



TÁRCIO SULLIVAN MOTA DORO

Expansão de Consultas Utilizando Ontologia

Juiz de Fora
Dezembro de 2010

TÁRCIO SULLIVAN MOTA DORO

Expansão de Consultas Utilizando Ontologia

Universidade Federal de Juiz de Fora
Instituto de Ciências Exatas
Departamento de Ciência da Computação
Bacharelado em Ciência da Computação

Orientadora: Profa. Regina Maria Maciel Braga Villela

Juiz de Fora
Dezembro de 2010

TÁRCIO SULLIVAN MOTA DORO

Expansão de Consultas Utilizando Ontologia

Monografia submetida ao corpo docente do Departamento de Ciência da Computação do Instituto de Ciências Exatas da Universidade Federal de Juiz de Fora, como parte integrante dos requisitos necessários para a obtenção do grau de Bacharel em Ciência da Computação.

Aprovada pela banca em 09/12/2010 constituída pelos seguintes professores:

Regina Maria Maciel Braga Villela – DCC/UFJF
orientadora

Alessandréia Marta de Oliveira Júlio – DCC/UFJF

Jairo Francisco de Souza – DCC/UFJF

Juiz de Fora
Dezembro de 2010

Dedicatórias

Dedico este trabalho aos meus pais Helber e Marta, a minha irmã Tércila e a minha namorada Marcela por se constituírem diferentemente enquanto pessoas, igualmente belas e admiráveis em essência, estímulos que me impulsionaram a buscar vida nova a cada dia, meus agradecimentos por terem aceito se privar de minha companhia pelos estudos, concedendo a mim a oportunidade de me realizar ainda mais.

Agradecimentos

Primeiramente a Deus, por me dar saúde e força para concluir mais essa etapa da minha vida. E também aos amigos e professores da Universidade Federal de Juiz de Fora, pelo incentivo que me deram durante todo o trabalho, tanto nesta pesquisa quanto no nosso cotidiano.

Resumo

Com o crescimento da web, está cada vez mais difícil para o usuário realizar pesquisas que retornam documentos que sejam considerados relevantes para a sua necessidade. Para tentar resolver esse problema foram desenvolvidas várias técnicas, entre elas a técnica de expansão de consultas. Tal técnica adiciona à consulta inicial termos tidos como relevantes para o contexto de pesquisa. Existem diversos modos de implementar a técnica de expansão de consultas. Esse trabalho descreve alguns trabalhos relacionados à técnica de expansão de consultas, tendo um enfoque maior nas técnicas que utilizam ontologias para decidir quais termos adicionar na consulta inicial. Para isso o trabalho também define ontologia e mostra suas principais características que podem ajudar na aplicação da expansão de consultas. Um protótipo que aplica a técnica de expansão de consultas que utiliza ontologia foi desenvolvido, seu funcionamento é detalhado e para finalizar é realizada uma análise dos resultados obtidos.

Palavras-chave: Ontologia. Expansão de consultas. Semântica.

Abstract

Considering the growth of web, it is increasingly difficult for the user to perform searches that return documents that are considered relevant to his needs. In order to consider this problem, several techniques were developed, including the query expansion technique. This technique adds to the initial query terms considered as relevant to the search context. There are several methods to implement the query expansion technique. This work describes some of them, with a focus on techniques that use ontologies to decide which terms to add to the initial query. For this, the work also defines ontology and shows its main features that can help in the implementation of query expansion. A prototype that applies the query expansion technique and uses ontologies was developed. Its operation is explained and some reflections are done.

Keywords: Ontology. Query Expansion. Semantics.

Lista de Figuras

Figura 1 – Comparação entre ontologia e base de conhecimento.....	14
Figura 2: Ontologia de produto.....	17
Figura 3: Ontologia com relacionamento “é-relacionado-a”.....	18
Figura 4: Relações taxonômicas na Word Net.....	27
Figura 5: Um screenshot de uma pesquisa no Google com o Surf Canyon instalado.....	30
Figura 6: Processo de expansão de consultas.....	34
Figura 7: Arquitetura simplificada do protótipo.....	36
Figura 8: Ontologia utilizada na aplicação.....	38
Figura 9: Código fonte da classe processaConsulta.....	43
Figura 10: Tela inicial do protótipo	44
Figura 11: Consulta expandida	45
Figura 12: Resultados da consulta “EXEMPLO DE CONSUMIDORES”.....	46
Figura 13: Resultados da consulta “ANIMAIS DA FLORESTA”.....	47

Sumário

CAPÍTULO 1 – INTRODUÇÃO.....	9
1.2. OBJETIVOS.....	10
1.3. METODOLOGIA.....	10
1.4. JUSTIFICATIVA.....	11
CAPÍTULO 2 – ONTOLOGIAS.....	12
2.1 DEFINIÇÃO:.....	12
2.2 COMPOSIÇÃO DE UMA ONTOLOGIA.....	14
2.3 APLICAÇÕES DE ONTOLOGIAS.....	15
2.4 RELAÇÕES.....	16
2.5 ATRIBUTOS.....	18
2.6 USO DE ONTOLOGIAS EM EXPANSÃO DE CONSULTAS.....	19
2.7 CONSIDERAÇÕES ADICIONAIS.....	20
CAPÍTULO 3 - EXPANSÃO DE CONSULTAS.....	21
3.1 DEFINIÇÃO:.....	21
3.2 TÉCNICAS DE EXPANSÃO DE CONSULTAS:.....	22
3.2.1 Retorno Relevante.....	22
3.2.2 Expansão Probabilística.....	23
3.2.2.1 Tipos de Métodos Probabilísticos.....	24
3.2.3 Outras Técnicas.....	25
3.2.4 Expansão Utilizando Ontologias.....	25
3.3 APLICAÇÕES EXISTENTES.....	28
3.3.1 Aplicação do Retorno Relevante.....	28
3.3.2 Aplicações que Utilizam Ontologia.....	30
3.3.2.1 Utilizando a Ontologia WordNet.....	30
3.3.2.2 Utilizando Ontologias de Domínio Especifico.....	31
CAPÍTULO 4 – ESTUDO DE CASO: USO DE TÉCNICAS DE EXPANSÃO DE CONSULTAS.....	36
4.1 CARACTERÍSTICAS GERAIS.....	36
4.2 ONTOLOGIA PARA ECOLOGIA.....	37
4.3 USO DO GOOGLE.....	39
4.4 DESENVOLVIMENTO.....	39
4.4.1 Ambiente de Desenvolvimento.....	39
4.4.2 Classes do Protótipo.....	40
4.5 EXEMPLOS DE USO DO PROTÓTIPO.....	43
4.6 ANÁLISE DA TÉCNICA APLICADA.....	48
CAPÍTULO 5 – CONCLUSÃO.....	49

REFERÊNCIAS BIBLIOGRÁFICAS.....	51
--	-----------

Capítulo 1 – Introdução

1.1 . Conceituação

Nos últimos anos o crescimento da internet, tanto em questão de conteúdo quanto em número de usuários unidos a grande melhoria na tecnologia dos mecanismos de busca, mudou radicalmente a forma como o conhecimento e as informações são coletadas e compartilhadas. O acesso a informação nunca foi tão fácil e aberto a um público tão grande como é hoje. Nesse contexto, é que encontrar informações precisas na internet vem se tornando uma tarefa cada vez mais importante e crítica.

Apesar das melhorias alcançadas pelos mecanismos de busca, há ainda um número significativo de casos em que os resultados retornados por esses mecanismos contêm um elevado número de resultados irrelevantes. Em muitos casos, os usuários não sabem como criar consultas eficientes, realizando consultas com poucos termos e sem sinônimos, esse fato ocorre principalmente quando a consulta pertence a um domínio desconhecido pelo usuário. Outro problema é o fato de que as palavras podem ter vários sentidos, por exemplo, se uma pessoa pesquisar pela palavra “manga” não tem como saber se ele se refere à fruta ou à parte de uma camisa.

Segundo (FARHOODI et al, 2009) a maioria dos mecanismos de busca realiza suas pesquisas com base em palavras-chave de uma consulta e as páginas da web são consideradas como um conjunto de palavras, sem a percepção do criador da página. Para melhorar a qualidade da consulta e conseqüentemente melhorar o resultado dos mecanismos de buscas que trabalham com palavras chaves, foram propostos vários métodos de expansão de consultas. A técnica de expansão de consulta adiciona termos à consulta original, assim os mecanismos de busca têm um número maior de termos para serem pesquisados, ampliando também o número de documentos relacionado com a consulta. O grande problema é saber quais termos adicionar, pois apenas a quantidade de termos não resolve o problema, os termos têm que ser semanticamente relacionados com a consulta, caso contrário os resultados obtidos podem ser piores do que com a aplicação da técnica. Estudos mostram que adicionando sinônimos, termos que pertencem ao mesmo contexto, termos com relação léxica e/ou semântica aprimoram o resultado da pesquisa.

As ontologias podem ser vistas como uma representação estruturada do conhecimento e são utilizadas para a verificação do grau e o tipo de relacionamento entre termos. Outras técnicas também podem ser utilizadas. Porém, com o uso da ontologia é possível que seja agregada semântica à consulta fornecida pelo usuário e o resultado final da pesquisa não depende de um resultado inicial favorável como ocorre em outras técnicas. Portanto, ontologias podem ser utilizadas para expandir as consultas, seja manualmente, automaticamente ou semi-automaticamente.

1.2. Objetivos

Os objetivos desse trabalho são analisar as diversas técnicas de expansão de consulta e compará-las com técnicas que utilizam ontologias para realizar a expansão, além de realizar um estudo comparativo entre essas técnicas que utilizam ontologia. Como objetivo secundário, podemos ressaltar o desenvolvimento de um protótipo que compara resultados obtidos por uma consulta fornecida pelo usuário com os resultados obtidos por uma consulta expandida através do uso de uma ontologia.

1.3. Metodologia

A primeira fase no desenvolvimento desse trabalho foi a realização do levantamento de material bibliográfico sobre ontologias, a fim de montar o capítulo 2, que define ontologias mostrando suas principais características e aplicações na área de ciência da computação e ressaltando seu uso na técnica de expansão de consultas.

A segunda fase abrangeu as pesquisas sobre as várias abordagens da técnica de expansão de consultas. Com a análise do material coletado foi desenvolvido o capítulo 3, que apresenta as técnicas de expansão de consultas, mostrando suas vantagens e desvantagens, além de exemplos de sistemas que usam tal técnica. Apesar do capítulo falar sobre expansão de consultas de modo geral, ele possui um enfoque especial às técnicas que utilizam ontologias.

A próxima fase foi o desenvolvimento de um protótipo que expande consultas fornecidas pelo usuário, a ontologia utilizada pertence ao domínio da ecologia. O último passo

foi desenvolver o capítulo 4, que descreve a protótipo criado, e faz uma análise dos resultados obtidos e propõe melhorias para a técnica utilizada.

1.4. Justificativa

Apesar do aprimoramento dos sistemas de busca que ocorreu nos últimos anos, muitas vezes um usuário faz inúmeras pesquisas para encontrar algo de seu interesse, tornando a fase da pesquisa muito demorada e em alguns casos sem sucesso. Nos dias de hoje esse tempo “perdido” na realização de pesquisas pode ser crucial para o usuário. Outra situação que ocorre com bastante frequência é o fato de muitas páginas, que podem ser consideradas relevantes para certo domínio, muitas vezes receberem poucos acessos, isso se deve ao fato dessa página não aparecer entre os primeiros resultados por um mecanismo de busca. Devido a esse cenário, existe a necessidade de se estudar novas técnicas de busca, que sejam capazes de retornar resultados mais relevantes para os usuários, entre essas técnicas está a expansão de consultas que utiliza ontologia.

Capítulo 2 – Ontologias

2.1 Definição:

O termo ontologia vem da filosofia, em que a ontologia é uma sistemática da existência. O radical "onto", do grego, tem o significado de indivíduo ou ser, e "logia", que significa estudo, tem-se que "ontologia" vem a ser o estudo investigativo e comparativo do indivíduo, no caso, um exemplar na espécie humana frente aos demais seres vivos, passando pela sua concepção, criação, evolução e extinção.

No campo da inteligência artificial, de acordo com (GRUBER, 1993), ontologia é o estudo das categorias de coisas que existem ou podem existir em algum domínio. Quando queremos representar o conhecimento de um domínio através de um formalismo, o conjunto de objetos que podem ser representados é chamado de universo de discurso. Nesse contexto de representação do conhecimento, podemos descrever uma ontologia definindo o conjunto de termos representacionais. Em uma ontologia, definições associam os nomes das entidades no universo do discurso (por exemplo, conceitos, relações, atributos, ou outros objetos) com textos que descrevem o significado dos nomes e axiomas formais que restringem a interpretação e o uso desses termos. Ontologias fornecem uma maneira estruturada de descrever o conhecimento.

