na expansão, o termo “*agencia*” foi corrigido para “*agência*”. A sugestão de termos para correção depende do dicionário de palavras utilizado no SpellChecker. O sistema pode facilmente incorporar novas palavras para serem utilizadas. Se houver mais de uma sugestão proveniente da análise sintática dos termos, todas as sugestões são enviadas para a expansão de consultas.

Usuário	A
Consulta	agencia turismo
Expansão	agência turismo

Tabela 5.1 – Consulta realizada pelo usuário A e sua expansão.

O resultado desta consulta pode ser visto na Figura 5.13, onde foram encontradas 19 páginas Web. A mesma consulta submetida ao GeoSen não teve a correção ortográfica da palavra “*agencia*”, desta forma, nenhum resultado foi retornado.



Resultados da Pesquisa: [agencia turismo](#) (Foram encontrados aproximadamente 19 resultados).

Agência de Viagens | Turismo de Aventura e Ecoturismo | Adventure Club (a)
Agência de Viagens | Turismo de Aventura e Ecoturismo | Adventure ... ir para o conteúdo Home **Agência** Nacionais Internacionais Mural do Cliente ...
<http://www.adventureclub.com.br/>
[Explain](#) [Geo Explain](#) [Cache](#)

Agência de Viagens | Turismo de Aventura e Ecoturismo | Adventure Club (b)
Agência de Viagens | Turismo de Aventura e Ecoturismo | Adventure ... ir para o conteúdo Home **Agência** Nacionais Internacionais Mural do Cliente ...
<http://www.adventureclub.com.br/default.asp>
[Explain](#) [Geo Explain](#) [Cache](#)

Agência Ecoturismo Brasil ? Viagens, turismo de aventura e expedições. (c)
Agência Ecoturismo Brasil – Viagens, turismo de aventura e expedições. WWW ...
<http://www.ecoturismobrasil.com.br/>
[Explain](#) [Geo Explain](#) [Cache](#)

TRAVESSIA ECOTURISMO - Agência Receptivo e Operadora de Turismo de Aventura na Chapada dos Veadeiros
 ... TRAVESSIA ECOTURISMO - **Agência** Receptivo e Operadora de Turismo ...
<http://www.travessia.tur.br/>
[Explain](#) [Geo Explain](#) [Cache](#)

Travessia Ecoturismo Chapada dos Veadeiros Agência Receptivo e Operadora de Turismo de Aventura Goiás
 ... Travessia Ecoturismo Chapada dos Veadeiros **Agência** Receptivo e Operadora de Turismo ...
<http://www.travessia.tur.br/pt-br/>
[Explain](#) [Geo Explain](#) [Cache](#)

Agência Ar Bonito MS e Pantanal
 ... viajar com a **Agência Ar** e uma **Agência** de turismo de pacotes e grupos? ALUGUEL ... Experiência" direto do Ministério do ...
<http://www.agenciaar.com.br/>
[Explain](#) [Geo Explain](#) [Cache](#)

Turismo Ecológico
 ... **Turismo** De Aventura - **Agência** De **Turismo** **Agência** Serra do Cipó **Turismo** 3486 6771 Serra do Cipó ... do Brasil: ... **Agência** ...
<http://www.holfrog.com.br/Produtos/Turismo-Ecol-gico>

TRAVESSIA ECOTURISMO - Agência Receptivo e Operadora de Turismo de Aventura na Chapada dos Veadeiros
 ... TRAVESSIA ECOTURISMO - **Agência** Receptivo e Operadora de Turismo ...
<http://www.travessia.tur.br/>
[Explain](#) [Geo Explain](#) [Cache](#)

Travessia Ecoturismo Chapada dos Veadeiros Agência Receptivo e Operadora de Turismo de Aventura Goiás
 ... Travessia Ecoturismo Chapada dos Veadeiros **Agência** Receptivo e Operadora de Turismo ...
<http://www.travessia.tur.br/pt-br/>
[Explain](#) [Geo Explain](#) [Cache](#)

Agência Ar Bonito MS e Pantanal
 ... viajar com a **Agência Ar** e uma **Agência** de turismo de pacotes e grupos? ALUGUEL ... Experiência" direto do Ministério do ...
<http://www.agenciaar.com.br/>
[Explain](#) [Geo Explain](#) [Cache](#)

Turismo Ecológico
 ... **Turismo** De Aventura - **Agência** De **Turismo** **Agência** Serra do Cipó **Turismo** 3486 6771 Serra do Cipó ... do Brasil: ... **Agência** ...
<http://www.holfrog.com.br/Produtos/Turismo-Ecol-gico>
[Explain](#) [Geo Explain](#) [Cache](#)

TURISMO ECOLÓGICO
 ... as opções de roteiros de **turismo** ecológico da **agência** UAI TRIP CAMINHADAS ECOLÓGICAS / CACHOEIRAS ... 00 Horas Baixo Incluso: Traslado **agência** ...
<http://www.tiradentesgerais.com.br/ecologico.htm>
[Explain](#) [Geo Explain](#) [Cache](#)

Revista Turismo - Parati - RJ
 ... Revista **Turismo** - Parati - RJ Parati - RJ - Fev ... a forte presença de **turismo** na região atraídos pela beleza
<http://www.revistaturismo.com.br/Ecoturismo/parati.html>
[Explain](#) [Geo Explain](#) [Cache](#)

Pantanal Ecoturismo, turismo no Pantanal e em Bonito - MS - Brasil
 ... em Bonito Mergulho Portal de **Turismo** de Bonito Licenciamento Ambiental Meio ... e Pantanal Viagem Iguassu Falls **Agência Turismo** ...
<http://www.pantanalecoturismo.tur.br/>
[Explain](#) [Geo Explain](#) [Cache](#)

[Primeira](#) [1](#) [2](#) [Próxima](#) [Última](#)

Figura 5.13 – Resultado da consulta 1 executada pelo usuário A.

A construção da folksonomia é iniciada a partir do momento que o usuário clica para visualizar algum dos resultados. No exemplo, o **usuário A** clicou nos links (a), (b), e (c), mostrados na Figura 5.13. Ao realizar esta ação o sistema associa os termos “*agência*” e “*turismo*”, referentes a consulta do usuário ao contexto do mesmo e ao documento clicado.

Com isso, indiretamente ele atribuiu anotações como “*agência*” e “*turismo*” aos documentos escolhidos.

Na segunda consulta realizada pelo **usuário B**, já existem termos na folksonomia, desta forma, ela será utilizada, em conjunto com a análise sintática, para realizar a expansão. Sendo assim temos na Tabela 5.2 a consulta do usuário B e sua expansão.

Usuário	B
Consulta	turismo
Expansão	agência turismo

Tabela 5.2 – Consulta realizada pelo usuário B e sua expansão.

Como pode ser visto, após uma consulta na folksonomia, a busca foi expandida com o termo “*agência*”, uma vez que este está relacionado ao termo “*turismo*”, já que eles co-ocorrem no mesmo documento. Além disso, tanto o **usuário A**, que utilizou os termos “*agência turismo*”, quanto o **usuário B**, que utilizou apenas o termo “*turismo*”, possuem o mesmo contexto geográfico. Na Figura 5.14, é mostrado o mapeamento da folksonomia gerada pela **consulta 1** do **usuário A**.

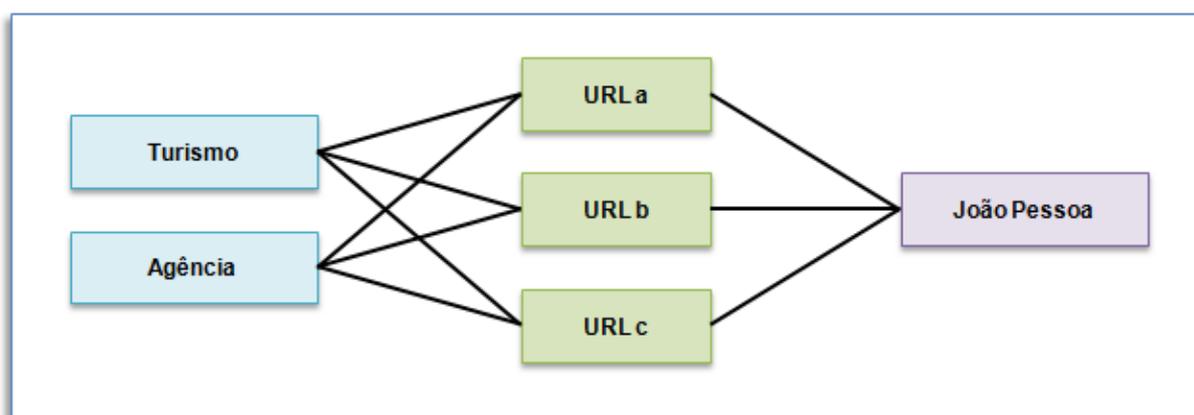


Figura 5.14 – Mapeamento da folksonomia gerada pela consulta 1 do usuário A.

Diante disto, o resultado da **consulta 2** feita pelo **usuário B** no GeoSen, retornaria muitos recursos, diferentemente do GeoSen_Tags, que acrescentou mais um termo a consulta, melhorando assim sua precisão.

Suponha agora que o **usuário B** efetuou a **consulta 3**, mostrada na Tabela 5.3.

Usuário	B
Consulta	eco turismo
Expansão	agência eco turismo

Tabela 5.3 – Consulta realizada pelo usuário B e sua expansão.

