

**IJKEM**  
International Journal of Knowledge Engineering  
and Management



---

## O USO DE ANOTAÇÕES SEMÂNTICAS E ONTOLOGIAS PARA A CLASSIFICAÇÃO DE DOCUMENTOS

Flavio Ceci\*

Cristiane Raquel Woszezenki\*\*

Alexandre Leopoldo Gonçalves\*\*\*

### Resumo

O problema da recuperação de informação pode ser visto como um problema de classificação de documentos. Um dos desafios enfrentados na classificação de documentos é classificar os textos de acordo com a semântica do seu conteúdo. Este trabalho apresenta um modelo para a classificação semântica de documentos fazendo uso de anotações semânticas e ontologias. A classificação realizada pelo modelo proposto foi comparada com a classificação manual realizada por especialistas, apresentando 51,6% de acerto. Levando em conta as limitações do estudo de caso, esse resultado atesta que anotações semânticas e ontologias são promissoras para a classificação de documentos.

**Palavras-Chave:** Anotações semânticas. Ontologias. Recuperação de Informação.

---

\* Doutorando em Engenharia e Gestão do Conhecimento (PPGEGC /UFSC). Mestre em Engenharia e Gestão do Conhecimento (PPGEGC/UFSC). Professor Titular da Universidade do Sul de Santa Catarina e desenvolvedor do Instituto Stela. flavio@stela.org.br

\*\* Doutoranda em Engenharia e Gestão do Conhecimento (PPGEGC /UFSC). Mestre em Ciência da Computação (PPGCC/PUCRS). Professora do Ensino Básico Técnico e Tecnológico no Instituto Federal de Santa Catarina. cristiane.raquel@ifsc.edu.br

\*\*\* Doutor em Engenharia de Produção (PPGEP/UFSC). Professor Adjunto da UFSC/Araranguá e Professor Permanente do Programa de Pós-Graduação em Engenharia e Gestão do Conhecimento da UFSC. alexandre.goncalves@ararangua.ufsc.br

## 1 INTRODUÇÃO

A Web é atualmente a maior fonte de pesquisa dos mais variados domínios. As informações nela contidas crescem a uma taxa exponencial (HIMMA, 2007), e a recuperação de documentos relevantes nesse vasto repositório exige processos cada vez mais sofisticados, apresentando novos desafios.

De acordo com Pokorný (2004), embora as pessoas acreditem que possam encontrar na Web facilmente o que procuram, os serviços de recuperação de informação apresentam muitos resultados inconsistentes, imprecisos, incompletos e irrelevantes. Um dos motivos disso é o fato de que os principais métodos de recuperação de informação na Web são baseados nos algoritmos tradicionais de recuperação de informação, que foram originalmente desenvolvidos para conjuntos menores e mais coerentes de informação do que os contidos na Web atualmente (LEVANDOWSKI, 2005). Devido ao crescimento exponencial do volume de informações na Web, às fontes de informação heterogêneas e ao conteúdo não-estruturado, as ferramentas de busca na Web hoje requerem novas técnicas.

O problema da recuperação de informação pode ser visto como um problema de classificação de documentos, onde, basicamente, um documento é classificado como “relevante” ou “não relevante” (LI; JAIN, 1998). Um documento é apresentado para o usuário apenas se ele for classificado como “relevante”. Contudo, a classificação de documentos apresenta muitos desafios e dificuldades, como por exemplo, considerar a semântica do conteúdo dos documentos a serem classificados.

Uma alternativa à classificação de documentos baseada nos algoritmos tradicionais de recuperação de informação é a anotação semântica dos recursos da Web. Existem técnicas, ferramentas e métodos que permitem o enriquecimento semântico de documentos a partir de anotações, de modo que essas anotações possam ser analisadas e exploradas por ferramentas de recuperação de informações para obtenção de resultados mais precisos. Uma abordagem comum para a criação de anotações semânticas é a utilização de ontologias para a descrição formal dos conceitos dos recursos da Web, ou documentos de um corpus.

Neste contexto, este trabalho apresenta um modelo para a classificação semântica de documentos fazendo uso de anotações semânticas e ontologias aliado a técnicas tradicionais de recuperação de informação, como a estrutura de índice invertido para sua persistência. O artigo está assim organizado: a seção 2 apresenta os aspectos conceituais levantados; a seção 3 descreve o modelo proposto; o estudo de caso e os resultados são apresentados na seção 4; as conclusões encerram o artigo na seção 5.