Ainda segundo (GRUBER, 1993) uma ontologia é “Uma ontologia é uma especificação formal explícita de uma conceitualização compartilhada”. Nessa definição a palavra “conceitualização” refere-se a um modelo abstrato de algum fenômeno que identifique conceitos relevantes desse fenômeno, a palavra “explícita” significa que os tipos de conceitos usados e as limitações do uso desses conceitos devem ser definidos de forma explícita. A palavra “formal” refere-se que a ontologia deve ser passível de ser processada por uma máquina, por fim, “compartilhada” reflete a noção de que a ontologia captura um conhecimento consensual, isto é, esse conhecimento não deve ser restrito a alguns indivíduos, mas aceito por um grupo de pessoas (FENSEL, 2001).

Uma definição que complementa bem a definição de Gruber é a definição dada por (GÓMEZ-PÉREZ, 1999): “Uma ontologia é um conjunto de termos ordenados hierarquicamente para descrever um domínio que pode ser usado como um esqueleto para

uma base de conhecimentos.”. Conforme essa descrição, uma ontologia deve possuir um conjunto de conceitos organizados com uma hierarquia associada, ou seja, uma taxonomia.

Ontologias podem ser vistas como tipos especiais de grafos indicando as entidades que existem em um domínio, seus conceitos e as relações entre eles. O produto da ontologia é um catálogo dos “tipos de coisas” que se presume existirem em um domínio (um domínio é apenas uma área específica ou área de conhecimento, como a medicina, a fabricação de ferramentas, imobiliário, reparação automóvel, gestão financeira, etc.) de interesse.

Ontologias possuem um grande poder de descrição do conhecimento, podendo descrever informações presentes em documentos distintos.

Existe diferença entre uma ontologia e uma base de conhecimento. A ontologia provê uma estrutura na qual se pode construir uma base de conhecimento, fornecendo um conjunto de conceitos para descrever um determinado domínio, enquanto a base de conhecimento usa esses conceitos para descrever uma determinada realidade. A figura 1 exemplifica essa diferença. Nessa figura temos uma ontologia para circuitos eletrônicos e uma base de conhecimento que utiliza a ontologia. Podemos observar que apesar da realidade descrita na base de conhecimento variar para o rádio, para a calculadora e para a televisão, a ontologia empregada continua sendo a mesma, ou seja, são utilizados os mesmos conceitos para descrever as duas realidades.

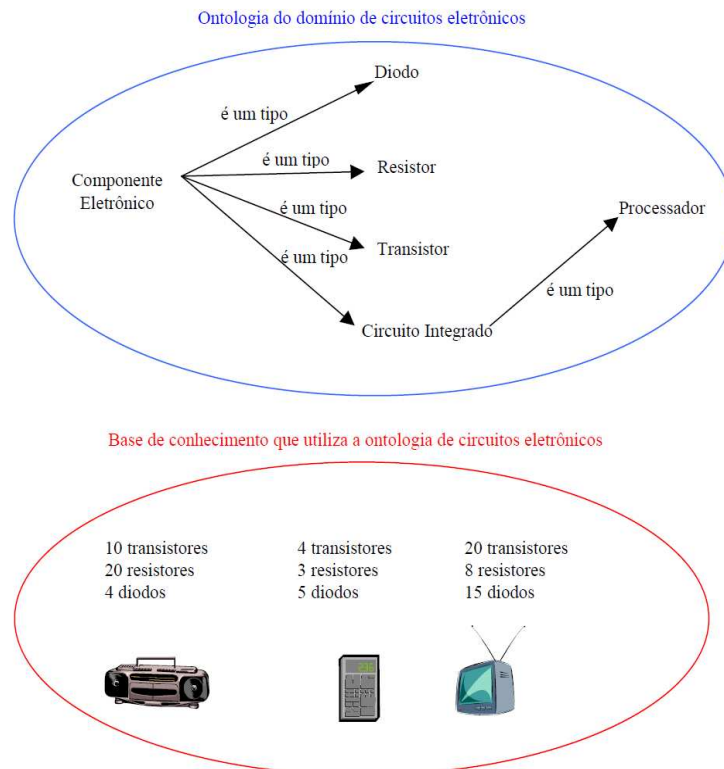


Figura 1 – Comparação entre ontologia e base de conhecimento

Fonte: GUIMARÃES, 2002, pág. 52

2.2 Composição de Uma Ontologia

Os blocos básicos de construção de ontologias são os conceitos e os relacionamentos. Conceitos (ou classes ou categorias ou tipos) podem ser pensados como conjuntos e aparecem como nós no gráfico de ontologias. Existe uma hierarquia entre esses conceitos, ou seja, uma taxonomia. Os conceitos podem ser abstratos (por ex.: força) ou concretos (por ex.: carro), elementares (por ex.: elétron) ou compostos (por ex.: átomo), reais ou fictícios. Um exemplo de taxonomia é o conceito homem ser um subconceito do conceito pessoa. Conceitos em ontologias têm geralmente um texto para defini-los. Quase sempre os conceitos são descritos por uma ou mais propriedades. Observe que cada conceito pode ter mais de uma propriedade que o descreve e que a propriedade não precisa corresponder a apenas um conceito. Por exemplo, para descrever o conceito de futebol, as propriedades "gol" e "jogadores" podem ser usadas. No entanto, a propriedade "jogadores" pode também se referir ao conceito de algum outro esporte. Normalmente, as ontologias incluem uma única propriedade para cada conceito. Relacionamentos são geralmente de um tipo específico e conectam dois ou mais

conceitos, a maioria das ontologias incluem relações é-um e parte-de entre conceitos, ou seja, "carro é um veículo e "Terra é-parte do Sistema Solar".

2.3 Aplicações de Ontologias

O uso de ontologia em Ciência da Computação apresenta muitas vantagens, e suas principais são:

- Fornecer um vocabulário para representação do conhecimento, esse vocabulário tem por trás uma conceitualização que o sustenta, evitando assim interpretações ambíguas.
- Compartilhar o conhecimento de modo que caso exista uma ontologia que modele adequadamente certo domínio de conhecimento, essa pode ser compartilhada e usada por pessoas que desenvolvam aplicações dentro desse domínio.
- Fornecer uma descrição exata do conhecimento, diferentemente da linguagem natural em que as palavras podem ter semântica totalmente diferente conforme o seu contexto, a ontologia por ser escrita em linguagem formal e também é possível estender o uso de uma ontologia genérica de forma a que ela se adéque a um domínio específico.

Ontologias têm sido utilizadas para descrever artefatos com diferentes graus de estrutura, estes vão desde simples taxonomias, a esquemas de metadados e teorias lógicas. Suas principais áreas de aplicação na ciência da computação são no processamento de linguagem natural, na gestão do conhecimento, no comércio eletrônico e na web semântica. A aplicação de maior interesse no contexto deste trabalho é na web semântica.

Em (ABURDENE; NAISBITT, 1990) é mostrado bem o estado atual da web dizendo que nós estamos nos afogando em informações e com fome de conhecimento, tal afirmação é verdadeira, pois existe muita informação disponível na internet, existem muitas pessoas interessadas nessas informações, porém ainda se encontra uma dificuldade muito grande em pesquisar essas informações na internet. Os mecanismos atuais não conseguem saber ao certo sobre o que se trata uma página da internet, heurísticas são utilizadas para tentar compreender

o significado do conteúdo da página. Como alternativa a essas heurísticas as ontologias têm sido utilizadas como uma forma de representar a semântica de documentos.

2.4 Relações

Uma ontologia pode possuir diversos tipos de relações entre seus conceitos, porém algumas são mais importantes do que as outras dependendo do contexto de utilização. No contexto da expansão de consultas, área de interesse desse trabalho, encontramos uma grande diversidade de técnicas que utilizam ontologia.

Um dos tipos de relacionamento mais relevantes para essas técnicas é a relação de sinonímia, ou seja, a ontologia contém a informação de que conceitos são equivalentes semanticamente, Salton realizou vários experimentos, dentre eles o Cranfield-I (CRAN-1), que é o de maior relevância para acervos específicos de domínio. Uma das etapas do experimento CRAN-1 analisou de que forma as relações de sinonímia impactariam a eficiência da recuperação de informação. Os resultados mostraram que estas relações produzem melhoras significativas na recuperação de informação. Por exemplo, se um usuário deseja realizar uma consulta para saber o preço do produto roupa, essa consulta pode ser expandida utilizando-se o termo vestuário já que é um sinônimo de roupa.

Uma relação que está presente na maior parte das ontologias seja qual for sua aplicação é a relação “é-um”, ela pode tanto especificar quanto generalizar um termo, por exemplo, quando utilizamos a relação “é-um” no termo veículo podemos ter o seguinte cenário, “carro é-um veículo”, portanto carro é uma especialização de veículo, então uma consulta que deseja informações sobre veículo pode ser expandida utilizando a palavra carro.

Outro tipo de relacionamento que aparece em grande parte das ontologias é o “parte-de”, essa relação mostra o conjunto de partes que constitui um elemento, por exemplo, “motor é parte-de um carro”. Porém essa relação pode apresentar vários sentidos. Para resolver esse problema, algumas ontologias como a WordNet utilizam diversos tipos de relacionamento “parte-de”. Essa prática auxilia na criação de uma hierarquia entre os conceitos. Alguns desses relacionamentos utilizados pela WordNet são:

- Sentido de membro: “professor é-membro do quadro de funcionários”

- Sentido de matéria: “chás são feitos-de água”
- Sentido de posse: “parque de diversões tem brinquedos”

Os relacionamentos “é-um” e “parte-de” já citados anteriormente são muito úteis para quando se deseja formar hierarquias, porém a transitividade implícita em hierarquias causa alguns problemas. Por exemplo, "Minha mão é parte de mim" e "Eu sou parte de um comitê", assim, somos levados a uma conclusão muito estranha de que "Minha mão é parte de um comitê". Este fenômeno é causado quando diferentes tipos do relacionamento “parte-de” são misturados, como é descrito em (WINSTON et al, 1987), porém em algumas situações esse fenômeno pode ser benéfico, como no domínio da biologia molecular, caso o usuário consulte por elementos constituídos por carbono, podem ser recuperados elementos constituídos por outros elementos que contenham carbono, embora o banco de dados não represente explicitamente esta informação.

A figura 2 ilustra uma ontologia de produto que utiliza todos os relacionamentos tratados até esse ponto do trabalho.

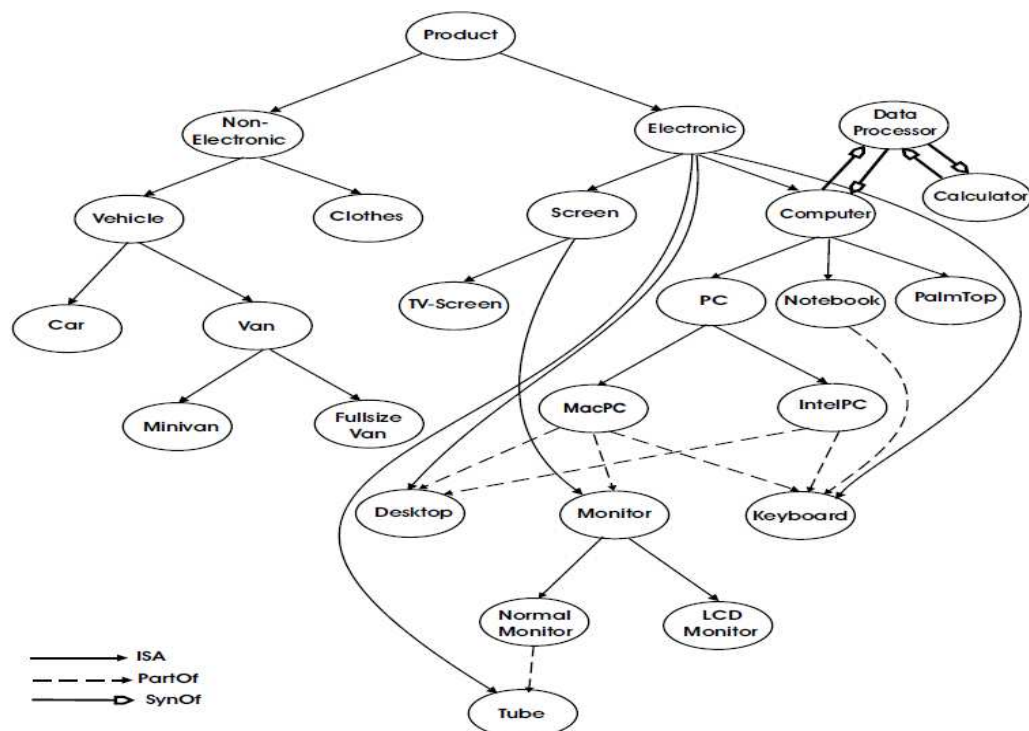


Figura 2: Ontologia de produto

Fonte: FREITAG; NECIB, 2004, pág. 6

Algumas ontologias utilizam tipos de relacionamentos um pouco mais gerais, com sentidos mais amplos, um exemplo desse tipo de relacionamento é a relação “é-relacionado-a”, nesse tipo de relação cada ligação entre dois elementos da ontologia possui um peso, esse peso demonstra o quão relacionado são esses elementos, a figura 3 mostra uma ontologia que possui somente esse tipo de relacionamento.