Após análise na folksonomia, o algoritmo de expansão acrescenta o termo “*agência*”, uma vez que ele co-ocorre com o termo “*turismo*” em um mesmo documento e o **usuário B** possui o mesmo contexto geográfico do **usuário A**, que anotou o recurso via feedback indireto, ao visitar as páginas dos resultados da **consulta 1**. O resultado da **consulta 3** é mostrado na Figura 5.15.



GeoSen_Tags

Resultados da Pesquisa: eco turismo (Foram encontrados aproximadamente 6 resultados).

Agência Ar Bonito MS e Pantanal
 ... Viajar com a Agência Ar e uma Agência de turismo de pacotes e grupos? ALUGUEL ... Experiência* direto do Ministério do ...
<http://www.agenciaar.com.br/>
[Explain](#) [Geo Explain](#) [Cache](#)

Turismo Ecológico
 ... Turismo De Aventura - Agência De Turismo Agência Serra do Cipó Turismo 3486 6771 Serra do Cipó ... do Brasil... Agência ...
<http://www.hotfrog.com.br/Produtos/Turismo-Ecol-gico>
[Explain](#) [Geo Explain](#) [Cache](#)

NATUREZA TERRA BRASIL ESCANDINÁVIA Excursões, Ecoturismo, Pacotes, Noruega, Finlândia, Suécia, Turismo Eco
 ... Polinésia, Consultoria e Assessoria em Turismo, Viagens, Europa, Pacotes, Agência Petrópolis, Rio de Janeiro, Brasil ... br Consultoria e Assessoria em ...
<http://www.naturezaterra.com.br/>
[Explain](#) [Geo Explain](#) [Cache](#)

Pantanal Ecoturismo, turismo no Pantanal e em Bonito - MS - Brasil
 ... em Bonito Mergulho Portal de Turismo de Bonito Licenciamento Ambiental Meio ... e Pantanal Viagem Iguassu Falls Agência Turismo ...
<http://www.pantanaletoturismo.tur.br/>
[Explain](#) [Geo Explain](#) [Cache](#)

BuscaCerta.com » Esportes » Esportes Radicais
 ... José dos Campos, SP. Andrenalina Turismo Aventura Turismo ecológico, aventura, rappel, trekking, grutas ... e outros. Atitude - Ecoturismo & Turismo ...
<http://www.buscacerta.com/site/279.php>
[Explain](#) [Geo Explain](#) [Cache](#)

Países::Brasil::Esportes::Radicais
 ... <http://www.abutre.com.br/Agência> volta ao parque Agência de turismo da Chapada Diamantina. (BR) <http://www.achei.com.br/01020314/028/radicais.html>
[Explain](#) [Geo Explain](#) [Cache](#)

Figura 5.15 – Resultado da consulta 3 do usuário B.

Por fim, caso o **usuário C** queira fazer a **consulta 4**, descrita na Tabela 5.4, percebe-se que não haverá expansão da mesma porque o contexto geográfico do **usuário C** é diferente do contexto do **usuário A**, que foi quem colaborou na construção da

folksonomia. Sendo assim, o resultado da consulta, com 23 páginas, é apresentado na Figura 5.16.

Usuário	C
Consulta	eco turismo
Expansão	-

Tabela 5.4 – Consulta realizada pelo usuário C.

GeoSen_Tags

Resultados da Pesquisa: eco turismo (Foram encontrados aproximadamente 23 resultados)

Patagonia Turismo de aventura y ecoturismo Argentina y Chile ? Sudamérica | Patagonia viajes organizados
 ... Cómo reservar | F.A.Q. | Site Map **ECO-FAMILY** - Ecoturismo - Viajes - Argentina - Chile ... Sudamérica en Argentina y Chile. **Eco-Family** ofrece viajes organizados de
<http://www.eco-family.com/espanol/>
[Explain](#) [Geo Explain](#) [Cache](#)

NATUREZA TERRA BRASIL ESCANDINÁVIA Excursões, Ecoturismo, Pacotes, Noruega, Finlândia, Suécia, Turismo Eco
 ... Poinésia, Consultoria e Assessoria em **Turismo**, Viagens, Europa, Pacotes, Agência Petrópolis ... Janeiro, Brasil-Home Natureza Terra **Turismo** Ecológico e Educativo Um novo ...
<http://www.naturezaterria.com.br/>
[Explain](#) [Geo Explain](#) [Cache](#)

Turismo Ecológico
 ... localização e pousadas **eco turismo** - **Turismo** De Aventura - **turismo** aventura - Bóia Cavernas (15) 3556 ... de Santana (ornamentações), Água Suja (... **eco** ...
<http://www.hotllog.com.br/Produtos/Turismo-Ecol-gico>
[Explain](#) [Geo Explain](#) [Cache](#)

Cia Eco - Viaje sempre em boa companhia
 ... **Cia Eco** - Viaje sempre em boa companhia ... Planejamento de Viagem Ecoturismo e **Turismo** de Aventura Portadores de Necessidades ...
<http://www.ciaeco.tur.br/>
[Explain](#) [Geo Explain](#) [Cache](#)

Bienvenidos a EcoTurismo Yucatán - Excursiones de Ecolog'a Maya - YUCATAN - MEXICO Eco Tours
 ... a Maya - YUCATAN - MEXICO **Eco Tours** Libro de Visitas Excursiones ... referencias. Bienvenidos los profesionales del **turismo** EcoTurismo Yucatán es miembro de ...
<http://www.ecoyuc.com/spanish.html>
[Explain](#) [Geo Explain](#) [Cache](#)

Ecoturismo | Turismo de Aventura | NaturaMexico
 ... Ecoturismo | **Turismo** de Aventura | NaturaMexico :::: Visitanos y ... Trip Mexico Mexcape Agencia de **Turismo** Alternativo ▶ Actividades -- Elige una Actividad
<http://www.naturamexico.com/>
[Explain](#) [Geo Explain](#) [Cache](#)

Turismo de Aventura e Esportes Radicais
 ... **Turismo** de Aventura , Ministério do Turismo ... Aventura - Ecoturismo ▶ - Esportes na Neve ▶ - **Eco-Resorts** ▶ - Bungee Jumping ▶ - Ciclismo ▶ - Mergulho ... igual. Bungee jumping, esporte radical.
<http://www.buscateamatica.net/aventura.htm>

Figura 5.16 – Resultado da consulta 4 do usuário C.

5.5. Considerações Finais

Através dos estudos de caso apresentados neste capítulo, foi mostrada uma visão geral das funcionalidades implementadas no protótipo GeoSen_Tags. Além disso, foi possível comprovar a eficácia das abordagens propostas na elaboração do mecanismo de busca semântico e ciente de contexto, quando comparadas com as do GeoSen.

No próximo capítulo, serão apresentadas as conclusões desta dissertação, além de alguns trabalhos futuros que poderão ser realizados para a continuidade desta pesquisa.

Capítulo 6

Conclusão

Estudos recentes na área de RI apontam para o emprego de novas abordagens na melhoria dos mecanismos de busca para Web. Dentre alguns dos aspectos considerados estão o uso da personalização e o contexto do usuário, uma vez que este possui um papel fundamental neste cenário.

Além disso, o rápido crescimento das redes sociais e anotações de recursos permitem que o conceito de folksonomia também seja empregado nos motores de busca, visto que as informações de anotações atribuídas a recursos podem ser de grande valia, já que os usuários foram bem treinados utilizando a Internet ao longo dos anos.

Outro fator importante que pode ser considerado é a localização geográfica, uma vez que ela é intrínseca a diversos conteúdos presentes na Web.

Diante disto, o principal objetivo deste trabalho foi a criação de um motor de busca semântico e ciente de contexto, aplicado em um Sistema de Recuperação de Informação Geográfica existente: o GeoSen.

O restante deste capítulo está dividido da seguinte maneira: na seção 6.1, são apresentadas as contribuições do trabalho aqui proposto e, na seção 6.2, são listadas algumas proposições para trabalhos futuros.

6.1. Contribuições

No capítulo 3, foram apresentados alguns trabalhos relacionados na área de RI, que tratam da melhoria dos mecanismos de busca através de anotações, folksonomias e ontologias. Além disso, foram elencadas oito características fundamentais para um mecanismo de busca semântico e ciente de contexto.

Diante disto, foi possível concluir que, os trabalhos citados focam em alguma técnica, enquanto que o GeoSen_Tags faz uso em conjunto de diversas abordagens, propondo ainda um novo método de expansão de consultas baseado em anotações.

Como resultado do trabalho aqui descrito tem-se:

- **Protótipo GeoSen_Tags:** para validar as abordagens aqui apresentadas, o Sistema de Recuperação de Informação Geográfica, GeoSen, foi estendido com as novas funcionalidades propostas. Sendo assim, o GeoSen_Tags é um motor de busca com enfoque geográfico, enriquecido semanticamente e que considera o contexto do usuário;
- **Folksonomias:** construção de uma folksonomia de páginas da Web através do feedback indireto do usuário. Sempre que este faz uma consulta e visualiza um resultado, os termos informados para a busca são utilizados como anotações para as páginas visitadas.
- **Expansão de Consultas:** para melhorar a eficácia do mecanismo de busca, foi proposto um novo método de expansão de consultas baseado em anotações. De modo geral, o seu funcionamento envolve três etapas: análise sintática, análise semântica e filtragem.
- **Ontologias:** para enriquecer semanticamente o mecanismo de busca apresentado, o conceito de ontologia foi empregado. Elas são usadas para inferir pesquisas relacionadas aos termos iniciais da consulta do usuário. Os termos escolhidos devem obedecer duas restrições: não fazer parte nem do conjunto de termos da expansão nem da consulta inicial fornecida pelo usuário.
- **Gazetteers:** o algoritmo de detecção de lugares, nos termos da consulta, proposto pelo GeoSen, não reconhece nomes de lugares como: museus, estádios, parques, praças, dentre outros. Para isso, o GeoSen_Tags utiliza gazetteers, sendo possível recuperar as coordenadas de latitude e longitude

do lugar pesquisado, e, de posse dela, fazer uma consulta a base de dados geográfica do sistema para retornar uma lista de localidades, que devem ser desambiguadas via o feedback direto do usuário.