## **2 FUNDAMENTAÇÃO TEÓRICA**

Esta seção trata dos principais conceitos explorados neste trabalho: Classificação de Documentos, Anotações Semânticas e Ontologias.

### **2.1 Classificação de Documentos**

A classificação de documentos é uma área que tem como função a identificação e atribuição de temas, assuntos ou conceitos pré-definidos para um documento. Neste processo é necessário utilizar como base o conteúdo completo do documento, para que seja possível a sua interpretação e classificação (MADUREIRA, 2009).

A área da classificação de documentos é importante para áreas como: ciência da informação, ciência da computação, jornalismo, entre outras (MADUREIRA, 2009). Segundo Sebastiani (2002), no contexto da computação, a classificação de documentos e textos pode ser aplicada às seguintes situações:

- auxiliar na indexação e recuperação de documentos a fim de trazer uma lista de documentos já categorizados, organizando melhor o resultado da consulta;
- organização de documentos utilizando uma base de conhecimento, facilitando a organização da base de documentos de uma organização, por exemplo a gestão automática de notícias de um portal jornalístico;
- filtro de conteúdo, possibilitando a partir da análise do conteúdo do documento uma melhor distribuição, como por exemplo entrega ou bloqueio de notícias para um usuário;

- desambiguação de temas, ou seja, verificar que um documento tem como foco “banco” instituição financeira e o outro “banco” como objeto para se sentar.

Segundo Mamakis et al. (2012), as estratégias de classificação de documentos baseiam-se, em sua grande maioria em métodos estatísticos, como Naïve Bayes, modelo de linguagem e algoritmos de regressão ou em técnicas de processamento de linguagem natural.

De acordo com Dharmadhikari et al. (2011), a classificação de documentos pode ser supervisionada, sendo treinada a partir de interações e validadas por um especialista, ou não supervisionada, realizada de forma automática sem a necessidade de intervenção humana.

## **2.2 Anotações Semânticas**

De acordo com Popov et al. (2003), a anotação semântica é um processo de geração de metadados específicos para possibilitar novos métodos de acesso à informação e mesmo estender os métodos existentes. De forma mais sucinta, Sánchez-Fernández e Fernández-García (2005) referem-se à anotação semântica como o ato de anotar documentos com metadados. Um metadado é um dado que carrega uma informação estruturada necessária para o entendimento do significado de outro dado contido em algum tipo de mídia. Assim, a adição de metadados, enriquece o conteúdo dos recursos semanticamente (ARANTES, 2010).

Pokorný (2004) diz que uma vantagem das anotações semânticas é que os motores de busca podem usá-las para inferir relações semânticas adicionais, realizando, assim, uma busca semântica. A busca semântica permite estender e melhorar os processos de busca tradicionais da área de Relações com Investidores, possibilitando a criação de serviços de informação inteligentes na Web, sites personalizados e motores de busca que levam em consideração o significado do que o usuário está buscando.

Oren et al. (2006) consideram que as anotações podem ser realizadas de forma manual, semi-automática ou automática. A anotação manual pode ser realizada por uma ou mais pessoas. Anotações semi-automáticas são realizadas por pessoas com a assistência de ferramentas que fazem sugestões automáticas. Anotações automáticas criam as anotações sem intervenção manual. Os autores ainda diferenciam três tipos de anotação: informal, formal e ontológica.

Anotações informais não são lidas por máquina, pois não utilizam uma linguagem formal no processo. Anotações formais são interpretáveis por máquina, mas não usam termos de ontologias. Anotações ontológicas utilizam uma terminologia de conceitualização compartilhada chamada ontologia. Afirmam ainda, que a representação das anotações podem ser intrusivas ou não intrusivas. As intrusivas são guardadas nos próprios documentos e não intrusivas são armazenadas em repositórios de anotações que apontam para os documentos que passaram pelo processo de anotação.

### **2.3 Ontologias**

Segundo Studer (1998) uma ontologia é uma especificação formal e explícita de uma conceitualização compartilhada. Nessa definição, “especificação explícita” diz respeito às definições de conceitos, instâncias, relações, restrições e axiomas; “formal” significa declarativamente definida, portanto compreensível para agentes e sistemas; “conceitualização” diz respeito a um modelo abstrato de parte do conhecimento; e “compartilhada” significa conhecimento consensual. De forma mais sucinta, o W3C (2012) conceitua ontologia como “a definição dos termos utilizados na descrição e na representação de uma área do conhecimento”.