Existem algoritmos que se baseiam nos valores dos pesos atribuídos ao relacionamento “é-relacionado-a” para determinar qual termo adicionar à consulta original, um desses algoritmos é descrito em (FARHOODI et al, 2009). Em tal algoritmo a consulta inicial é representada por um vetor e a ontologia é representada por uma matriz, esse algoritmo é apresentado com maiores detalhes na seção 3.2.3.

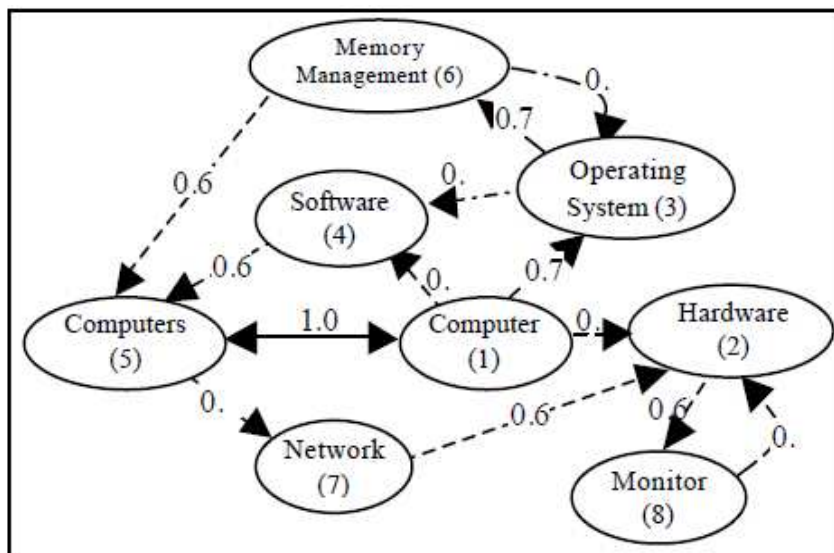


Figura 3: Ontologia com relacionamento “é-relacionado-a”

Fonte: FARHOODI et al, 2009, pág. 4

2.5 Atributos

Para descrever conceitos em uma ontologia atributos também podem ser utilizados. Cada atributo tem pelo menos um nome e um valor, e é utilizado para armazenar informações que são específicas de conceito. Por exemplo, um objeto carro pode ter como atributos:

- Nome: sem marca XT

- Número_de_portas: 4
- Motor: {1.0, 1.3}
- Câmbio: 5-marchas

Assim como os relacionamentos os atributos também podem ser utilizados para realizar a expansão de consultas. Existem mecanismos que analisam os atributos de um domínio para inferir novas informações e utilizá-las na expansão da consulta. Por exemplo, dada uma ontologia que descreva frutas, um atributo pode estabelecer que maçãs vermelhas sejam maçãs de cor vermelha, assim, caso o usuário deseje buscar as instâncias de maçã vermelha, a consulta pode ser expandida para recuperar instâncias de maçã associadas ao atributo cor vermelha.

2.6 Uso de Ontologias em Expansão de Consultas

Muitas técnicas de expansão de consultas fazem proveito da semântica da semântica que é inerente as ontologias. Tal técnica consiste em adicionar termos à consulta original fornecida por usuários que procuram por informações relevantes presentes em algum repositório de dados ou até na internet. Tais termos são escolhidos de acordo com as informações presentes na ontologia.

Na aplicação da expansão de consultas, muitas vezes é necessário retirar a ambiguidade de algum termo pertencente à consulta inicial, por exemplo, no caso em que o usuário deseja pesquisar pelo termo “banco”, se não for retirada a ambiguidade a pesquisa pode retornar muitos resultados com banco no sentido de móvel para sentar, sendo que o usuário estava interessado em realizar pesquisa sobre banco no sentido de instituição financeira ou vice-versa.

Ontologia também tem aplicação em algoritmos que realizam a desambiguação do termo. Um desses algoritmos é descrito em (MIHALCEA et al, 2004). Nesse algoritmo a ontologia é tratada como um grafo e utilizam Pagerank (PAGE et al 1999) para retirar a ambiguidade.

Outra maneira de desambiguar um termo baseia-se no uso de ontologia para apresentar ao usuário os diferentes contextos envolvidos de forma que ele seja capaz de escolher qual o

contexto mais apropriado. Os diferentes contextos são apresentados ao usuário, que é responsável por decidir qual deles é o mais adequado para sua consulta.

As ontologias vêm se mostrando serem cruciais para as aplicações que pesquisam informações através de diversos documentos, já que o mesmo termo pode ser usado com significados diferentes em contextos diferentes e termos diferentes podem ser usados para itens que têm o mesmo significado.

2.7 Considerações Adicionais

Além dos tópicos estudados no presente trabalho, existem ainda muitas outras características de ontologias, como: formas de representação, ferramentas para criação e edição, técnicas de desenvolvimento, entre outras. Essas características não são estudadas nesse trabalho, pois foi dado um foco maior para as características essenciais para a melhoria da técnica de expansão de consultas.

Ontologias possuem uma grande aplicação não só na ciência da Computação, mas no mundo da ciência como um todo. Seu poder de representar o conhecimento é essencial para o desenvolvimento de pesquisas realizadas em diversas áreas, para cada área de atuação, tipos diferentes de relacionamentos e atributos são utilizados.

No próximo capítulo será abordada a técnica de expansão de consultas. Especificamente na seção 3.2.3, poderemos ver com mais detalhes a importância de utilizarmos ontologias como ferramenta para aplicação desta técnica.

Capítulo 3 - Expansão de Consultas

3.1 Definição:

Os sistemas de busca atuais se baseiam nas teorias da área de Recuperação da Informação e são capazes de realizar consultas por palavras-chave em um acervo de documentos. Os resultados são exibidos ao usuário por ordem de relevância, ou seja, os documentos que o sistema determina serem os mais importantes são colocados no topo da lista de resultado. Para definir a relevância de um documento são levados em consideração diversos fatores, como por exemplo, o número de vezes que um termo aparece no acervo, porém a semântica dos termos presentes nos documentos não é levada em conta. Segundo (BRAGONHOLO; CAMPOS; ELIAS, 2008) esse fato faz com que, na maioria das vezes, resultados importantes não sejam retornados para o usuário, isso acontece quando um documento não contém o termo utilizado na busca, mas sim um sinônimo ou um termo mais genérico. Por exemplo, se um usuário busca pelo termo “livro”, documentos que contenham o termo “romance”, mas não “livro”, não serão retornados para o usuário. O sistema de busca não sabe que “romance” é um tipo de “livro”. Para melhorar esse cenário, técnicas de expansão de consultas podem ser utilizadas.

A expansão da consulta é o processo de reformulação de uma consulta para melhorar o desempenho dos sistemas de busca. Com o uso dessa técnica, tenta-se estender a cobertura da recuperação de documentos relevantes, que não necessariamente contêm as mesmas palavras que a consulta. No exemplo anterior, a consulta poderia ser expandida para consultar “livro” e “romance” ao invés de apenas “livro”. Porém as técnicas de expansão de consulta podem ter efeitos negativos sobre o resultado da pesquisa, ou seja, retornar um número menor de documentos relevantes. Vários parâmetros influenciam no resultado da aplicação dessa técnica, como a quantidade de termos que devem ser adicionados à consulta inicial e de onde retirar os termos que serão adicionados, (BRAGONHOLO; CAMPOS; ELIAS, 2008) acreditam que as relações semânticas presentes em uma ontologia podem ajudar na expansão de consultas, aumentando a cobertura e a precisão do sistema.

De acordo com (BOGHAL; MACFARLANE; SMITH, 2006) o processo de adição de termos pode ser manual, automático ou assistido pelo usuário. A expansão de consulta manual

depende da experiência do usuário para tomar decisões sobre quais os termos a serem incluídos na nova consulta. No caso da expansão de consulta automática, ponderações são calculadas para todos os termos possíveis de serem adicionados, os que possuem a maior ponderação são adicionados à consulta inicial. Existem diversas funções de ponderação, portanto, a relevância dos documentos contidos no resultado depende da função escolhida. Com a expansão da consulta assistida pelo usuário, o sistema gera os termos possíveis de serem adicionados e o usuário seleciona qual desses incluir.

Segundo (BRODER, 2002) e (KANG; KIM, 2003) existem três classificações de consultas, são elas: de navegação, informação e transacionais. Consultas de navegação são tipicamente usadas para encontrar páginas, as de informação são utilizadas para encontrar informações relevantes para um determinado tópico e as de transação permitem ao usuário localizar um site que ofereça serviços, como compras, por exemplo. De acordo com Broder, o melhor tipo de consulta para expansão é a de informação. A razão para isso é que essas consultas são menos específicas do que as outras.

3.2 Técnicas de Expansão de Consultas:

Existem diversas formas para se implementar a expansão de consultas, cada qual com suas vantagens e desvantagens dependendo do domínio de aplicação. A principal diferença entre as implementações é de onde retirar os termos a serem adicionados na consulta inicial e como selecionar tais termos.

Também existem outras diferenças, como o número de termos a serem adicionados, se usa ou não alguma técnica de desambiguação, a função de ponderação utilizada, entre outras. Essas implementações são melhores explicadas nas seções seguintes.

3.2.1 Retorno Relevante

A técnica do retorno relevante é uma das mais utilizadas. Nessa técnica os termos a serem adicionados na consulta são escolhidos a partir dos resultados da consulta inicial, a fim de formar uma segunda consulta que retornará resultados mais relevantes para o usuário.

Essa técnica possui três variações: a explícita, a pseudo e a implícita. Na técnica explícita a consulta inicial retorna um conjunto de documentos (geralmente títulos ou resumos). A partir desta tela, o usuário julga os documentos como relevantes ou não relevantes. As palavras chaves dos documentos selecionados como relevantes são adicionados à consulta inicial. Em contrapartida termos que aparecem em documentos marcados como não relevantes são retirados da consulta inicial. (GEIST; SCHWEIGHOFER, 2008) pensam que o potencial de realimentação dessa técnica encontra-se no contexto da pesquisa dos diferentes usuários.

Já a técnica pseudo obtém o conjunto de documentos retornados pela consulta inicial e, sem a ajuda do usuário, determina que os primeiros N documentos serão considerados como relevantes.

A técnica implícita visa melhorar a precisão e o retorno da informação, utilizando as ações do usuário para inferir a relevância ou a não relevância dos documentos. Muitos sinais demonstrados pelo comportamento dos usuários podem contribuir para uma avaliação probabilística da relevância do documento. Os principais sinais do usuário utilizados por essa técnica são os tipos de sites que o usuário acessa e principalmente quais documentos ele acessou em alguma pesquisa realizada.

A determinação de relevância obtida pelo método explícito é mais precisa, porém a técnica implícita é mais facilmente aplicada já que não exige esforço algum do usuário.

O grande problema da técnica do retorno relevante é quando o resultado retornado pela consulta inicial já não apresenta documentos que são relevantes para o usuário, pois nesse caso os termos que serão adicionados à consulta inicial só tendem a piorar o resultado da pesquisa.

3.2.2 Expansão Probabilística

O método probabilístico geralmente é baseado no cálculo de co-ocorrências de termos nos documentos e na seleção de termos que são mais relacionados aos termos da consulta inicial.

De acordo com (ANDREOU, 2005) existem os métodos probabilísticos globais e locais. Os métodos globais extraem suas estatísticas de toda a coleção de documentos e seus cálculos podem ser realizados off-line.

Técnicas locais extraem as estatísticas dos primeiros N documentos retornados por uma consulta inicial. Todos os cálculos são realizados on-line, após a entrada da consulta inicial e antes da apresentação do resultado ao usuário.

O grande problema desses métodos é que não levados em consideração a semântica dos termos, o que pode afetar negativamente o resultado da consulta. Uma alternativa à expansão probabilística são os métodos ontológicos que utilizam relações semânticas para a seleção dos termos. Tais métodos serão estudados mais a frente.