- **Contexto do Usuário:** para desenvolver a busca ciente de contexto, foi escolhida a localização geográfica como sendo a informação relevante da entidade usuário. Com isso, é formada uma rede onde usuários estão conectados através de suas anotações a recursos e de suas posições espaciais. Tal informação é usada para auxiliar no método de expansão de consultas.

6.2. Trabalhos Futuros

Para continuidade das pesquisas iniciadas neste trabalho, são enumeradas proposições de alguns trabalhos futuros:

- Desenvolver um processo de *crawling*, indexação e busca de imagens e vídeos;
- Manipular localizações geográficas a nível global, visto que, atualmente, só se tem informações sobre lugares do Brasil na base de dados geográficas.
- Melhorar o desempenho do método de inferência nas ontologias, uma vez que ainda não é possível fazer uso de ontologias com um grande volume de informações como, por exemplo, WordNet.
- Avaliar a utilização de informações pessoais (preferências musicais, esportivas, etc) do usuário, além do seu contexto geográfico, no processo de expansão da consulta.
- Analisar como o sistema se comporta com milhares de anotações na folksonomia dos documentos.
- Realizar uma avaliação experimental do método de expansão de consultas aqui proposto, através dos procedimentos adotados pela comunidade Internacional de RI nas Text Retrieval Conferences⁴¹ (TREC).

⁴¹ <http://trec.nist.gov>

Referências Bibliográficas

- [1] SHADBOLT, N.; BERNERS-LEE, T.; HALL, W. The Semantic Web Revisited. *IEEE Intelligent Systems* 21(3): 96-101 (2006).
- [2] BHOGAL, J.; MACFARLANE, A.; SMITH, P. A Review Of Ontology Based Query Expansion. *Inf. Process. Manage.* 43(4): 866-886 (2007).
- [3] GAMMA, E.; HELM, R.; JOHNSON, R.; VLISSIDES, J. *Design Patterns: Elements Of Reusable Object-Oriented Software*. Addison-Wesley Reading, MA, 1995.
- [4] XU, J.; CROFT, W.B. Query Expansion Using Local And Global Document Analysis. In: *Proceedings of the 19th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp.4-11, 1996, Zurich, Switzerland.
- [5] VOORHEES, E. M. Query Expansion Using Lexical-Semantic Relations. In: *Proceedings of the 17th annual international ACM SIGIR conference on Research and development in information retrieval*, pp.61-70, 1994, Dublin, Ireland.
- [6] YANBE, Y.; JATOWT, A.; NAKAMURA, S.; TANAKA, K. Can Social Bookmarking Enhance Search In The Web?. In: *Proceedings of the 7th ACM/IEEE-CS joint conference on Digital libraries*, pp. 107–116, 2007, Vancouver, BC, Canada.
- [7] BIANCALANA, C.; MICARELLI, A. Social Tagging In Query Expansion: A New Way For Personalized Web Search. In: *Proceedings of SocialCom-09 the 2009 IEEE International Conference on Social Computing 2009*, 29th - 31st August 2009, Vancouver, Canada.
- [8] HORROCKS, I.; PATEL-SCHNEIDER, P. F. Reducing Owl Entailment To Description Logic Satisfiability. In: *D. Fensel, K. Sycara, and J. Mylopoulos, editors, Proc. Of the 2003 International Semantic Web Conference (ISWC 2003)*, number 2870 in Lecture Notes in Computer Science, pages 17–29. Springer, 2003.
- [9] LASSILA, O.; SWICK, R. R. *Resource Description Framework (Rdf) Model And Syntax Specification*. W3C Recommendation, 1999. Disponível em: <http://www.w3.org/TR/1999/RECrdf-syntax-19990222>.
- [10] HILL, L.: *Building Georeferenced Collections Gazetteer Services*. Taxonomy Authority file Workshop, Washington, DC; June (1998).

-
- [11] BAEZA-YATES, R.; RIBEIRO-NETO, B. *Modern Informtion Retrieval*. ACM Press Book: New York, 1999.
- [12] MORRIS, M., HORVITZ, E. 2007. Searchtogether: An Interface For Collaborative Web Search. *In: Proceedings of the 20th annual ACM symposium on User interface software and technology*, 3-12, 2007, Newport, Rhode Island, USA.
- [13] JARVELIN, K.; KEKALAINEN, J.; NIEMI, T. *Expansiontool: Concept-Based Query Expansion And Construction*. *Information Retrieval*, 4(3/4):231–255, 2001.
- [14] CARPINETO, C.; MORI, R.; ROMANO, G.; BIGI, B. *An Information-Theoretic Approach To Automatic Query Expansion*. *ACM Transactions on Information Systems*, 19(1):1–27, 2001.
- [15] CUI, H.; WEN, J.; LI, M. *A Statistical Query Expansion Model Based On Query Logs*. *Journal of Software*, 14(9):1593–1599, 2003.
- [16] ATTAR, R.; FRAENKEL, A. S. *Local Feedback In Full-Text Retrieval Systems*. *Journal of the ACM*, 24(3):397–417, July 1977.
- [17] CAI, D.; RIJSBERGEN, C. J.; JOSE, J. M. Automatic Query Expansion Based On Divergence. *In: Proceedings of the tenth international conference on Information and knowledge management (CIKM-01)*, pages 419–426, 2001, Atlanta, Georgia, USA.
- [18] FU, G.; JONES, C. B.; ABDELMOTY, A. I. Ontology-Based Spatial Query Expansion In Information Retrieval. *In: Lecture Notes in Computer Science, Volume 3761, On the Move to Meaningful Internet Systems: ODBASE 2005*.
- [19] CARMAN, M.; BAILLIE, M.; CRESTANI, F. Tag Data And Personalized Information Retrieval. *In: Proceeding of the 2008 ACM workshop on Search in social media*, pages 27-34, 2008, Napa Valley, California, USA.
- [20] ZANARDI, V.; CAPRA, L. Social Ranking: Uncovering Relevant Content Using Tag-Based Recommender Systems. *In Proceedings of the 2008 ACM conference on Recommender systems*, pages 51-58, 2008, Lausanne, Switzerland.
- [21] BERTIER, M.; GUERRAOUI, R.; KERMARREC, A. Toward Personalized Query Expansion. *In: Proceedings of the Second ACM EuroSys Workshop on Social Network Systems*, pages 7-12, 2009, Nuremberg, Germany.
- [22] CHIRITA, P. A.; FIRAN, C. S.; NEJDL, W. Personalized Query Expansion For The Web. *In: Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 7–14, 2007, Amsterdam, The Netherlands.

- [23] CUCERZAN, S.; WHITE, R. W. Query Suggestion Based On User Landing Pages. In *Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 875–876, 2007, Amsterdam, The Netherlands.
- [24] WANG, J.; DAVISON, B. D. Explorations In Tag Suggestion And Query Expansion. In: *Proceeding of the 2008 ACM workshop on Search in social media*, October 30-30, 2008, Napa Valley, California, USA.
- [25] GRUBER, T. Ontology Of Folksonomy. A Mash-Up Of Apples And Oranges, In: *First On-Line Conference on Metadata and Semantics Research 2005*.
- [26] VANDER WAL, T. *Explaining And Showing Broad And Narrow Folksonomies*. Disponível em: <http://www.vanderwal.net/random/category.php?cat=153>.
- [27] XU, S.; BAO, S.; FEI, B.; SU, Z.; YU, Y. Exploring Folksonomy For Personalized Search. In: *Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval*, July 20-24, 2008, Singapore, Singapore.
- [28] DI MATTEO N.; PERONI S.; TAMBURINI F.; VITALI F. A Parametric Architecture For Tags Clustering In Folksonomic Search Engines. In *Proceedings of 9th International Conference on Intelligent Systems Design and Applications (ISDA) - IEEE*, 279-282, 2009, Pisa.
- [29] ANTONIOU G.; VAN HARMELEN F. *A Semantic Web Primer*. Cambridge, Mass.: MIT Press, p. 288, 2008.
- [30] BREITMAN K.; CASANOVA M. A.; TRUSZKOWSKI W. *Semantic Web. Concepts, Technologies And Applications*. London: Springer, p. 330, 2008.
- [31] WELLER, K. Folksonomies And Ontologies- Two New Players In Indexing And Knowledge Representation. In: *Proceedings of Online Information*, pp. 108-115, 2007, London, Great Britain.
- [32] WELLER K.; PETERS I. Reconsidering Relationships For Knowledge Representation. In: *Proceedings of I-Know '07, Graz*, 5–7 September, 501–504, 2007.
- [33] ANGELETOU, S.; SABOU, M.; MOTTA, E. Folksonomy Enrichment And Search. In: *Proceedings of the 6th European Semantic Web Conference on The Semantic Web: Research and Applications*, pp. 801-805, 2009, Heraklion, Crete, Greece.
- [34] GOODCHILD, M. *Distributed Geolibraries: Spatial Information Resources*. National Academy Press, USA, p. 132, 1999.