3.2.2.1 Tipos de Métodos Probabilísticos

A técnica da co-ocorrência se baseia na proximidade de dois ou mais termos dentro de um documento pertencente ao domínio considerado, adicionando à consulta termos que ocorrem frequentemente em conjunto com os termos da consulta original. De acordo com (FARHOODI et al, 2009) uma das desvantagens dessa técnica é a probabilidade de deteriorar a precisão, pois não é considerada a semântica dos termos durante a verificação das ocorrências. Por exemplo, se os termos “computador” e “software” aparecerem juntos várias vezes em um documento e acontecer o mesmo para os termos “hardware” e “software” e não ocorrer essa mesma situação para os termos “hardware” e “computador”, a relação entre essas duas palavras seria muito pequena ou até mesmo nenhuma, o que não é verdade se for levado em consideração o contexto semântico das palavras.

Outro método é descrito em (HANG et al, 2002). Tal método utiliza os logs das pesquisas já realizadas para realizar a expansão de consultas. Nessa abordagem os termos adicionados são as palavras chaves dos documentos que foram retornados e com alta taxa de acesso em consultas do mesmo domínio. No entanto para utilizar essa técnica é necessário que o sistema possua uma base de logs confiável e com um volume de informações considerável, o que impede a utilização dessa abordagem em sistemas que ainda estão em fase de desenvolvimento, ou possuem pouco tempo de uso. O foco dessa abordagem é realizar melhorias em sistemas que já estão em uso há algum tempo.

3.2.3 Outras Técnicas

A técnica da derivação consiste em adicionar á consulta inicial, variações léxicas de termos presentes na consulta dada pelo usuário. Essas variações são geradas por adição ou remoção de prefixos e sufixos conforme o caso, por exemplo, uma consulta que possui o termo “justo” poderia ser expandida adicionando os termos “injustos”, “injusto”, “justamente”, “injustamente”, “justíssimo” e outros mais.

Já a expansão de consultas usando clusters é uma técnica que foi estudada por muitos pesquisadores durante o final dos anos 60 e início dos anos 70. Nessa técnica documentos semelhantes são colocados em um cluster, e faz-se a suposição que esses documentos são relevantes para todas as pesquisas referentes ao domínio do cluster.

Se os termos da uma consulta forem mapeados para um ou mais clusters, os termos-chaves do(s) cluster (s) são usados para expandir a consulta. Segundo (BOGHAL; MACFARLANE; SMITH, 2006) essa técnica não se mostrou muito útil quando se tem clusters pequenos resultantes de pequenas coleções de documentos ou pequenas diferenças no vocabulário entre documentos relevantes e não relevantes. Outro problema com o agrupamento é que ele trabalha com a hipótese de que um termo só pode pertence a um cluster, quando existem termos ambíguos na consulta os termos adicionados podem não ter relação alguma com os termos da consulta original.

3.2.4 Expansão Utilizando Ontologias

De acordo com (BOGHAL; MACFARLANE; SMITH, 2006) ontologias têm sido utilizadas para uma série de tarefas na área de recuperação de informação, tais como sumarização de temas, desambiguação do sentido das palavras, indexação, classificação de texto, recuperação de informações de diferentes línguas, e para a expansão da consulta. Neste trabalho o foco principal é a expansão de consulta.

Ainda de acordo com (BOGHAL; MACFARLANE; SMITH, 2006) quando comparamos as técnicas de expansão de consultas que utilizam ontologia com aquelas que não utilizam, vemos que existem algumas semelhanças. Ambos os conjuntos de técnicas ainda

precisam adquirir um retorno seja ele manual, automático ou semi-automático. Outra semelhança entre os dois conjuntos de abordagens é que os novos termos sugeridos ainda precisam ser ponderados para o algoritmo de classificação. As suas abordagens são baseadas no coletivo já que as outras técnicas dependem da informação de uma comunidade de usuários que compartilham do mesmo interesse e uma ontologia é geralmente uma representação "coletiva" de um domínio que tem sido derivado de uma comunidade de usuários especialistas do domínio. A principal diferença é que as técnicas de expansão de consulta ontológicas vão obter os termos sugeridos a partir de um conhecimento modelado e as outras técnicas confiam em ter um conjunto razoável de documentos relevantes.

Ontologias vêm sendo utilizadas como ferramenta para as técnicas de expansão de consultas desde o início dos anos 90, a WordNet foi e ainda continua sendo uma ontologia geral muito popular usada na área de expansão de consultas, isso pode ser comprovado na seção 3.3.2.1.

O problema com ontologias de domínio geral, tais como WordNet, é sua ampla área de cobertura, aumentando a probabilidade de existirem termos ambíguos na ontologia, o que geraria um problema para a expansão. Para pesquisas mais restritas, ontologias de domínio específico são a melhor escolha, pois essas ontologias foram construídas em aplicações de diferentes áreas como direito, medicina, arqueologia, agricultura, geografia, multimídia, negócios, economia, história, apenas para citar alguns, portanto tem uma probabilidade muito pequena de conter termos ambíguos.

Muitas aplicações expandem as consultas utilizando sinônimos e hiperônimos de termos da consulta inicial, porém (NAVIGLI; VELARDI, 2003) argumentam que expansão de consultas que utilizam sinônimos e hiperônimos têm um efeito limitado sobre o desempenho da recuperação da informação na web. Eles sugerem que outros tipos de informações semânticas derivadas a partir de uma ontologia são mais eficazes como definições dos conceitos. Isto se deve ao fato de que palavras no mesmo domínio semântico e no mesmo nível de generalidade são as melhores candidatas para a expansão.

(ANDREOU, 2005) diz que o primeiro passo para aplicar a técnica de expansão de consultas é desambiguar os termos da consulta inicial e o passo seguinte é adicionar à consulta os conceitos relacionados na ontologia. Em uma ontologia um conceito é frequentemente descrito por vários sinônimos e em alguns casos o inverso também é verdade, um termo é usado para descrever vários conceitos. Por exemplo, na WordNet "humano" é

tanto um irmão do "animal" abaixo de "organismo" e um hipônimo de "animal". Os métodos ontológicos não distinguem tais casos, se uma versão do conceito é um hipônimo de um conceito da consulta, esse conceito é utilizado na expansão como se fosse um termo sem ambiguidade. Nesse caso deve ser feita a desambiguação para que o termo corresponda a somente um conceito ontológico. Ainda existe uma discussão sobre a melhor maneira de realizar essa desambiguação, mas não é objetivo desse trabalho.

Depois de retirada a ambiguidade, os termos são selecionados através de relações diretas na ontologia. Usualmente tipos específicos de relações são verificados (“sinonímia”, “sinonímia e hiponímia”, “sinonímia, hiponímia e hiperonímia”), juntos com um método que combina diversas relações.

(ANDREOU, 2005) propôs um cálculo da medida de similaridade semântica, nesse cálculo dois termos são dados para uma função que retornará um valor. Essa função pode usar a distribuição dos termos ao longo de documentos ou da ontologia. Usando ontologia a função pode levar em conta o caminho entre os termos da ontologia ou a definição dos termos, o melhor método faz uma mescla entre essas duas abordagens.

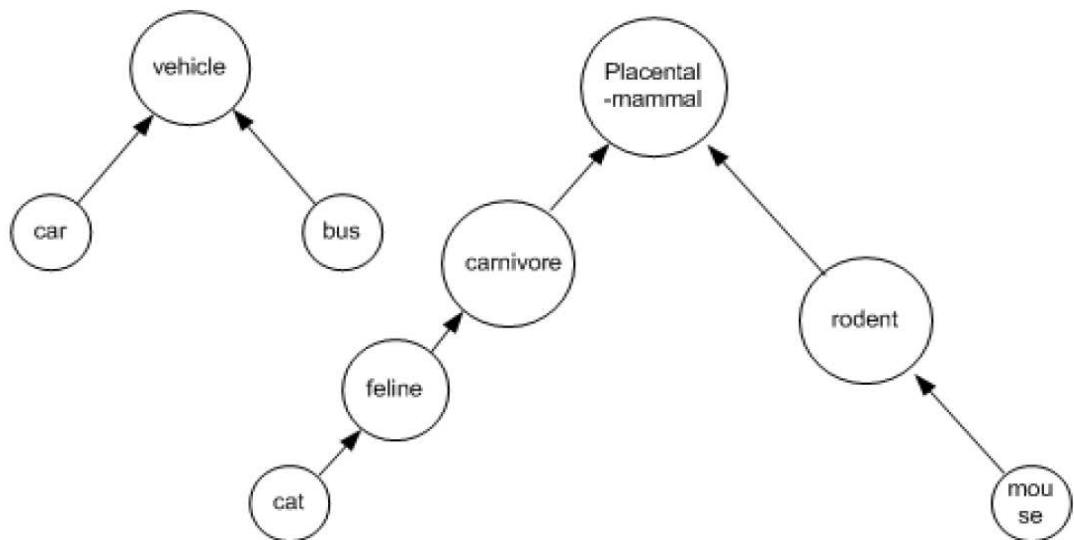


Figura 4: Relações taxonômicas na Word Net

Fonte: ANDREOU, 2005, pág. 17

O método estrutural mais simples usa a relação *é_um* e calcula a similaridade de acordo com o tamanho do caminho. Como podemos ver na figura 4 *car é_um vehicle* e *bus é_um vehicle*, nesse caso o comprimento do caminho é um, pois existe *car- vehicle -bus*. Em contrapartida *cat é_um feline*, *feline é_um carnivore*, *carnivore é_um placental-mammal*, *rodent é_um placental-mammal* e um *mouse é_um rodent*. Esse caminho possui comprimento quatro, pois possui quatro nós intermediários.

O método mais simples baseado na definição calcula a similaridade simplesmente contando o número de palavras comuns entre as definições dos termos e assumindo esse valor como a pontuação.

3.3 Aplicações Existentes

3.3.1 Aplicação do Retorno Relevante

Surf Canyon é uma extensão de browser que monitora o comportamento do usuário ao navegar na internet. Essa extensão tem como base vários estudos como (KELLY; TEEVAN, 2003) que diz que um clique através de um resultado específico retornado por um sistema de busca pode ser interpretado como uma indicação do usuário sobre a relevância do documento e (HEMBROOKE et al, 2005) que concluíram que se o usuário clica em um documento retornado por uma pesquisa resultado depois de ter ignorado outros documentos. Analisando estas pesquisas, pode-se considerar que em 80% dos casos o documento clicado é mais relevante do que os ignorados. O download dessa extensão pode ser realizado no site <http://www.surfcanyon.com/extension.jsp>.

A tecnologia Surf Canyon modifica dinamicamente a página de resultados de sistemas de busca comerciais como o Google, Yahoo e Live Search. O usuário entra em um sistema de busca web e realiza sua pesquisa, assim como faria para qualquer pesquisa (usando a barra de busca embutido no navegador ou navegando o URL do mecanismo de busca). Após a consulta inicial, o sistema de pesquisa retorna a página de resultados exatamente como ela seria se o Surf Canyon não tivesse instalado, com apenas uma pequena modificação, pequenos ícones são adicionados ao lado do título do link para cada resultado da pesquisa (ver Figura 5).

A figura 5 mostra uma pesquisa realizada no Google em que a consulta de entrada foi "expansão de consultas". Foram retornados 10 resultados (apenas seis são mostrados na Figura 5) dos 1.670.000 resultados que satisfazem a consulta. O usuário então selecionou o primeiro, o terceiro e o quarto resultados da pesquisa, portanto esses resultados podem entrar em futuras listas de resultados recomendados.

Com base no comportamento do usuário já monitorado pelo Surf Canyon, é recomendado um resultado para a pesquisa, que de acordo com o algoritmo utilizado considera esse resultado mais pertinente às necessidades do usuário neste determinado momento que os resultados apresentados pelo Google. As recomendações podem ser geradas de forma automática como no exemplo da figura 5 ou serem geradas quando um usuário clica no ícone ao lado do título do resultado. Assume-se que essa seleção indica que o usuário entende que este documento é semelhante a, mas não exatamente o que o usuário está procurando.

The screenshot shows a Google search interface with the query "expansão de consultas". The search results are as follows:

- Result 1:** "Recuperação de Informação e Expansão Automática de Consulta com ..." (PDF/Adobe Acrobat). Description: "Os resultados indicam que a expansão de consulta, com a abordagem utilizada, traz ... busca com e sem expansão automática de consulta, e apresentamos os ...". URL: www.inf.pucre.br/~gonzalez/docs/expansconsul.pdf.
- Surf Canyon recommendation:** "Surf Canyon recomenda 1 resultado buscado: Ontologias". Description: "Reflete a estrutura conceitual de cada consulta à Web ... O exemplo apresentado demonstrou uma melhoria de 30% pela expansão da consulta inicial ...". URL: www.di.ufpe.br/~fab/publications/slides.ppt.
- Result 2:** "Expansão de Consultas Utilizando Indexação Semântica Latente" (Adobe PostScript). Description: "de expansão de consultas e formular uma consulta mais elaborada a partir da ... expansão de consultas e util porque a linguagem natural permite que as ...". URL: homepages.dcc.ufmg.br/~nivio/papers/sbbd02.ps.
- Result 3:** "Sistema FOQuE para Expansão Semântica de Consultas baseada em ..." (PDF/Adobe Acrobat). Description: "tipos de expansão de consultas com o intuito de recuperar resultados semanticamente ...". URL: www2.dc.ufscar.br/~cristiane_yaguinuma/.../DissCAY.pdf.
- Result 4:** "Expansão Automática de Consultas utilizando Ontologias Automatic ..." (PDF/Adobe Acrobat). Description: "expansão automática de consultas em Sistemas de Recuperação de Informação. Na próxima seção é apresentado o algoritmo para expansão de consultas, ...". URL: www.uff.br/ontologia/artigos/35.pdf.
- Result 5:** "Expansão de consultas com realimentação e pseudo realimentação de ..." (Web page). Description: "Últimos Artigos, Teses e Dissertações sobre Expansão de consultas com realimentação e pseudo realimentação de relevantes em um sistema que utiliza o modelo ...". URL: www.radarcencia.org/.../expansao-de-consultas.../o2ScYKO1L3WmYzWlYGVkAQR=/.
- Result 6:** "Índice Nacional SCPC de Crédito ao Consumidor registra forte ..." (Web page). Description: "14 set. 2010 ... De acordo com o economista-chefe da ACSP, Marcel Solimeo, os principais fatores que explicam a forte expansão das consultas aos SCPCs/SPCs ...". URL: www.administradores.com.br/.../expansao.../37914/.

Figura 5: Um screenshot de uma pesquisa no Google com o Surf Canyon instalado

3.3.2 Aplicações que Utilizam Ontologia

3.3.2.1 Utilizando a Ontologia WordNet

A) Em (GONZALO et al, 1998) foi retirada a ambiguidade dos termos da consulta original de forma manual e foram adicionados termos à consulta original de acordo com o conjunto de sinônimos da WordNet, segundo ele essa técnica melhorou em 29% o retorno da informação.

- B)** Em (NAVIGLI; VELARDI, 2003) a ontologia é utilizada para extrair o domínio semântico de uma palavra e em seguida, a consulta é expandida usando a co-ocorrência de palavras. A eficácia do uso de ontologias para melhorar os resultados depende do tipo e do comprimento da consulta. Eles concordam com a visão de que a expansão da consulta é apropriada para consultas com poucos termos. Seus experimentos utilizaram a ontologia WordNet 1.6 e o Google como sistema de busca. No algoritmo proposto uma rede semântica é criada para cada sentido da palavra, depois é realizada a interseção entre as redes relevantes e então uma pontuação é atribuída de acordo com o número de nós em comum. Nós em comum, são aqueles na rede semântica que podem ser alcançados por ambos os centros da rede semântica através de caminhos dirigidos. Os resultados experimentais mostraram uma melhora em relação as consultas não expandidas.
- C)** (BAZIZ et al, 2005) usaram uma pequena coleção de documentos. O conteúdo do documento é representado através de uma rede semântica ótima chamada de núcleo semântico. Conceitos são extraídos da WordNet e, em seguida, globalmente desambiguados com referência aos termos do documento para produzir a rede semântica ótima. As medidas de similaridade entre os nós conectados são os pesos das ligações. Quatro medidas de similaridade são usadas, das quais as três primeiras são baseadas na relação “é-um” e a quarta está baseada na definição do conceito.

3.3.2.2 Utilizando Ontologias de Domínio Específico

- A)** (SCHWEIGHOFER; GEIST, 2008) realizaram um trabalho usando uma ontologia do domínio legal. O modelo deles não pode substituir, mas pode suplementar os sistemas já existentes de recuperação de informação na área legal. Como a qualidade da consulta digitada pelo usuário é o principal problema então a melhoria da consulta foi o primeiro passo.

Uma ontologia léxica foi construída provendo uma base de conhecimento com 5500 termos, definições e relações entre os conceitos, esses componentes recebem pesos diferenciados. Os termos da consulta inicial são pesquisados na ontologia e recebem pesos, o caso mais simples consiste em procurar por um sinônimo, se o termo

é encontrado e existe uma relação de sinonímia então o peso 1 é atribuído. Mais complicado é quando existem vários subtermos, nesse caso é atribuído o peso 0,5, todos os termos significativos na definição são selecionados e recebem peso 0,25. Todos os pesos atribuídos a um mesmo termo são somados, então termos que possuem a soma maior do que 1 receberam peso 1, termos que possuem soma maior do que 0,5 e menor do que 1 também receberam peso 1 e termos que possuem outros valores de soma não serão levados em consideração.

- B)** (NAVIGLI; VELARDI, 2003) colocaram em prática um método de expansão usando termos que constam nas definições dos conceitos desambiguados e relataram uma melhora significativa em relação aos métodos probabilísticos. No entanto, em alguns poucos casos termos que estavam na definição de algum termo piorou o resultado da consulta, por exemplo, em uma consulta sobre "uniformes nas escolas públicas", onde a definição de escola pública é “uma escola gratuita sustentada pelos impostos e dirigida por um conselho de escola”, a palavra "imposto" gera resultados indesejados ao usuário.
- C)** Em (BARTH; TIMOSZCZUK, 2008) cada termo da consulta submetida pelo usuário é pré-processado tendo em vista a remoção de acentos, de letras maiúsculas e a redução ao singular. Em seguida os termos são confrontados com a ontologia e seus sinônimos. Se os termos ou um de seus sinônimos for encontrado na ontologia todos são conectados pelo operador OR que irá formar uma expressão que vai substituir o termo original na consulta. Todos os termos expandidos ou não são conectados pelo operador AND formando a consulta expandida.

O usuário pode escolher focalizar ou generalizar sua consulta. Na generalização os termos da consulta são substituídos por seus sinônimos e termos correspondentes as superclasses e seus sinônimos. Na focalização são adicionados seus sinônimos e termos correspondentes as subclasses e seus sinônimos.

- D)** (FARHOODI et al, 2009) implementaram uma técnica de expansão de consultas tendo como base uma ontologia da Wikipédia persa, a Wikipédia apresenta vários artigos que definem e descrevem uma entidade e possui links para outras páginas dentro e fora da Wikipédia. O mais importante é a categorização e a ligação dos conteúdos, páginas da Wikipédia são pertencentes a uma ou mais categorias, a idéia central foi associar as páginas da Wikipédia a tópicos ontológicos. Os links entre as páginas

mostram a relação semântica entre várias entidades, ou seja, os documentos são organizados de maneira hierárquica, portanto a Wikipédia pode ser considerada como um gráfico no qual conceitos são ligados um ao outro. Essas características foram utilizadas para criar uma ontologia.

Ainda segundo (FARHOODI et al, 2009) a expansão de consultas envolve técnicas como:

- Encontrar sinônimos
- Realizar derivações das palavras
- Corrigir erros de grafia
- Ponderar os termos na consulta original

Cada página da Wikipédia consiste das seguintes partes que foram usadas para construção da ontologia:

- Título: introduz o conceito que o conteúdo da página descreve
- Palavras chaves: algumas páginas possuem um número de palavras chaves que são relacionadas ao título da página
- Texto: o texto de cada página descreve seu título, alguns dos conceitos existentes no texto são ligados a páginas com o mesmo título. Por exemplo, se o título da página é “chip”, conceitos como “capacitor”, “resistor” e “transistor” aparecem no texto como links para páginas com mesmos títulos
- Veja também: essa parte introduz outras páginas com tópicos relacionados com a página corrente
- Categoria: essa parte indica a categoria da página, cada página pode ser classificada em uma ou mais categorias.

As ligações entre as páginas são significativas e mostram seus relacionamentos conceituais, com essas características foi possível construir uma ontologia com relacionamentos do tipo “é-relacionado-a”.

Para construir a ontologia é muito importante considerar um sistema de pesos correto para diferentes relações entre as páginas. No algoritmo proposto três variações de relacionamentos foram propostas:

- Relacionamento entre os títulos e as palavras chaves
- Relacionamento entre o título e os conceitos que aparecem como link na página
- Relacionamento entre o título e os links do “Veja também”

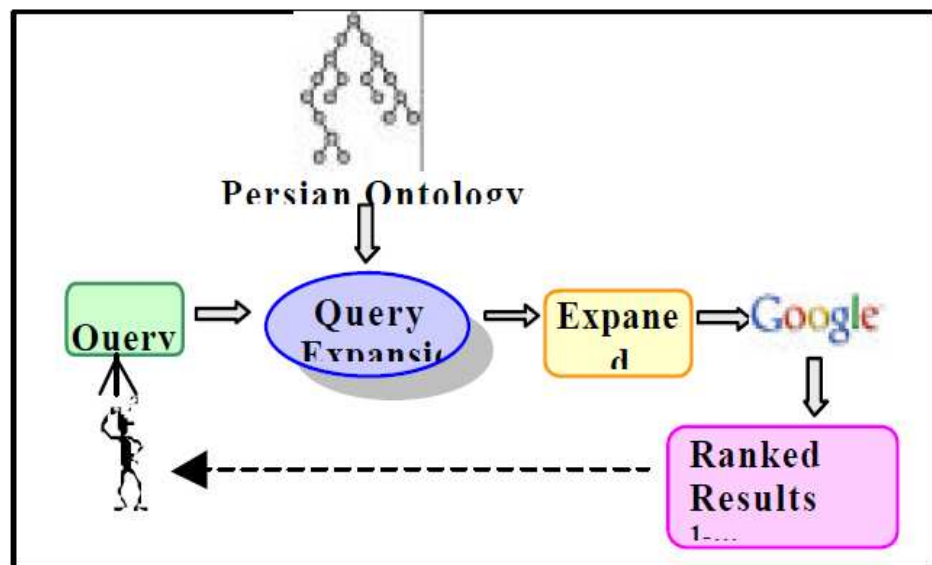


Figura 6: Processo de expansão de consultas

Fonte: FARHOODI et al, 2009, pág. 4

Como mostrado na Figura 6 o usuário entra com a consulta, essa é processada e logo após são verificados os conceitos existentes nela. Os conceitos relacionados da consulta são retirados utilizando a ontologia e considerando o peso das relações e finalmente a consulta expandida é dada como entrada no mecanismo de busca.

Para expandir a consulta, foi utilizado o método de vetor. O vetor de consulta foi criado com base nos conceitos existentes na consulta e o tamanho comprimento desse vetor é equivalente ao número de conceitos existentes na ontologia. O índice de cada elemento no vetor da consulta representa um conceito. Se o elemento existe na consulta então o valor do seu campo correspondente recebe 1, caso contrário recebe 0.

A ontologia foi representada em forma de matriz onde os índices das linhas e colunas representam os conceitos e o valor do campo da matriz representa o peso da relação entre 2 conceitos. Foi utilizada a seguinte fórmula para realizar a expansão de consulta:

$$Q_e = Q + \alpha QO + \beta QO^2 + \gamma OQ + \delta O^2Q ;$$

$$\alpha, \beta, \gamma, \delta \geq 0;$$

$$\alpha + \beta + \gamma + \delta = 1$$

Q = vetor da consulta

Q_e = vetor da consulta expandida

O = matriz da ontologia

$\alpha, \beta, \gamma, \delta$ = coeficientes definidos para mostrar a importância da distância no gráfico entre os conceitos e o tipo de ligação existente entre eles.

Capítulo 4 – Estudo de Caso: Uso de Técnicas de Expansão de Consultas

4.1 Características Gerais

Com o objetivo de verificar a aplicabilidade das técnicas estudadas, foi desenvolvido um protótipo que aplica a técnica de expansão de consultas. O objetivo desse protótipo é verificar na prática o resultado que tal técnica pode proporcionar e comparar com o resultado retornado por uma consulta que não foi expandida, verificando se existem mais resultados relevantes com ou sem a aplicação da técnica. Na figura 7 é apresentada uma arquitetura simplificada do protótipo.