- [35] COHEN, S. B. *The Columbia Gazetteer Of The World*. Columbia University Press, 1998. Disponível em <http://www.columbiagazetteer.org/main/Home.html>
- [36] HILL, L.; QI, Z. Indirect Geospatial Referencing Through Place Names In The Digital Library: Alexandria Digital Library Experience With Developing And Implementing Gazetteers. In: *Proceedings of the American Society for information Science Annual Meeting*, pp. 57-69, November , 1999, Washington, USA.
- [37] HILL, L. Core Elements Of Digital Gazetteers: Placenames, Categories, And Footprints. In: *Proceedings of the 4th European Conference - ECDL*, pp.280-290, 2000.
- [38] HILL, L.; FREW, J.; ZHENG, Q. *Geographic Names: The Implementation Of A Gazetteer In A Georeferenced Digital Library*. D-Lib Magazine, January, 1999.
- [39] SCHLIEDER, C.; VÖGELE, T.; VISSER, U. Qualitative Spatial Representation For Information Retrieval By Gazetteers. In: *Proceedings of the International Conference on Spatial Information Theory: Foundations of Geographic Information Science – COSIT*, California: Springer , 2001.
- [40] FRAKES, W. B.; BAEZA-YATES, R. *Information Retrieval Data Structures & Algorithms*, Prentice Hall, 1992.
- [41] GEY, F. Models In Information Retrieval. In: *Proceedings of the 19th ACM Conference on Research and Development in Information Retrieval (SIGIR)*, 1992.
- [42] BIWAS, G.; BEZDEK, J.; MARQUES, M.; SUBRAMANIAN, V. *Knowledge-Assisted Document Retrieval: Ii. The Retrieval Process*. Journal of the American Society for Information Science (JASIS), Vol. 38, No 2, 1987.
- [43] BOOKESTEIN, A. *Fuzzy Re Quests: An Approach To Weighted Boolean Searches*. Journal of the American Society for Information Science (JASIS), Vol. 31, No 7, 1980.
- [44] KWOK, K. L. A Network Approach To Probabilistic Information Retrieval. *ACM Transactions on Information Systems*, Vol. 13, No 3, 1995.
- [45] SALTON, G.; BUCCKLEY, C. Term-Weighting Approaches In Automatic Retrieval. *Information Processing & Management*, Vol. 24, No 5, 1988.
- [46] VAN RIJSBERGEN, C. J. *Information Retrieval*. Butterworths, 2nd edition, 1979.
- [47] ROBERTSON, S. E; SPARK JONES, K. *Relevance Weighting Of Search Terms*. Journal of the American Society for Information Science (JASIS), Vol. 27, No 3, 1976.

- [48] CALLAN, J. P. Passage-Level Evidence In Document Retrieval. *In: Proceedings of the 19th ACM Conference on Research and Development in Information Retrieval (SIGIR)*, pp. 302-310, 1996, Dublin, Ireland.
- [49] SPACCAPIETRA, S.; CULLOT, N.; PARENT, C.; VANGENOT, C. On Spatial Ontologies. *In: Proceedings of the VI Brazilian Symposium on Geoinformatics*. Campos do Jordão, SP, Brasil, 2004.
- [50] ROCHE C. Ontology: Survey. *In: Proceedings of the 8th Symposium on Automated Systems Based on Human Skill and Knowledge*, 22-24, Goteborg, Sweden, 2003.
- [51] GRUBER, T. R. *A Translation Approach To Portable Ontology Specifications*. Technical Report KLS 92-7, Knowledge Systems Laboratory, Stanford University, Stanford, CA, 1992.
- [52] GRUBER, T. R. *Toward Principles For The Design Of Ontologies Used For Knowledge Sharing*. *International Journal of Human and Computer Studies*, 43 (5-6): 907-28, 1995.
- [53] KHAN, L. R. *Ontology-Based Information Selection*. Tese PhD, Faculty of the Graduate School, University of Southern Califórnia, 2000.
- [54] SWARTOUT, W.; TATE, A. Guest Editors' Introduction: Ontologies. *IEEE Intelligent Systems*, 14 (1): 18-19, 1999.
- [55] KARP, R.; CHAUDHRI, V.; THOMERE, J. *Xol: An Xml-Based Ontology Exchange Language*. 1999. Disponível em: <http://www.xml.com/pub/r/888>
- [56] HEFLIN, J.; HENDLER, J.; LUKE, S. SHOE. *A Knowledge Representation Language For Internet Application*. Technical Report CS-TR-4078 (UMIACS TR-99-71), 1999.
- [57] GENESERETH, M. R.; FIKES, R. E. *Knowledge Interchange Format*. Reference Manual, Logic Group Report Logic-92-1 Computer Science Department, Stanford University, Stanford, California, USA, 1992.
- [58] LASSILA, O.; SWICK, R. R. *Resource Description Framework (RDF) Model And Syntax Specification*. Disponível em: <http://www.w3.org/TR/1999/REC-rdf-syntax-19990222/>
- [59] FENSEL, D.; VAN HARMELEN, F.; HORROCKS, I.; MCGUINNESS, D.; PATEL-SCHNEIDER, P. OIL: An Ontology Infrastructure For The Semantic Web. *IEEE Intelligent Systems*, 16(2): 38-44, 2001.

- [60] HORROCKS, I.; VAN HARMELEN, F. *Reference Description of the DAML+OIL Ontology Markup Language*, 2000.
- [61] MCGUINNESS D. L.; VAN HARMELEN, F. *OWL Web Ontology Language Overview*. Disponível em <http://www.w3.org/TR/owl-features>.
- [62] CARDOSO, O. N. P. *Recuperação de Informação*. 6 p. Departamento de Ciência da Computação – Universidade Federal de Lavras, Lavras, 2002.
- [63] YONGGANG, Q.; FREI, H. P. Concept Based Query Expansion. *In: XVI Annual International Acm Conference On Research And Development In Information Retrieval*, 1993, Pittsburgh. Nova York: ACM Press, 1993. p. 160 - 169.
- [64] BILLERBECK, B.; ZOBEL, J. Techniques for Efficient Query Expansion. *In: Apostolico, A.; Melucci, M.. Proc. String Processing and Information Retrieval Symposium*. Padova: Springer, 2004. p. 30-42.
- [65] XU, J.; CROFT, W. B. Improving the Effectiveness of Information Retrieval with Local Context Analysis. *ACM Transactions On Information Systems, Amherst* v. 1, n. 18, p.79-112, 2000.
- [66] SALTON, G.; BUCKLEY, C. *Improving retrieval performance by relevance feedback*. Readings in information retrieval, São Francisco, Ca, 1997.
- [67] SILVA, F. C. C.; BLATTMANN, U. A Colaboração e a interação na Web 2.0. *In: Congresso Brasileiro De Biblioteconomia, Documentação E Ciência Da Informação*, 22, 2007, Brasília. Revista ACB: Biblioteconomia em Santa Catarina, Florianópolis, v.12, n.2, p. 191-215, jul./dez., 2007.
- [68] AQUINO, M. C. *A folksonomia como hipertexto potencializador de memória coletiva: um estudo dos links e das tags no de.licio.ou e no Flickr*. *Liinc em Revista*, v. 4, n. 2, set. 2008, p. 303-320.
- [69] CAMPELO, C.E.C; BAPTISTA, C. S; FERNANDES, R. M. *Spatial Search Engines* Mehdi Khosrow-Pour (Editor), *Encyclopedia of Information Science and Technology*, Second Edition, IGI Global Publishing, ISBN 978-1-60566-026-4, 2008.
- [70] CAMPELO, C.E.C; BAPTISTA, C. S. Geographic Scope Modeling for Web Documents. *In: Proceedings of Fifth Workshop on Geographic Information Retrieval, ACM 17th Conference on Information and Knowledge Management (CIKM 2008)*, Napa Valley, California October 26-30, 2008, ACM Press.
- [71] CAMPELO, C.E. *GeoSen, Um Motor de Busca com Enfoque Geográfico*. Dissertação de Mestrado, Universidade Federal de Campina Grande, 2008.

[72] CHRISTOPHER D. M.; RAGHAVAN, R.; SCHÜTZE, H. *Introduction to Information Retrieval*, Cambridge University Press. 2008.

[73] BORGES, K. A. V., LAENDER, A. H. F., MEDEIROS, C. B., SILVA, A. S., DAVIS JR., C. A., 2003, *The Web as a data source for spatial databases*, V Simpósio Brasileiro de Geoinformática - GeoInfo 2003, Campos do Jordão (SP).

[74] PERRY M., SHETH A., ARPINAR I. B., *Geospatial and Temporal Semantic Analytics*, To appear in Encyclopedia of Geoinformatics, Hassan A. Karimi (Ed), Idea-Group Inc., 2007.

[75] ARASU, A.; CHO, J.; GARCIA-MOLINA, H.; PAEPCKE, A.; RAGHAVAN, S. *Searching the web*. ACM Transactions on Internet Technology, v. 1, n. 1, p. 2-43, August. 2001.

[76] LARSON, R.R., *Geographical information retrieval and spatial browsing*. Geographical Information Systems and Libraries: Patrons, Maps, and Spatial Information. pp. 81-124. 1996.

[77] LANA-SERRANO S.; VILLENA-ROMÁN J.; GOÑI-MENOYO J. M.; *Miracle at GeoCLEF Query Parsing 2007: Extraction and Classification of Geographical Information*, 8th CLEF Workshop, September 2007.

[78] PAGE, L.; BRIN, S.; MOTWANI, R.; WINOGRAD, T. *The PageRank Citation Ranking: Bringing Order to the Web*. 1999. Disponível em: <http://dbpubs.stanford.edu/pub/1999-66>.