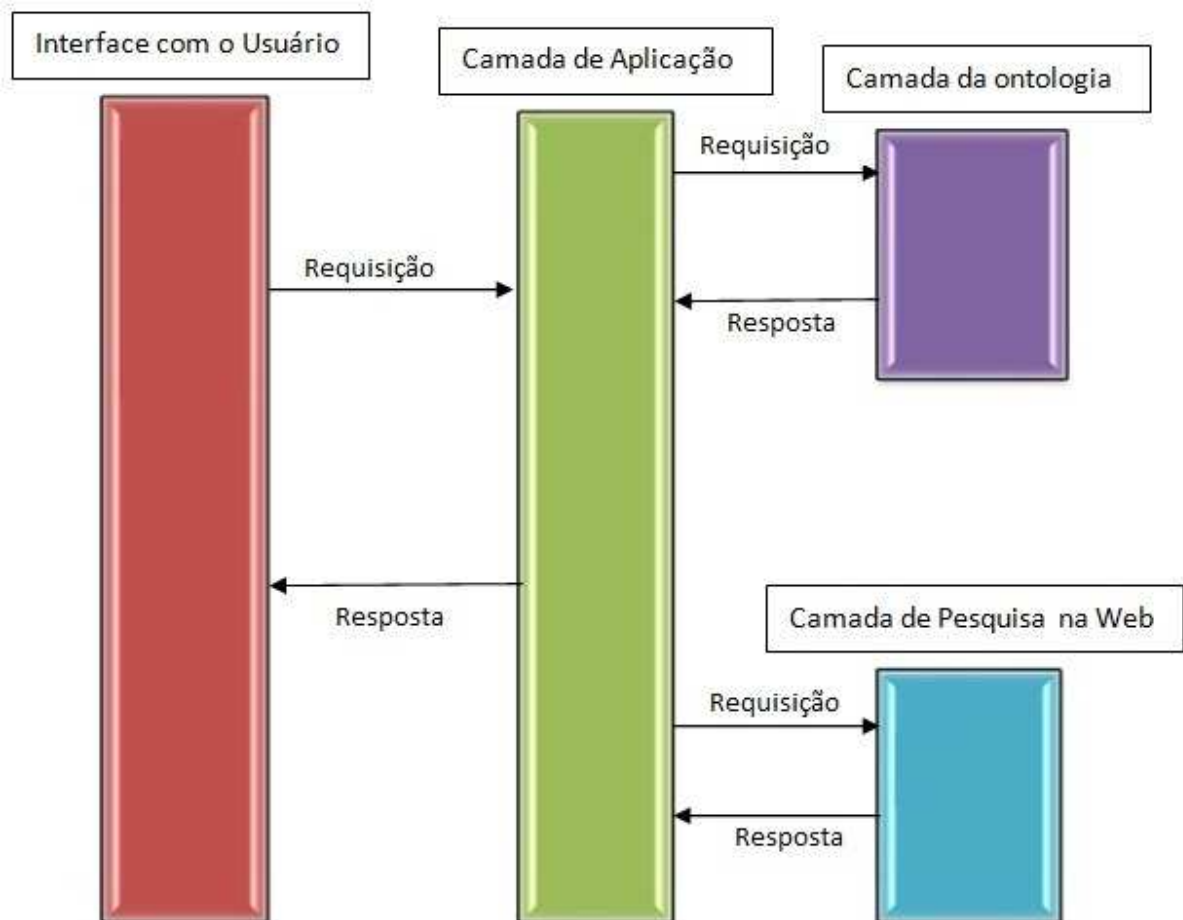


Figura 7: Arquitetura simplificada do protótipo

O protótipo possui um campo para que o usuário forneça os termos que ele deseja consultar. Com os termos da consulta, é estabelecida uma conexão HTTP com o servidor do Google para que seja realizada a consulta através desse mecanismo de busca. O resultado da consulta é mostrado ao usuário na forma de links. Paralelamente, o aplicativo aplica a técnica da expansão de consultas na consulta fornecida pelo usuário e envia a consulta expandida para o Google da mesma forma que foi feito com a consulta inicial. Considerando os dois resultados, o usuário poderá definir qual técnica retornou mais documentos relevantes para o tópico pesquisado.

Neste estudo de caso, a consulta fornecida pelo usuário deverá ser relacionada ao domínio da ecologia. Isso se deve ao fato de que a ontologia escolhida para realizar a expansão pertence a esse domínio.

4.2 Ontologia para Ecologia

A ontologia selecionada para o estudo de caso faz parte do projeto Bloc-Eco (<http://www.inf.pucrs.br/~ontolp/downloads-eco.php>) que se destina a criação de uma base de conhecimento com informações ontológicas para termos de Ecologia na língua portuguesa. O endereço da página do projeto é <http://www.nilc.icmc.usp.br/nilc/projects/bloc-eco.htm> acessado dia 12/11/2010 . O download da ontologia escolhida pode ser realizado no seguinte endereço <http://www.inf.pucrs.br/~ontolp/downloads/Projeto%20Bloc-Eco/EcoCom01.rdf-xml.owl> acessado dia 12/11/2010 .

Tal ontologia possui uma lista de conceitos pertencentes ao domínio da ecologia. Os conceitos possuem uma descrição em português, e a informação de qual seria o seu equivalente em italiano, porém essas propriedades não foram exploradas na expansão de consultas, no contexto do protótipo, cujo objetivo é apenas de verificar a viabilidade da proposta. A estrutura da ontologia pode ser vista na figura 8.



Figura 8: Ontologia utilizada na aplicação

Como podemos perceber os termos mais comuns na área de ecologia estão presentes na base da ontologia. Por exemplo, o termo “CORREGO”, que faz parte do dia-a-dia de muita gente, é um “HABITAT_AQUATICO_DOCE_LOTICO” que é um “HABITAT_AQUATICO_DOCE” e esse por sua vez é um tipo de “HABITAT_AQUATICO” que é uma especialização de “HABITAT” que é um tipo de “BIOTOPO”, então a partir de um termo específico conseguimos chegar a uma forma mais geral navegando através das relações.

4.3 Uso do Google

O Google foi escolhido como mecanismo de busca da aplicação devido a vários motivos, dentre eles, a velocidade com que os resultados são retornados e sua popularidade entre os usuários. De acordo com a página http://www.google.com.br/why_use.html acessada no dia 13/11/2010, o algoritmo utilizado pelo Google é o PageRank(TM), um sistema para dar notas para páginas na web. Através desse algoritmo, um link da página A para a página B é interpretado como um voto da página A para a página B. É verificado mais do que o volume de votos, que uma página recebe, também é levada em consideração a página que dá o voto. Os votos dados por páginas "importantes" pesam mais e ajudam a tornar outras páginas "importantes." Outra razão foi o fato de que com o Google, os usuários podem pesquisar mais de um bilhão de URLs.

4.4 Desenvolvimento

4.4.1 Ambiente de Desenvolvimento

A IDE escolhida para o desenvolvimento da aplicação foi o Eclipse, pois possui código aberto e provê suporte ao Java (linguagem utilizada para o desenvolvimento do protótipo), além de permitir a adição de extensões à plataforma e disponibilizar funcionalidades para o desenvolvimento e gerenciamento do software em todo seu ciclo de vida. O Eclipse possui um portal no endereço <http://www.eclipse.org/> acessado dia 09/11/2010 onde pode ser

encontrada documentação e mais informações, além de ser possível realizar o download da IDE.

4.4.2 Classes do Protótipo

Como a linguagem Java permite, o desenvolvimento da aplicação foi dividido em quatro classes: principal, HyperlinkLabel, pesquisa e processaConsulta. Essa abordagem permite que as informações de cada classe sejam encapsuladas, trazendo maior segurança. Outra vantagem é a facilitação da manutenção e de verificação de erros.

A classe principal é a responsável pela troca de informações com o usuário, além de ser a responsável pelo envio e recebimento de informações das outras classes. Essa classe possui um construtor, parte em que é montada a interface gráfica do aplicativo. A interface é bem simples, possui um JTextField que receberá a consulta do usuário e um JButton que é responsável pelo envio da consulta para a classe que realiza o processamento da consulta e a pesquisa.

Os resultados são mostrados ao cliente em forma de links, dentro de duas janelas internas, a primeira mostra os resultados da consulta inicial e a segunda mostra os resultados da consulta expandida. Também é mostrada uma mensagem informando ao usuário como que ficou a consulta expandida.

Na função de tratamento de eventos, é passada para a classe Pesquisa a consulta inicial fornecida pelo usuário, a resposta é em forma de um JSONObject, que de acordo com a página <http://www.json.org/java/> acessada dia 15/11/2010 nada mais é um texto formatado em JSON. Com o auxílio da biblioteca Json-simple, a resposta é decodificada para uma String comum contendo somente os dados que interessam à aplicação. São mostrados dez resultados para o usuário. Logo após, a consulta inicial é passada para a classe processaConsulta, essa retorna a consulta expandida que passará pelo mesmo processo aplicado na consulta inicial.

A classe HyperlinkLabel é herdada da classe JLabel nativa da linguagem Java. Ela cria um objeto HyperlinkLabel, que na verdade é um objeto JLabel que possui um tratador de eventos implementado e o atributo url que é do tipo URI.

O tratador de eventos muda o cursor do mouse toda para “HAND_CURSOR” toda vez que o mouse passa sobre o HyperlinkLabel, além disso, toda vez que o HyperlinkLabel é clicado o browser é aberto e direcionado para a página contida no atributo url, os HyperlinkLabel que já foram acessados aparecem para o usuário na cor vermelha.

A classe Pesquisa é responsável por enviar a consulta para o servidor Google e obter a resposta. O seu construtor é bem simples, e sua única função é receber a consulta por parâmetro e substituir seus espaços em branco pelo caractere “+”, assim a consulta fica no padrão aceito pelo Google.

A função realizaPesquisa cria uma conexão HTTP, e através de um método GET envia a consulta para o servidor. O método GET é utilizado porque ele utiliza a própria url para enviar dados ao servidor. Quando um formulário é enviado pelo método GET, o navegador captura as informações do formulário e as adicionam à url de destino, separando o endereço e os dados do formulário pelo caractere “?”.

Quando é realizada uma consulta no Google, que é o caso desse trabalho, é feita uma requisição utilizando o método GET. É possível copiar a url da consulta do Google e acessar outras vezes que o resultado será o mesmo, lembrando que de tempos em tempos o Google atualiza a posição das páginas, o que pode fazer que o resultado da url acessada seja diferente.

Para realizar a requisição ao servidor do Google foram utilizadas informações presentes na seguinte página http://code.google.com/intl/pt-BR/apis/websearch/docs/reference.html#_class_GwebSearch acessada dia 12/11/2010 . Nessa página é fornecido o endereço padrão das consultas ao servidor, que é <https://ajax.googleapis.com/ajax/services/search/web> .

A url enviada pelo método GET do aplicativo é a seguinte: http://ajax.googleapis.com/ajax/services/search/web?v=1.0&hl=pt-br&gl=countryBR&q=CONSULTA&rsz=filtered_cse . Portanto faz-se uma consulta em que a versão do protocolo é 1.0, a linguagem do host do aplicativo é português do Brasil, o resultado é adequado ao Brasil, a consulta é fornecida pelo aplicativo, e são retornados dez resultados.

A classe ProcessaConsulta é a responsável pela principal função do aplicativo, ou seja, ela é responsável por expandir as consultas fornecidas pelo usuário.

Essa classe possui apenas dois métodos, o construtor e o `expande`. O construtor apenas fragmenta a consulta em termos individuais que são armazenados em um vetor. O método `expande` acessa a ontologia e procura por cada elemento presente no vetor criado no construtor. Quando um termo é encontrado na ontologia a relação `subClass` é consultada, caso o termo retornado ainda não pertence à consulta então ele é adicionado. Quando um termo não é encontrado na ontologia ou até é encontrado, porém sem possui relacionamento do tipo `subClass` então é adicionado o termo “`ECOLOGIA`” à consulta, apenas para contextualizar a consulta no domínio em questão, essa adição somente é realizada se o termo “`ECOLOGIA`” não estiver presente na consulta anteriormente. Quando o usuário fornece termos repetidos na consulta inicial, esses termos são retirados restando apenas um.

O método `expande`, retorna uma `String` que é a junção de todos os termos da consulta através do caractere “`+`”.

O código fonte da classe `processaConsulta` pode ser observado na figura 9.

```

1 import com.hp.hpl.jena.rdf.model.*;
2 import com.hp.hpl.jena.util.FileManager;
3 import java.io.*;
4
5 public class processaConsulta extends Object {
6
7     private static final String arquivo = "EcoCom01.rdf-xml.owl.xml";
8     private String[] consulta;
9
10    processaConsulta(String consulta){
11        this.consulta = consulta.split(" ");
12    }
13    public String expande(){
14        int tamanho = consulta.length;
15        int tamanho2;
16        String termo = null;
17        Model model = ModelFactory.createDefaultModel();
18        InputStream in = FileManager.get().open( arquivo );
19        if (in == null) {
20            throw new IllegalArgumentException( "Arquivo: " + arquivo + " não encontrado" );
21        }
22        model.read(in, "");
23        String expandido = "";
24        String[] palavras;
25        for (int i = 0; i < tamanho; i++){
26            termo = consulta[i];
27            Resource conceito = model.getResource("http://www.owl-ontologies.com/unnamed.owl#" + termo);
28            try{
29                if ((expandido == "") || (!expandido.contains(termo + "+"))){expandido = termo + "+" + expandido;
30                    palavras = conceito.getProperty(model.getProperty("http://www.w3.org/2000/01/rdf-schema#subClassOf")).getObject().toString()
31                                                                .substring(42).split("_");
32                    tamanho2 = palavras.length;
33                    for (int j = 0; j < tamanho2; j++){
34                        if (!expandido.contains(palavras[j] + "+")) expandido = palavras[j] + "+" + expandido;
35                    } catch ( Exception e){
36                        if (!expandido.contains("ECOLOGIA" + "+")) expandido = "ECOLOGIA" + "+" + expandido;
37                    }
38                }
39            return expandido.substring(0,expandido.length()-1);
40        }
41    }

```

Figura 9: Código fonte da classe processaConsulta

4.5 Exemplos de Uso do Protótipo

A tela inicial do protótipo é bem simples e possui apenas um campo para o usuário digitar a consulta desejada e um botão para enviar a consulta, a figura 10 mostra a tela inicial do protótipo já com uma consulta fornecida pelo usuário.