[79] GILES, C. L.; BOLLACKER, K.; LAWRENCE, S. CiteSeer: An automatic citation indexing system. In: *Proceedings of Digital Libraries 98 - The Third ACM Conference on Digital Libraries*, Pittsburgh, PA. ACM Press. 1998.

[80] BUYUKKOKTEN O.; CHO J.; GARCIA-MOLINA H.; GRAVANO L.; SHIVAKUMAR N. Exploiting geographical location information of web pages. In *Proceedings of the ACM SIGMOD Workshop on the Web and Databases, WebDB*, 1999.

[81] JONES, C. B., H. ALANI, TUDHOPE D., Geographical Information Retrieval with Ontologies of Place. In. *Proceedings of COSIT 2001 International Conference on Spatial Information Theory*, Morro Bay, CA, USA, September 19-23. D. R. Montello (eds.), Springer-Verlag: 322-335. 2001.

[82] BORGES K. A. V. *Uso de uma Ontologia de Lugar Urbano para Reconhecimento e Extração de Evidências Geo-espaciais na Web*. Belo Horizonte: Instituto de Ciências Exatas, 2006. 195p. (Tese, Doutorado em Ciências da Comunicação).

- [83] JONES. C. *Geographical Information Retrieval*. GeoInfo 2006. Disponível em: http://www.geoinfo.info/geoinfo2006/presentation/Christopher_Jones.ppt
- [84] MARTINS B.; SILVA M. J.; FREITAS S. E AFONSO A. P.; *Handling Locations in Search Engine Queries*, GIR-2006, the Workshop on Geographic Information Retrieval at SIGIR 2006, Agosto 2006.
- [85] SANTOS, D., CHAVES, M.S., The place of place in geographical IR. *In: Proc. of the 3rd Workshop on Geographic Information Retrieval, SIGIR'06*, Seattle, USA 5–8. 2006.
- [86] SALTON. G., *Automatic Text Processing: The Transformation, Analysis, and Retrieval of Information by Computer*. Addison-Wesley, Reading, Pennsylvania, 1989.
- [87] MCCURLEY, S.K. Geospatial mapping and navigation of the web. *In Proceedings of the Tenth International WWW Conference*, Hong Kong, 1-5 May, 221-229. 2001.
- [88] LI, Z., WANG, C., XIE, X., WANG X. AND MA, W.-Y., *Indexing implicit locations for geographical information retrieval*. In Workshop on Geographic Information Retrieval, SIGIR. Seattle, USA. 2006.
- [89] FONSECA, F., EGENHOFER, M., BORGES K. A. V., *Ontologias e Interoperabilidade Semântica entre SIGs*. In: II Workshop Brasileiro em Geoinformática - GeoInfo2000, São Paulo. 2000.
- [90] TOCHTERMANN, K., RIEKERT, W-F., WIEST, G., SEGELKE, J. MOHAUPTJAHR, B., Using Semantic, Geographical, and Temporal Relationships to Enhance Search and Retrieval in Digital Catalogs. *In: Peters, C., Thanos, C. (eds.): Research and Advanced Technology for Digital Libraries. Proceedings ECDL'97*. Springer-Verlag Berlin. 1997.
- [91] VIEIRA, V.; SOUZA, D.; SALGADO, A. C.; TEDESCO, P. *Uso e Representação de Contexto em Sistemas Computacionais*, Mini-curso apresentado no Simpósio de Fatores Humanos em Sistemas Computacionais (IHC 2006), Natal, Brasil, 2006.
- [92] WEISER, M. *The Computer for the 21st Century*. Scientific American, v. 265, n. 3, p.94-104, 1991, reprinted in IEEE Pervasive Computing, Jan-Mar. 2002, PP. 19-25.

- [93] SANTOS JÚNIOR, L. B.; GOULARTE, R.; MOREIRA, E. S.; FARIA, G. The Modeling of Structured Context-Aware Interactive Environments. *In: Transactions of the SDPS – Journal of Integrated Design and Process Sciences*, v. 5, n. 4, 2001, p. 77-93.
- [94] MORAN, T. O.; DOURISH, O. *Introduction to This Special Issue on Context-Aware Computing, Special Issue of Human-Computer Interaction*, Volume 16, 2001.
- [95] TRUONG, K. N.; ABOWD, G. D.; BROTHERTON, J. A. Who, What, When, Where, How: Design Issues of Capture Access Applications. *In: Ubicomp 2001: Ubiquitous Computing, Third International Conference*, Atlanta, Georgia, USA, 2001, PP. 209-224.
- [96] BAZIRE, M.; BRÉZILLON, O. Understanding Context Before Using It. *In: 5th International and Interdisciplinary Conference, CONTEXT-05*, v. LNAI 3554, PP.29-40, Springer Verlag, Paris, France.
- [97] DEY, A. K. *Understanding and Using Context*. In ACM Personal and Ubiquitous Computing Journal, v.5, n. 1, p. 4-7, 2001.
- [98] GOULARTE, R. *Personalização e adaptação de conteúdo baseadas em contexto para TV Interativa*. Tese de doutorado. Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos, 2003.
- [99] Borges, K. A. V., Laender, A. H. F., Medeiros, C. B., Davis Jr., C. A. Discovering Geographic Locations in Web Pages Using Urban Addresses. In: Workshop on Geographic Information Retrieval, CIKM 2007, 2007, Lisboa, Portugal. *In: Proceedings of the 4th Workshop on Geographic Information Retrieval*, 2007.
- [100] Delboni, T. M., Borges, K. A. V., Laender, A. H. F. Davis Jr., C. A. Semantic Expansion of Geographic Web Queries Based on Natural Language Positioning Expressions. *Transactions in GIS*, v. 11, p. 377-397, 2007.