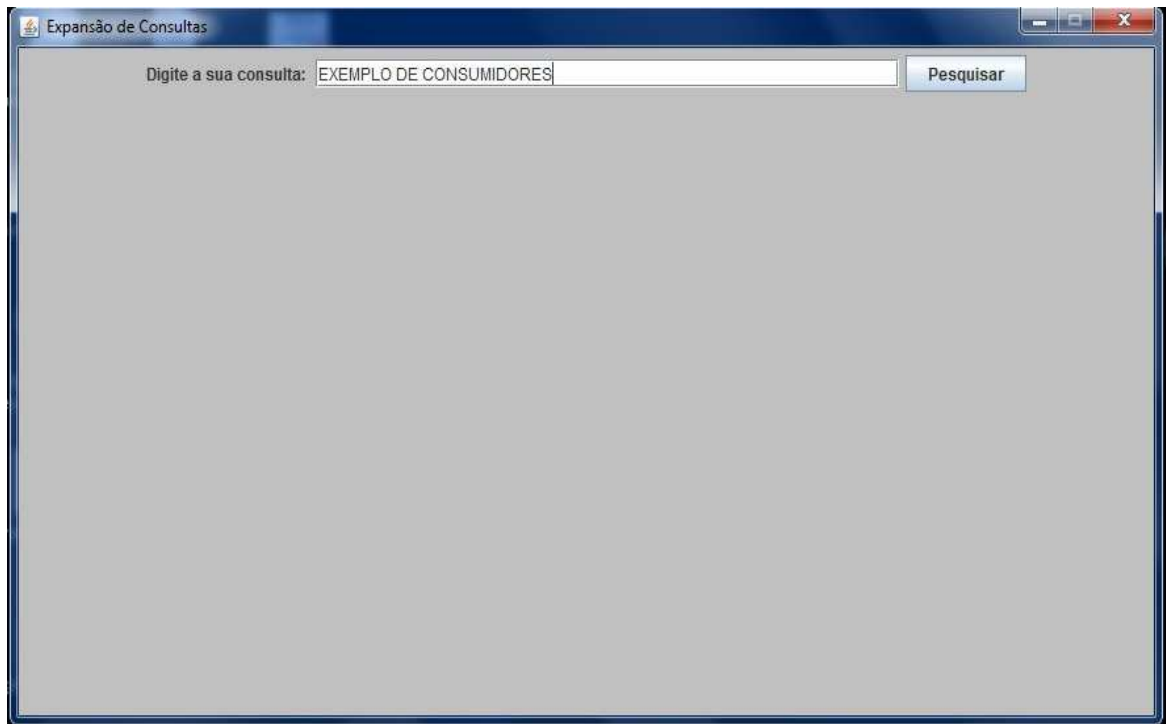


Figura 10: Tela inicial do protótipo

Como podemos ver na figura 11, a consulta fornecida pelo usuário foi “EXEMPLO DE CONSUMIDORES”. A princípio, essa é uma consulta bem ambígua, pois sem um contexto não temos como saber a qual domínio pertence o termo “CONSUMIDORES”. A figura 11 nos mostra como ficou a consulta expandida.

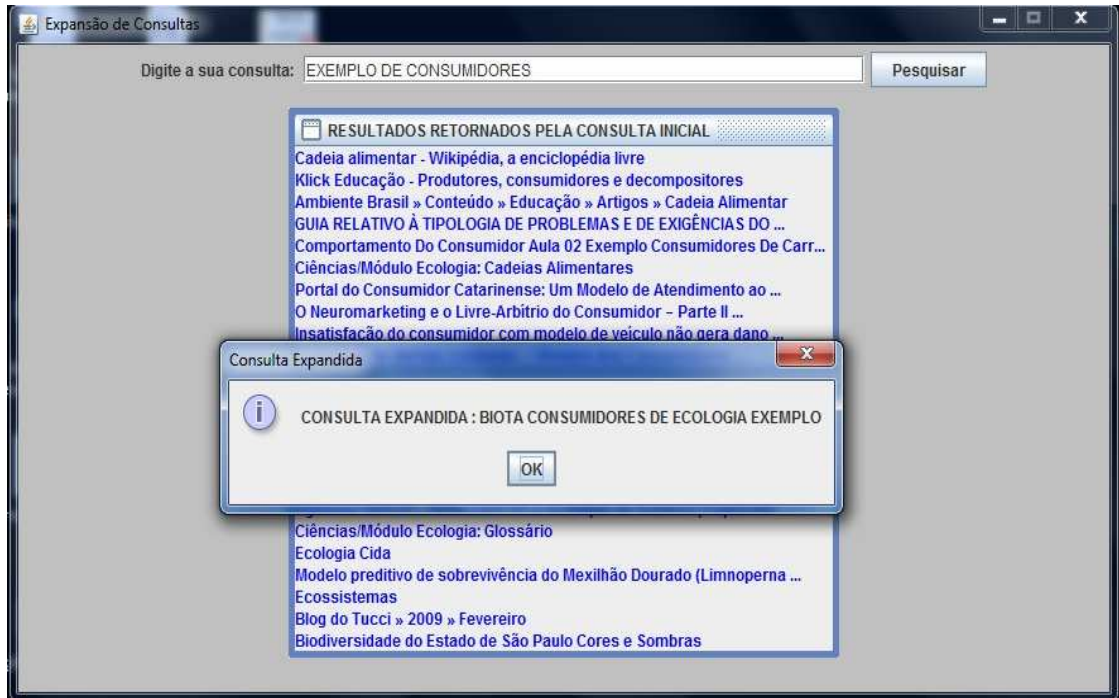


Figura 11: Consulta expandida

Conforme a figura 12 demonstra, a consulta “EXEMPLO DE CONSUMIDORES” foi expandida para “BIOTA CONSUMIDORES DE ECOLOGIA EXEMPLO”. Essa expansão ocorreu de acordo com o algoritmo presente na classe `processaConsulta`. Esse algoritmo acessa a ontologia através de métodos do framework Jena que de acordo com a página <http://jena.sourceforge.net/> acessada dia 13/11/2010 é um framework para desenvolvimento de aplicações de web semântica e procura pelo hiperônimo do termo digitado, esse hiperônimo é adicionado à consulta. Nos casos em que algum termo da consulta não possui classe na ontologia, o termo “ECOLOGIA” é adicionado à consulta. Essa expansão tem o intuito de contextualizar a consulta, nos casos em que nenhum termo da consulta possui classe na ontologia, e nos demais casos o intuito é de adicionar mais um termo que possui grandes chances de estar presente em documentos relevantes para a pesquisa, já que tal pesquisa é feita no domínio da ecologia. O primeiro termo analisado pelo algoritmo foi “EXEMPLO”, como esse termo não possui classe na ontologia e o termo “ECOLOGIA” ainda não está presente na consulta, este é adicionado à consulta. O segundo termo analisado foi “DE”, este também não possui classe na ontologia, mas como o termo “ECOLOGIA” agora já faz parte da consulta, então nenhum outro termo é adicionado. O último termo analisado é “CONSUMIDORES”, esse possui classe na ontologia, então o algoritmo acessa a ontologia e

verifica qual classe está relacionada a ela através da relação subClass, no caso é a classe “BIOTA”, portanto o termo “BIOTA” é adicionado à consulta, resultando na consulta expandida. Ainda poderia ser verificado quem é o hiperônimo de “BIOTA” a fim de expandir ainda mais a consulta, porém essa expansão não é realizada devido ao risco de generalizar demais a consulta tornando-a ineficiente. A figura 12 mostra os resultados retornados pela consulta.

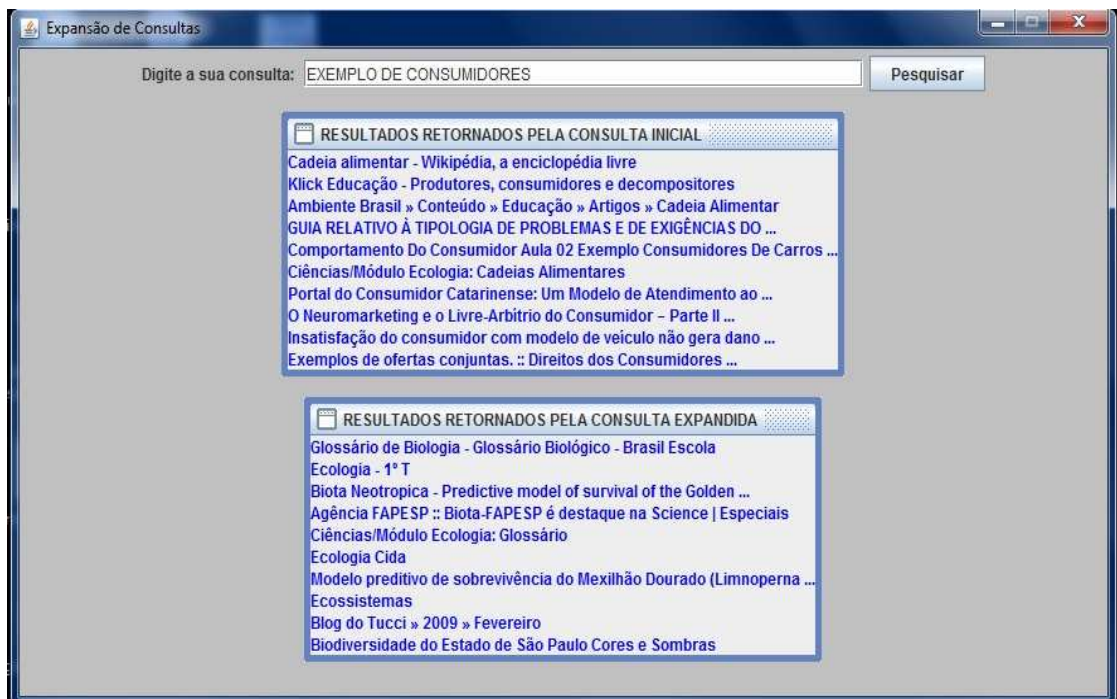


Figura 12: Resultados da consulta “EXEMPLO DE CONSUMIDORES”

Na figura 13 podemos ver que são mostradas duas janelas internas, a primeira contém os resultados retornados pela consulta inicial fornecida pelo usuário, a segunda contém os resultados retornados pela consulta expandida pelo protótipo. Fazendo uma análise dos resultados, percebemos que dos dez documentos retornados pela consulta inicial apenas quatro pertencem ao domínio de interesse, em contrapartida todos os documentos retornados pela consulta expandida pertencem ao domínio da ecologia.

Outra consulta realizada foi “ANIMAIS DA FLORESTA”, essa consulta foi submetida ao mesmo processo da consulta do exemplo anterior, portanto a expansão se deu da seguinte maneira: o primeiro termo analisado foi “ANIMAIS” como esse termo não possui

classe na ontologia e o termo “ECOLOGIA” ainda não está presente na consulta, este é adicionado à consulta. O segundo termo analisado é “DA”, este também não possui classe na ontologia, mas como o termo “ECOLOGIA” agora já faz parte da consulta, então nenhum outro termo é adicionado. O último termo analisado é “FLORESTA”. Esse possui classe na ontologia, então o algoritmo acessa a ontologia e verifica qual classe está relacionada a ela através da relação subClass, no caso é a classe “HABITAT_TERRESTRE”, como o nome da classe possui dois termos, esses são separados e adicionados à consulta. Portanto consulta expandida é “TERRESTRE HABITAT FLORESTA DA ECOLOGIA ANIMAIS”. Os resultados retornados por essa consulta são mostrados na figura 13.