Apêndice

Ontologia directoryOntology.owl

```
1      <?xml version= "1.0 "?>
2      <rdf:RDF
3          xmlns=
"http://www.dsc.ufcg.edu.br/~madeira/ontology/directoryOntology.owl# "
4          xmlns:rdf= "http://www.w3.org/1999/02/22-rdf-syntax-ns# "
5          xmlns:protege= "http://protege.stanford.edu/plugins/owl/protege#
"
6          xmlns:xsp= "http://www.owl-ontologies.com/2005/08/07/xsp.owl# "
7          xmlns:owl= "http://www.w3.org/2002/07/owl# "
8          xmlns:xsd= "http://www.w3.org/2001/XMLSchema# "
9          xmlns:swrl= "http://www.w3.org/2003/11/swrl# "
10         xmlns:swrlb= "http://www.w3.org/2003/11/swrlb# "
11         xmlns:rdfs= "http://www.w3.org/2000/01/rdf-schema# "
12         xml:base=
"http://www.dsc.ufcg.edu.br/~madeira/ontology/directoryOntology.owl ">
13     <owl:Ontology rdf:about= " "/>
14     <owl:Class rdf:ID= "Estacoes_do_ano ">
15         <rdfs:seeAlso>
16             <owl:Class rdf:ID= "Temperatura "/>
17         </rdfs:seeAlso>
18         <rdfs:seeAlso>
19             <owl:Class rdf:ID= "Clima "/>
20         </rdfs:seeAlso>
21     </owl:Class>
22     <owl:Class rdf:ID= "Conferencias ">
23         <rdfs:subClassOf>
24             <owl:Class rdf:ID= "Viagens "/>
25         </rdfs:subClassOf>
26     </owl:Class>
27     <owl:Class rdf:ID= "Esportes "/>
28     <owl:Class rdf:ID= "Organizacoes ">
29         <rdfs:subClassOf>
30             <owl:Class rdf:about= "#Viagens "/>
31         </rdfs:subClassOf>
32     </owl:Class>
33     <owl:Class rdf:ID= "Turismo_sustentavel ">
34         <rdfs:subClassOf>
35             <owl:Class rdf:ID= "Turismo "/>
36         </rdfs:subClassOf>
37     </owl:Class>
38     <owl:Class rdf:about= "#Viagens ">
39         <rdfs:seeAlso>
40             <owl:Class rdf:about= "#Turismo "/>
41         </rdfs:seeAlso>
42     </owl:Class>
43     <owl:Class rdf:ID= "Turismo_ecologico ">
44         <rdfs:seeAlso>
45             <owl:Class rdf:ID= "Eco_turismo "/>
46         </rdfs:seeAlso>
47         <rdfs:subClassOf>
48             <owl:Class rdf:about= "#Turismo "/>
49         </rdfs:subClassOf>
50     </owl:Class>
```

```
51 <owl:Class rdf:ID= "Transporte ">
52   <rdfs:subClassOf rdf:resource= "#Viagens "/>
53 </owl:Class>
54 <owl:Class rdf:ID= "Alojamento ">
55   <rdfs:subClassOf rdf:resource= "#Viagens "/>
56 </owl:Class>
57 <owl:Class rdf:ID= "Outono ">
58   <rdfs:subClassOf rdf:resource= "#Estacoes_do_ano "/>
59   <rdfs:seeAlso>
60     <owl:Class rdf:ID= "Verao "/>
61   </rdfs:seeAlso>
62   <rdfs:seeAlso>
63     <owl:Class rdf:ID= "Primavera "/>
64   </rdfs:seeAlso>
65   <rdfs:seeAlso>
66     <owl:Class rdf:ID= "Inverno "/>
67   </rdfs:seeAlso>
68   <rdfs:seeAlso rdf:resource= "#Estacoes_do_ano "/>
69   <rdfs:seeAlso rdf:resource= "#Clima "/>
70   <rdfs:seeAlso rdf:resource= "#Temperatura "/>
71 </owl:Class>
72 <owl:Class rdf:about= "#Inverno ">
73   <rdfs:subClassOf rdf:resource= "#Estacoes_do_ano "/>
74   <rdfs:seeAlso>
75     <owl:Class rdf:about= "#Verao "/>
76   </rdfs:seeAlso>
77   <rdfs:seeAlso>
78     <owl:Class rdf:about= "#Primavera "/>
79   </rdfs:seeAlso>
80   <rdfs:seeAlso rdf:resource= "#Estacoes_do_ano "/>
81   <rdfs:seeAlso rdf:resource= "#Outono "/>
82   <rdfs:seeAlso rdf:resource= "#Clima "/>
83   <rdfs:seeAlso rdf:resource= "#Temperatura "/>
84 </owl:Class>
85 <owl:Class rdf:about= "#Eco_turismo ">
86   <rdfs:seeAlso rdf:resource= "#Turismo_ecologico "/>
87   <rdfs:subClassOf>
88     <owl:Class rdf:about= "#Turismo "/>
89   </rdfs:subClassOf>
90 </owl:Class>
91 <owl:Class rdf:ID= "Bagagem ">
92   <rdfs:subClassOf rdf:resource= "#Viagens "/>
93 </owl:Class>
94 <owl:Class rdf:ID= "Jogos_de_tabuleiro ">
95   <rdfs:subClassOf>
96     <owl:Class rdf:ID= "Jogos "/>
97   </rdfs:subClassOf>
98 </owl:Class>
99 <owl:Class rdf:ID= "Hoteis ">
100   <rdfs:subClassOf rdf:resource= "#Viagens "/>
101 </owl:Class>
102 <owl:Class rdf:ID= "Agentes ">
103   <rdfs:subClassOf rdf:resource= "#Viagens "/>
104 </owl:Class>
105 <owl:Class rdf:about= "#Primavera ">
106   <rdfs:subClassOf rdf:resource= "#Estacoes_do_ano "/>
107   <rdfs:seeAlso>
108     <owl:Class rdf:about= "#Verao "/>
109   </rdfs:seeAlso>
110   <rdfs:seeAlso rdf:resource= "#Estacoes_do_ano "/>
111   <rdfs:seeAlso rdf:resource= "#Inverno "/>
112   <rdfs:seeAlso rdf:resource= "#Outono "/>
113   <rdfs:seeAlso rdf:resource= "#Clima "/>
114   <rdfs:seeAlso rdf:resource= "#Temperatura "/>
115 </owl:Class>
116 <owl:Class rdf:about= "#Verao ">
117   <rdfs:seeAlso rdf:resource= "#Estacoes_do_ano "/>
118   <rdfs:seeAlso rdf:resource= "#Inverno "/>
```

```
119     <rdfs:seeAlso rdf:resource= "#Outono "/>
120     <rdfs:seeAlso rdf:resource= "#Primavera "/>
121     <rdfs:seeAlso rdf:resource= "#Clima "/>
122     <rdfs:seeAlso rdf:resource= "#Temperatura "/>
123     <rdfs:subClassOf rdf:resource= "#Estacoes_do_ano "/>
124   </owl:Class>
125   <owl:Class rdf:ID= "Bolsas_e_malas ">
126     <rdfs:subClassOf rdf:resource= "#Viagens "/>
127   </owl:Class>
128   <owl:Class rdf:about= "#Turismo ">
129     <rdfs:seeAlso rdf:resource= "#Viagens "/>
130   </owl:Class>
131 <owl:Class rdf:ID= "Computadores_e_Internet ">
132 </owl:Class>
133 <owl:Class rdf:ID= "Comunicacoes_e_Redes ">
134 <rdfs:seeAlso rdf:resource= "#Computadores_e_Internet "/>
135 <rdfs:subClassOf rdf:resource= "#Computadores_e_Internet "/>
136 </owl:Class>
137 <owl:Class rdf:ID= "Internet ">
138 <rdfs:seeAlso rdf:resource= "#Computadores_e_Internet "/>
139 <rdfs:subClassOf rdf:resource= "#Computadores_e_Internet "/>
140 </owl:Class>
141 <owl:Class rdf:ID= "Seguranca_e_Criptografia ">
142 <rdfs:seeAlso rdf:resource= "#Computadores_e_Internet "/>
143 <rdfs:subClassOf rdf:resource= "#Computadores_e_Internet "/>
144 </owl:Class>
145 <owl:Class rdf:ID= "Formatos_de_dados ">
146 <rdfs:seeAlso rdf:resource= "#Computadores_e_Internet "/>
147 <rdfs:subClassOf rdf:resource= "#Computadores_e_Internet "/>
148 </owl:Class>
149 <owl:Class rdf:ID= "Hardware ">
150 <rdfs:seeAlso rdf:resource= "#Computadores_e_Internet "/>
151 <rdfs:subClassOf rdf:resource= "#Computadores_e_Internet "/>
152 </owl:Class>
153 <owl:Class rdf:ID= "Multimidia ">
154 <rdfs:seeAlso rdf:resource= "#Computadores_e_Internet "/>
155 <rdfs:subClassOf rdf:resource= "#Computadores_e_Internet "/>
156 </owl:Class>
157 <owl:Class rdf:ID= "Noticias_e_Media ">
158 <rdfs:seeAlso rdf:resource= "#Computadores_e_Internet "/>
159 <rdfs:subClassOf rdf:resource= "#Computadores_e_Internet "/>
160 </owl:Class>
161 <owl:Class rdf:ID= "Software ">
162 <rdfs:seeAlso rdf:resource= "#Computadores_e_Internet "/>
163 <rdfs:subClassOf rdf:resource= "#Computadores_e_Internet "/>
164 </owl:Class>
165 <owl:Class rdf:ID= "Programacao_e_Desenvolvimento ">
166 <rdfs:seeAlso rdf:resource= "#Computadores_e_Internet "/>
167 <rdfs:subClassOf rdf:resource= "#Computadores_e_Internet "/>
168 </owl:Class>
169 <owl:Class rdf:ID= "MPEG ">
170 <rdfs:seeAlso rdf:resource= "#Multimidia "/>
171 <rdfs:seeAlso rdf:resource= "#Formatos_de_dados "/>
172 <rdfs:subClassOf rdf:resource= "#Multimidia "/>
173 </owl:Class>
174 <owl:Class rdf:ID= "CD-ROM ">
175 <rdfs:seeAlso rdf:resource= "#Multimidia "/>
176 <rdfs:subClassOf rdf:resource= "#Multimidia "/>
177 </owl:Class>
178 <owl:Class rdf:ID= "Flash_and_Shockwave ">