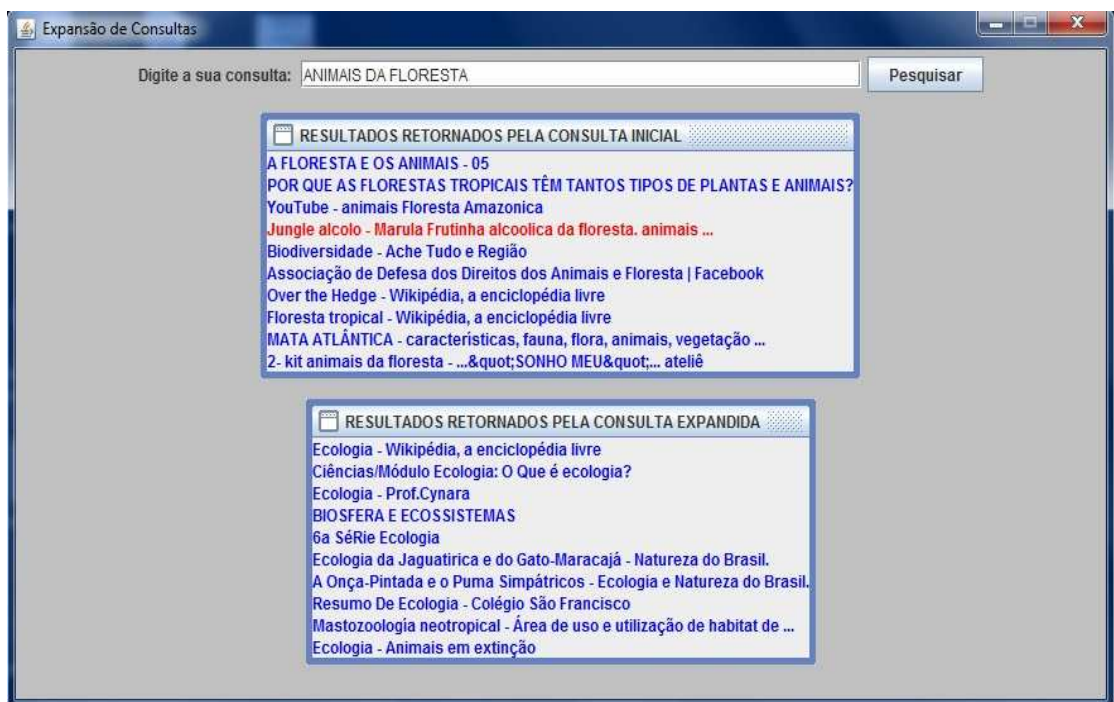


Figura 13: Resultados da consulta “ANIMAIS DA FLORESTA”

Observando os resultados mostrados na figura 13, podemos perceber que dentre os resultados retornados pela consulta inicial existem vários documentos irrelevantes para o domínio em questão como vídeos e imagens de animais, sendo grande parte de humor. Já a consulta expandida obteve como resultados documentos que falam sobre os animais e seus habitat tornando esses resultados mais relevantes para o domínio em questão.

4.6 Análise da Técnica Aplicada

Observando os exemplos da seção anterior e fazendo outras consultas é possível perceber que o protótipo desenvolvido apresentou melhorias nos resultados retornados. No entanto, devemos ressaltar que este é um protótipo bastante simples, cujo objetivo foi apenas de apresentar a viabilidade da técnica de uma maneira simplificada. Existem várias possibilidades de aperfeiçoamentos. Nos exemplos citados na seção anterior existem alguns documentos que foram retornados pela consulta inicial que podem ser considerados muito relevantes para o domínio e que não foram retornados pela consulta inicial, esse fato reitera a necessidade de aperfeiçoar a técnica aplicada.

Capítulo 5 – Conclusão

Nos dias de hoje é cada vez mais comum realizarmos alguma pesquisa na internet e não encontrarmos o que estávamos procurando. Isso se deve ao crescimento da Internet. A cada dia que passa mais e mais documentos se tornam disponíveis na grande rede. Esse crescimento dificulta o trabalho dos mecanismos de busca, que passam a ter que analisar cada vez mais documentos antes de retornar a resposta ao usuário. É nesse contexto que se desenvolveram várias técnicas para aprimorar esses mecanismos de busca. Uma delas é a expansão de consultas utilizando ontologia, objeto de estudo da presente monografia.

Foi realizado um estudo sobre ontologias, esse estudo teve um enfoque nas principais características que interferem no resultado das técnicas de expansão de consultas. Desse estudo pode concluir-se que os tipos de relações e atributos presentes em uma ontologia são importantes para o sucesso da expansão de consultas.

Uma série de técnicas de expansão de consultas foi analisada, como por exemplo, as técnicas de co-ocorrência, derivações, retorno relevante, entre outras. Um enfoque maior foi dado para as técnicas de expansão que utilizam ontologias. Estas técnicas possuem a vantagem de agregar à consulta a semântica presente nas ontologias.

Foi possível perceber que no geral essas técnicas sempre melhoram o resultado das pesquisas, porém alguns problemas podem ocorrer quando se utilizam uma ontologia de domínio geral, como a WordNet, o usuário pode estar procurando pelo termo “MANGA” relacionado a fruta e o mecanismo de busca entender que o usuário está procurando por uma parte de uma blusa, esse problema pode ser resolvido utilizando ontologias de domínio específico.

Por fim, foi desenvolvido um protótipo que aplica a técnica de expansão de consultas utilizando ontologias. A ontologia escolhida para realizar a expansão é do domínio da ecologia, e a relação explorada para a realização da expansão é a hiperonímia. O protótipo conseguiu melhorias no resultado das pesquisas, porém existem pontos falhos como não retornar alguns documentos relevantes que foram retornados sem a aplicação da técnica. Esse fato mostra que a técnica aplicada ainda tem que ser aprimorada para que retorne documentos com relevância maior para o usuário.

Pode se concluir que as técnicas de expansão de consultas que utilizam ontologias são muito promissoras e podem revolucionar a forma como os dados são pesquisados na internet em um futuro próximo.

Trabalhos futuros poderiam aprimorar a técnica utilizada pelo protótipo desenvolvido de diversas maneiras, entre elas:

- A utilização de outros tipos de relacionamentos, o que ampliaria as opções de termos a serem adicionados à consulta inicial.
- A utilização de mais de uma ontologia, o que permitiria o usuário realizar consultas pertencentes a diversos domínios.
- Para evitar conflitos entre termos presentes em mais de uma ontologia poderia ser dada uma opção ao usuário para escolher em qual domínio realizar a consulta dentre os disponíveis.
- Mesclar a técnica com métodos probabilísticos, a fim de ter documentos com maiores relevância dentre os retornados, dentre outros.

Referências Bibliográficas

ABURDENE, Patricia; NAISBITT, John. **Megatrends 2000**. 3. Ed. São Paulo : Amana-Key, 1990.

ANDREOU, Agissilaus. **Ontologies and Query expansion**. 2005. Dissertação (Master of Science) - School of Informatics, University of Edinburgh, 2005.

BARTH, Fabrício J. ; TIMOSZCZUK, A. P. . **Expansão Automática de Consultas utilizando Ontologias**. In: Seminário de Pesquisa em Ontologia no Brasil, 2008, Niterói, Rio de Janeiro, Brasil, 2008.

BAZIZ, M. et al. . **Conceptual indexing based on document content representation information context: nature, impact, and role**. In: 5th International conference on conceptions of library and information sciences, Glasgow, Reino Unido, 2005.

BERNABÉ, Flávio de Souza; LAZILHA, Fabrício Ricardo; PASCUTTI, Márcia Cristina Debalto. **Ontologias e Banco de Dados**. 1999.

BHOGAL, J.; MACFARLANE, A. ; SMITH, P. .A review of ontology based query expansion, 2006.

BRAGANHOLO, Vanessa; CAMPOS, Maria Luiza Machado; ELIAS André Bechara. **Explorando relações semânticas para expansão de consultas em acervos específicos de domínio**. In: Seminário de pesquisa em ontologia no Brasil, [2008], Niterói.

BRODER, Andrei. **A taxonomy of Web search**. [2002].

CRAMER, Mark; HARDTKE, David; WERTHEIM, Mike. **Demonstration of Improved Search Result Relevancy Using Real-Time Implicit Relevance Feedback**. In: SIGIR '09 Proceedings of the 32th annual international ACM SIGIR conference on Research and development in informaion retrieval, Boston, Massachusetts, Estados Unidos, 2009.

Class Reference - Google Web Search API - Google Code. Disponível em <http://code.google.com/intl/pt-BR/apis/websearch/docs/reference.html#_class_GwebSearch >. Acesso em 12 Nov 2010.

Eclipse – The Eclipse Foundation open source community website. Disponível em <<http://www.eclipse.org/>> . Acesso em 09 Nov. 2010.

FARHOODI, Mojgan et al. **Query Expansion Using Persian Ontology Derived from Wikipedia**, 2009.

FENSEL, Dieter; **Ontologies: Silver Bullet for Knowledge Management and Electronic Commerce**, [2001].

FREE CSS TEMPLATES. **OntoLP | Portal de Ontologias**. Disponível em <<http://www.inf.pucrs.br/~ontolp/downloads-eco.php>> . Acesso em 12 Nov. 2010.

FREYTAG, Johann-Christoph; NECIB, Chokri Ben. **Using Ontologies for Database Query Reformulation.** [2004].

GEIST, Anton; SCHWEIGHOFER, Erich. **Legal Query Expansion using Ontologies and Relevance Feedback.** [200-].

GÓMEZ-PÉREZ, A. **Evaluation of taxonomic knowledge in ontologies and knowledge bases.** In: TWELFTH WORKSHOP ON KNOWLEDGE ACQUISITION, MODELING AND MANAGEMENT, Alberta, Canadá, 1999.

GONZALEZ, Marco; LIMA, Vera Lúcia Strube. **Recuperação de Informação e Expansão Automática de Consulta com Thesaurus: uma avaliação.** [200-]

GRUBER, Thomas R. . **A Translation Approach to Portable Ontology Specifications,** 1993.

GUIMARÃES, Francisco José Zamith. **Utilização de ontologias no domínio B2C.**2002. Dissertação(Mestrado pelo Programa de Pós-Graduação em Informática da PUC-Rio) – Departamento de Informática, Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, 2002.

GONZALO, J. et al. . **Indexing with WordNet synsets can improve text retrieval.** 1998.

HANG, C. et al. **Probabilistic query expansion using query logs.** In: Proceedings of the eleventh international conference on World Wide Web, Honolulu, Hawaii, Estados Unidos , 2002.

HEMBROOKE, Helene et al. **Accurately interpreting clickthrough data as implicit feedback.** In: SIGIR '05 Proceedings of the 28th annual international ACM SIGIR conference on Research and development in informaion retrieval, Salvador, Bahia, Brasil, 2005.

JSON in Java. Disponível em <<http://www.json.org/java/>>. Acesso em 15 Nov. 2010.

KANG, In-Ho; KING, GilChang. **Query type classification for web document retrieval.** In: SIGIR '03 Proceedings of the 26th annual international ACM SIGIR conference on Research and development in informaion retrieval, Toronto, Canadá, Ontario, 2003.

KELLY, D.; TEEVAN, J.. **Implicit feedback for inferring user preference.** In: SIGIR '03 Proceedings of the 26th annual international ACM SIGIR conference on Research and development in informaion retrieval, Toronto, Canadá, Ontario, 2003.

LOPES, Lucelene. VIEIRA, Renata. **Aprendizagem de Ontologias a partir de Textos.** 2009.

MIHALCEA, R. et al . **Pagerank on semantic networks, with application to word sense disambiguation.** In: Proceedings of the 20st International Conference on Computational Linguistics, Coling, 2004.

NAVIGLI, Roberto; VELARDI Paola. **An Analysis of Ontology-based Query Expansion Strategies,** [2004?]

PAGE, L. et al. **The pagerank citation ranking: Bringing order to the web,** 1999.

Razões para se usar o Google. Disponível em < http://www.google.com.br/why_use.html >. Acesso em 13 Nov. 2010.

SPINK, Amanda et al. **Searching the Web: The Public and Their Queries.** 2001.

VILAR, Bruno Siqueira Campos Mendonça. **Processamento de consultas baseado em ontologias para sistemas de biodiversidade.** 2009. Dissertação (Mestrado em ciência da computação) – Instituto de Computação, Universidade Estadual de Campinas, Campinas, 2009.

WINSTON, M. et al. **A taxonomy of part-whole relations :** Cognitive Science, 1987.

YAGUINUMA, Cristiane Akemi. **Sistema FOQuE para Expansão Semântica de Consultas baseada em Ontologias Difusas.** 2007. Dissertação (Mestrado em ciência da computação) – Centro de Ciências Exatas e Tecnologia, Universidade Federal de São Carlos, São Carlos, 2007.

ZAVAGLIA, Cláudia. **BLOC-Eco.** Disponível em <<http://www.nilc.icmc.usp.br/nilc/projects/bloc-eco.htm>>. Acesso em 12 Nov. 2010.