179 <rdfs:seeAlso rdf:resource= "#Multimidia "/>
180 <rdfs:subClassOf rdf:resource= "#Multimidia "/>
181 </owl:Class>
182 <owl:Class rdf:ID= "QuickTime ">
183 <rdfs:seeAlso rdf:resource= "#Multimidia "/>
184 <rdfs:subClassOf rdf:resource= "#Multimidia "/>
185 </owl:Class>
186 <owl:Class rdf:ID= "Video ">
```

```
187 <rdfs:seeAlso rdf:resource= "#Multimedia "/>
188 <rdfs:subClassOf rdf:resource= "#Multimedia "/>
189 </owl:Class>
190 <owl:Class rdf:ID= "Revistas ">
191 <rdfs:seeAlso rdf:resource= "#Multimedia "/>
192 <rdfs:subClassOf rdf:resource= "#Multimedia "/>
193 </owl:Class>
194 <owl:Class rdf:ID= "Institutos ">
195 <rdfs:seeAlso rdf:resource= "#Multimedia "/>
196 <rdfs:subClassOf rdf:resource= "#Multimedia "/>
197 </owl:Class>
198 <owl:Class rdf:ID= "Microsoft_Silverlight ">
199 <rdfs:seeAlso rdf:resource= "#Multimedia "/>
200 <rdfs:subClassOf rdf:resource= "#Multimedia "/>
201 </owl:Class>
202 <owl:Class rdf:ID= "Hipermedia ">
203 <rdfs:seeAlso rdf:resource= "#Multimedia "/>
204 <rdfs:subClassOf rdf:resource= "#Multimedia "/>
205 </owl:Class>
206 <owl:Class rdf:ID= "MIME ">
207 <rdfs:seeAlso rdf:resource= "#Multimedia "/>
208 <rdfs:subClassOf rdf:resource= "#Multimedia "/>
209 </owl:Class>
210 <owl:Class rdf:ID= "Software ">
211 <rdfs:seeAlso rdf:resource= "#Multimedia "/>
212 <rdfs:subClassOf rdf:resource= "#Multimedia "/>
213 </owl:Class>
214 <owl:Class rdf:ID= "Programacao ">
215 <rdfs:seeAlso rdf:resource= "#Multimedia "/>
216 <rdfs:subClassOf rdf:resource= "#Multimedia "/>
217 </owl:Class>
218 <owl:Class rdf:ID= "Audio ">
219 <rdfs:seeAlso rdf:resource= "#Multimedia "/>
220 <rdfs:subClassOf rdf:resource= "#Multimedia "/>
221 </owl:Class>
222 <owl:Class rdf:ID= "Conferencias ">
223 <rdfs:seeAlso rdf:resource= "#Multimedia "/>
224 <rdfs:subClassOf rdf:resource= "#Multimedia "/>
225 </owl:Class>
226 <owl:Class rdf:ID= "Graficos ">
227 <rdfs:seeAlso rdf:resource= "#Multimedia "/>
228 <rdfs:subClassOf rdf:resource= "#Multimedia "/>
229 </owl:Class>
230 <owl:Class rdf:ID= "Formatos_de_datos ">
231 <rdfs:seeAlso rdf:resource= "#Multimedia "/>
232 <rdfs:subClassOf rdf:resource= "#Multimedia "/>
233 </owl:Class>
234 <owl:Class rdf:ID= "XML ">
235 <rdfs:seeAlso rdf:resource= "#Formatos_de_datos "/>
236 <rdfs:subClassOf rdf:resource= "#Formatos_de_datos "/>
237 </owl:Class>
238 <owl:Class rdf:ID= "HTML ">
239 <rdfs:seeAlso rdf:resource= "#Formatos_de_datos "/>
240 <rdfs:subClassOf rdf:resource= "#Formatos_de_datos "/>
241 </owl:Class>
242 <owl:Class rdf:ID= "OWL ">
243 <rdfs:seeAlso rdf:resource= "#Formatos_de_datos "/>
244 <rdfs:subClassOf rdf:resource= "#Formatos_de_datos "/>
245 </owl:Class>
246 <owl:Class rdf:ID= "ASCII ">
247 <rdfs:seeAlso rdf:resource= "#Formatos_de_datos "/>
248 <rdfs:subClassOf rdf:resource= "#Formatos_de_datos "/>
249 </owl:Class>
250 <owl:Class rdf:ID= "XDR ">
251 <rdfs:seeAlso rdf:resource= "#Formatos_de_datos "/>
252 <rdfs:subClassOf rdf:resource= "#Formatos_de_datos "/>
253 </owl:Class>
254 <owl:Class rdf:ID= "GEDCOM ">
```

```
255 <rdfs:seeAlso rdf:resource= "#Formatos_de_dados "/>
256 <rdfs:subClassOf rdf:resource= "#Formatos_de_dados "/>
257 </owl:Class>
258 <owl:Class rdf:ID= "VRML ">
259 <rdfs:seeAlso rdf:resource= "#Formatos_de_dados "/>
260 <rdfs:subClassOf rdf:resource= "#Formatos_de_dados "/>
261 </owl:Class>
262 <owl:Class rdf:ID= "SMIL ">
263 <rdfs:seeAlso rdf:resource= "#Formatos_de_dados "/>
264 <rdfs:subClassOf rdf:resource= "#Formatos_de_dados "/>
265 </owl:Class>
266 <owl:Class rdf:ID= "SGML ">
267 <rdfs:seeAlso rdf:resource= "#Formatos_de_dados "/>
268 <rdfs:subClassOf rdf:resource= "#Formatos_de_dados "/>
269 </owl:Class>
270 <owl:Class rdf:ID= "DOC ">
271 <rdfs:seeAlso rdf:resource= "#Formatos_de_dados "/>
272 <rdfs:subClassOf rdf:resource= "#Formatos_de_dados "/>
273 </owl:Class>
274 <owl:Class rdf:ID= "MathML ">
275 <rdfs:seeAlso rdf:resource= "#Formatos_de_dados "/>
276 <rdfs:subClassOf rdf:resource= "#Formatos_de_dados "/>
277 </owl:Class>
278 <owl:Class rdf:ID= "PDF ">
279 <rdfs:seeAlso rdf:resource= "#Formatos_de_dados "/>
280 <rdfs:subClassOf rdf:resource= "#Formatos_de_dados "/>
281 </owl:Class>
282 <owl:Class rdf:ID= "Museus ">
283 <rdfs:seeAlso rdf:resource= "#Historia "/>
284 <rdfs:subClassOf rdf:resource= "#Historia "/>
285 </owl:Class>
286 <owl:Class rdf:ID= "Historia ">
287 </owl:Class>
288 <owl:Class rdf:ID= "Paises ">
289 </owl:Class>
290 <owl:Class rdf:ID= "America_do_sul ">
291 <rdfs:subClassOf rdf:resource= "#Paises "/>
292 </owl:Class>
293 <owl:Class rdf:ID= "Pais ">
294 <rdfs:seeAlso rdf:resource= "#Argentina "/>
295 <rdfs:seeAlso rdf:resource= "#Bolivia "/>
296 <rdfs:seeAlso rdf:resource= "#Brasil "/>
297 <rdfs:seeAlso rdf:resource= "#Chile "/>
298 <rdfs:seeAlso rdf:resource= "#Colombia "/>
299 <rdfs:seeAlso rdf:resource= "#Equador "/>
300 <rdfs:seeAlso rdf:resource= "#Guiana "/>
301 <rdfs:seeAlso rdf:resource= "#Paraguai "/>
302 <rdfs:seeAlso rdf:resource= "#Peru "/>
303 <rdfs:seeAlso rdf:resource= "#Suriname "/>
304 <rdfs:seeAlso rdf:resource= "#Uruguai "/>
305 <rdfs:seeAlso rdf:resource= "#Venezuela "/>
306 </owl:Class>
307 <owl:Class rdf:ID= "Bolivia ">
308 <rdfs:seeAlso rdf:resource= "#America_do_sul "/>
309 <rdfs:seeAlso rdf:resource= "#Pais "/>
310 <rdfs:subClassOf rdf:resource= "#America_do_sul "/>
311 </owl:Class>
312 <owl:Class rdf:ID= "Brasil ">
313 <rdfs:seeAlso rdf:resource= "#America_do_sul "/>
314 <rdfs:seeAlso rdf:resource= "#Pais "/>
315 <rdfs:subClassOf rdf:resource= "#America_do_sul "/>
316 </owl:Class>
317 <owl:Class rdf:ID= "Chile ">
318 <rdfs:seeAlso rdf:resource= "#America_do_sul "/>
319 <rdfs:seeAlso rdf:resource= "#Pais "/>
320 <rdfs:subClassOf rdf:resource= "#America_do_sul "/>
321 </owl:Class>
322 <owl:Class rdf:ID= "Venezuela ">
```

```
323 <rdfs:seeAlso rdf:resource= "#America_do_sul "/>
324 <rdfs:seeAlso rdf:resource= "#Pais "/>
325 <rdfs:subClassOf rdf:resource= "#America_do_sul "/>
326 </owl:Class>
327 <owl:Class rdf:ID= "Colombia ">
328 <rdfs:seeAlso rdf:resource= "#America_do_sul "/>
329 <rdfs:seeAlso rdf:resource= "#Pais "/>
330 <rdfs:subClassOf rdf:resource= "#America_do_sul "/>
331 </owl:Class>
332 <owl:Class rdf:ID= "Argentina ">
333 <rdfs:seeAlso rdf:resource= "#America_do_sul "/>
334 <rdfs:seeAlso rdf:resource= "#Pais "/>
335 <rdfs:subClassOf rdf:resource= "#America_do_sul "/>
336 </owl:Class>
337 <owl:Class rdf:ID= "Paraguai ">
338 <rdfs:seeAlso rdf:resource= "#America_do_sul "/>
339 <rdfs:seeAlso rdf:resource= "#Pais "/>
340 <rdfs:subClassOf rdf:resource= "#America_do_sul "/>
341 </owl:Class>
342 <owl:Class rdf:ID= "Uruguai ">
343 <rdfs:seeAlso rdf:resource= "#America_do_sul "/>
344 <rdfs:seeAlso rdf:resource= "#Pais "/>
345 <rdfs:subClassOf rdf:resource= "#America_do_sul "/>
346 </owl:Class>
347 <owl:Class rdf:ID= "Guiana ">
348 <rdfs:seeAlso rdf:resource= "#America_do_sul "/>
349 <rdfs:seeAlso rdf:resource= "#Pais "/>
350 <rdfs:subClassOf rdf:resource= "#America_do_sul "/>
351 </owl:Class>
352 <owl:Class rdf:ID= "Suriname ">
353 <rdfs:seeAlso rdf:resource= "#America_do_sul "/>
354 <rdfs:seeAlso rdf:resource= "#Pais "/>
355 <rdfs:subClassOf rdf:resource= "#America_do_sul "/>
356 </owl:Class>
357 <owl:Class rdf:ID= "Peru ">
358 <rdfs:seeAlso rdf:resource= "#America_do_sul "/>
359 <rdfs:seeAlso rdf:resource= "#Pais "/>
360 <rdfs:subClassOf rdf:resource= "#America_do_sul "/>
361 </owl:Class>
362 <owl:Class rdf:ID= "Equador ">
363 <rdfs:seeAlso rdf:resource= "#America_do_sul "/>
364 <rdfs:seeAlso rdf:resource= "#Pais "/>
365 <rdfs:subClassOf rdf:resource= "#America_do_sul "/>
366 </owl:Class>
367 <owl:Class rdf:ID= "Eleicoes ">
368 </owl:Class>
369 <owl:Class rdf:ID= "Eleicoes_no_Brasil ">
370 <rdfs:seeAlso rdf:resource= "#Eleicoes "/>
371 <rdfs:subClassOf rdf:resource= "#Eleicoes "/>
372 </owl:Class>
373 <owl:Class rdf:ID= "Calendarios ">
374 <rdfs:seeAlso rdf:resource= "#Eleicoes "/>
375 <rdfs:subClassOf rdf:resource= "#Eleicoes "/>
376 </owl:Class>
377 <owl:Class rdf:ID= "urna_eletronica ">
378 <rdfs:seeAlso rdf:resource= "#Eleicoes "/>
379 <rdfs:subClassOf rdf:resource= "#Eleicoes "/>
380 </owl:Class>
381 <owl:Class rdf:ID= "Os_produtos_comerciais_@ ">
382 <rdfs:seeAlso rdf:resource= "#Eleicoes "/>
383 <rdfs:subClassOf rdf:resource= "#Eleicoes "/>
384 </owl:Class>
385 <owl:Class rdf:ID= "congresso_nacional ">
386 <rdfs:seeAlso rdf:resource= "#Eleicoes "/>
387 <rdfs:subClassOf rdf:resource= "#Eleicoes "/>
388 </owl:Class>
389 <owl:Class rdf:ID= "politicos ">
390 <rdfs:seeAlso rdf:resource= "#Eleicoes "/>
```

```
391 <rdfs:subClassOf rdf:resource= "#Eleicoes "/>
392 </owl:Class>
393 <owl:Class rdf:ID= "Os_direitos_de_voto ">
394 <rdfs:seeAlso rdf:resource= "#Eleicoes "/>
395 <rdfs:subClassOf rdf:resource= "#Eleicoes "/>
396 </owl:Class>
397 <owl:Class rdf:ID= "Reforma ">
398 <rdfs:seeAlso rdf:resource= "#Eleicoes "/>
399 <rdfs:subClassOf rdf:resource= "#Eleicoes "/>
400 </owl:Class>
401 <owl:Class rdf:ID= "Eletronico_de_votacao_e_Internet ">
402 <rdfs:seeAlso rdf:resource= "#Eleicoes "/>
403 <rdfs:subClassOf rdf:resource= "#Eleicoes "/>
404 </owl:Class>
405 <owl:Class rdf:ID= "Propaganda_politica ">
406 <rdfs:seeAlso rdf:resource= "#Eleicoes "/>
407 <rdfs:subClassOf rdf:resource= "#Eleicoes "/>
408 </owl:Class>
409 <owl:Class rdf:ID= "capital_do_brasil ">
410 <rdfs:seeAlso rdf:resource= "#Eleicoes "/>
411 <rdfs:subClassOf rdf:resource= "#Eleicoes "/>
412 </owl:Class>
413 <owl:Class rdf:ID= "Sistemas_ Eleitorais ">
414 <rdfs:seeAlso rdf:resource= "#Eleicoes "/>
415 <rdfs:subClassOf rdf:resource= "#Eleicoes "/>
416 </owl:Class>
417 <owl:Class rdf:ID= "Universidades ">
418 </owl:Class>
419 <owl:Class rdf:ID= "Universidades_Federais ">
420 <rdfs:seeAlso rdf:resource= "#Universidades "/>
421 <rdfs:subClassOf rdf:resource= "#Universidades "/>
422 </owl:Class>
423 <owl:Class rdf:ID= "Universidades_Estaduais ">
424 <rdfs:seeAlso rdf:resource= "#Universidades "/>
425 <rdfs:subClassOf rdf:resource= "#Universidades "/>
426 </owl:Class>
427 <owl:Class rdf:ID= "Universidades_Privadas ">
428 <rdfs:seeAlso rdf:resource= "#Universidades "/>
429 <rdfs:subClassOf rdf:resource= "#Universidades "/>
430 </owl:Class>
431 <owl:Class rdf:ID= "UFCG ">
432 <rdfs:seeAlso rdf:resource= "#Universidade_Federal_de_Campina_Grande
"/>
433 </owl:Class>
434 <owl:Class rdf:ID= "UFPB ">
435 <rdfs:seeAlso rdf:resource= "#Universidade_Federal_da_Paraiba "/>
436 </owl:Class>
437 <owl:Class rdf:ID= "Universidade_Federal_de_Campina_Grande ">
438 <rdfs:seeAlso rdf:resource= "#Universidades_Federais "/>
439 <rdfs:seeAlso rdf:resource= "#UFCG "/>
440 <rdfs:subClassOf rdf:resource= "#Universidades_Federais "/>
441 </owl:Class>
442 <owl:Class rdf:ID= "Universidade_Federal_da_Paraiba ">
443 <rdfs:seeAlso rdf:resource= "#Universidades_Federais "/>
444 <rdfs:seeAlso rdf:resource= "#UFPB "/>
445 <rdfs:subClassOf rdf:resource= "#Universidades_Federais "/>
446 </owl:Class>
447 <owl:Class rdf:ID= "IFPB ">
448 <rdfs:seeAlso rdf:resource= "#Instituto_Federal_da_Paraiba "/>
449 </owl:Class>
450 <owl:Class rdf:ID= "Instituto_Federal_da_Paraiba ">
451 <rdfs:seeAlso rdf:resource= "#Universidades_Federais "/>
452 <rdfs:seeAlso rdf:resource= "#IFPB "/>
453 <rdfs:subClassOf rdf:resource= "#Universidades_Federais "/>
454 </owl:Class>
455 <owl:Class rdf:ID= "UEPB ">
456 <rdfs:seeAlso rdf:resource= "#Universidade_Estadual_da_Paraiba "/>
457 </owl:Class>
```

```

458 <owl:Class rdf:ID= "Universidade_Estadual_da_Paraiba ">
459 <rdfs:seeAlso rdf:resource= "#Universidades_Estaduais "/>
460 <rdfs:seeAlso rdf:resource= "#UEPB "/>
461 <rdfs:subClassOf rdf:resource= "#Universidades_Estaduais "/>
462 </owl:Class>
463 <owl:Class rdf:ID= "FIP ">
464 <rdfs:seeAlso rdf:resource= "#Faculdades_Integradas_de_Patos "/>
465 </owl:Class>
466 <owl:Class rdf:ID= "Faculdade_de_Ciencias_Sociais_Aplicadas ">
467 <rdfs:seeAlso rdf:resource= "#Universidades_Privadas "/>
468 <rdfs:seeAlso rdf:resource= "#FACISA "/>
469 <rdfs:subClassOf rdf:resource= "#Universidades_Privadas "/>
470 </owl:Class>
471 <owl:Class rdf:ID= "Faculdade_de_Medicina_Nova_Esperanca ">
472 <rdfs:seeAlso rdf:resource= "#Universidades_Privadas "/>
473 <rdfs:seeAlso rdf:resource= "#FAMENE "/>
474 <rdfs:subClassOf rdf:resource= "#Universidades_Privadas "/>
475 </owl:Class>
476 <owl:Class rdf:ID= "Faculdade_de_Ciencias_Medicas_de_Campina_Grande
">
477 <rdfs:seeAlso rdf:resource= "#Universidades_Privadas "/>
478 <rdfs:seeAlso rdf:resource= "#FCMCG "/>
479 <rdfs:subClassOf rdf:resource= "#Universidades_Privadas "/>
480 </owl:Class>
481 <owl:Class rdf:ID= "FCMCG ">
482 <rdfs:seeAlso                                rdf:resource=
"#Faculdade_de_Ciencias_Medicas_de_Campina_Grande "/>
483 </owl:Class>
484 <owl:Class rdf:ID= "Faculdades_Integradas_de_Patos ">
485 <rdfs:seeAlso rdf:resource= "#Universidades_Privadas "/>
486 <rdfs:seeAlso rdf:resource= "#FIP "/>
487 <rdfs:subClassOf rdf:resource= "#Universidades_Privadas "/>
488 </owl:Class>
489 <owl:Class rdf:ID= "FACISA ">
490 <rdfs:seeAlso                                rdf:resource=
"#Faculdade_de_Ciencias_Sociais_Aplicadas "/>
491 </owl:Class>
492 <owl:Class rdf:ID= "FAMENE ">
493 <rdfs:seeAlso  rdf:resource= "#Faculdade_de_Medicina_Nova_Esperanca
"/>
494 </owl:Class>
495 <owl:Class rdf:ID= "Instituto_de_Educacao_Superior_da_Paraiba ">
496 <rdfs:seeAlso rdf:resource= "#Universidades_Privadas "/>
497 <rdfs:seeAlso rdf:resource= "#IESP "/>
498 <rdfs:subClassOf rdf:resource= "#Universidades_Privadas "/>
499 </owl:Class>
500 <owl:Class rdf:ID= "Faculdade_de_Enfermagem_Nova_Esperanca ">
501 <rdfs:seeAlso rdf:resource= "#Universidades_Privadas "/>
502 <rdfs:seeAlso rdf:resource= "#FACENE "/>
503 <rdfs:subClassOf rdf:resource= "#Universidades_Privadas "/>
504 </owl:Class>
505 <owl:Class rdf:ID= "Centro_Universitario_de_Joao_Pessoa ">
506 <rdfs:seeAlso rdf:resource= "#Universidades_Privadas "/>
507 <rdfs:seeAlso rdf:resource= "#UNIPE "/>
508 <rdfs:subClassOf rdf:resource= "#Universidades_Privadas "/>
509 </owl:Class>
510 <owl:Class rdf:ID= "UNIPE ">
511 <rdfs:seeAlso  rdf:resource= "#Centro_Universitario_de_Joao_Pessoa
"/>
512 </owl:Class>
513 <owl:Class rdf:ID= "FACENE ">
514 <rdfs:seeAlso rdf:resource= "#Faculdade_de_Enfermagem_Nova_Esperanca
"/>
515 </owl:Class>
516 <owl:Class rdf:ID= "IESP ">
517 <rdfs:seeAlso                                rdf:resource=
"#Instituto_de_Educacao_Superior_da_Paraiba "/>
518 </owl:Class>

```

```
519    </rdf:RDF>
520
521    <!-- Created with Protege (with OWL Plugin 3.4.4, Build 579)
http://protege.stanford.edu -->
```