De acordo com Bouramoul (2011), a ontologia é usada na área de RI para representar descrições compartilhadas relacionadas a um ou mais domínios de forma a adicionar aspectos semânticos aos sistemas de RI. O autor diz que o conhecimento representado pela ontologia pode ser usado em três níveis diferentes no processo de RI: pode ajudar a indexar documentos (indexação semântica); pode assistir a formulação da necessidade do usuário; e finalmente, a ontologia pode ser usada no modelo de busca para encontrar e organizar os documentos que correspondem à necessidade do usuário.

Os métodos de consulta baseados em ontologia exploram a relação entre os conceitos e não apenas a sintaxe exata ou parcial do termo consultado como presumem os modelos tradicionais de recuperação de informação (PARALIC, 1999).

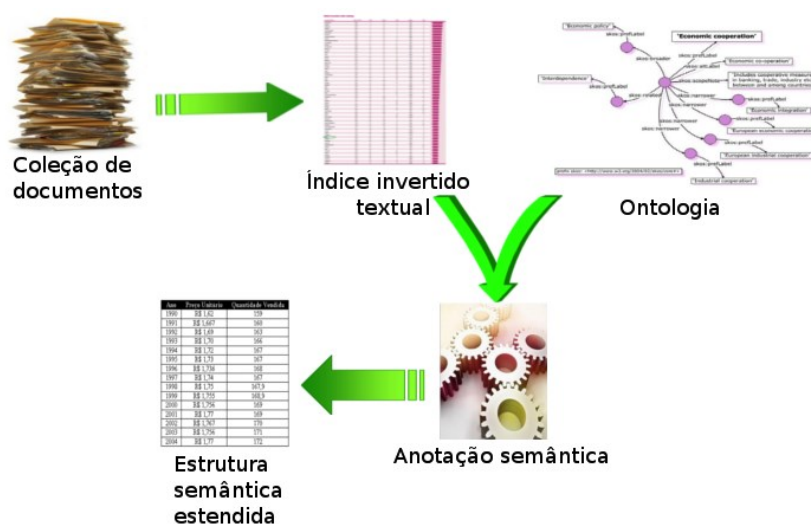
O uso de uma ontologia possibilita definir conceitos e suas relações representando o conhecimento sobre um documento em termos específicos de um domínio. Para representar o

conteúdo de um documento explicitamente, é necessário criar *links* (associação) entre o documento e partes relevantes de um modelo de domínio, ou seja, associar aqueles elementos do modelo de domínio que são relevantes ao conteúdo do documento.

### 3 MODELO PROPOSTO

A partir do referencial teórico desenvolveu-se um modelo para viabilizar a classificação não supervisionada de documentos utilizando anotações semânticas e ontologias para conceitualizar o domínio. A Figura 1 ilustra o modelo proposto.

Figura 1 - Modelo proposto para a anotação semântica.



Fonte: os autores.

Conforme mostra a Figura 1, o primeiro passo consiste em realizar a indexação dos documentos. Nesta etapa é gerado o índice invertido sobre a base de documentos de modo a facilitar a sua recuperação. O índice invertido consiste em uma lista de palavras ligadas a uma lista de documentos, onde na sua intersecção são mantidas informações como frequência em que a palavra ocorre no documento, posições da palavra, entre outras informações relevantes ao processo de recuperação (KNUTH, 1997).

O processo de anotação semântica recebe como entrada o índice invertido e uma ontologia de domínio e produz uma estrutura semântica estendida de forma a incorporar conceitos e propriedades da ontologia ao índice invertido. Esse processo é explicado a seguir.

### 3.1 Anotação Semântica

De acordo com a classificação das anotações semânticas apresentada na seção 2.2, o processo de anotação utilizado neste trabalho é automático, ontológico e não intrusivo. Os processos que compõem a anotação semântica dos documentos são ilustrados na Figura 2.

Figura 2 – Processos para a anotação semântica.



Fonte: os autores.

O processo “Gerador de Vetor de contexto” é baseado na estratégia de formulação de um vetor de contexto (vetor de termos mais representativos de um documento) proposta por Jing e Tzoukermann (1999) que utiliza a distância morfológica dos termos para a criação do vetor. Segundo Jing e Tzoukermann (1999), o vetor de contexto baseia-se em todas as ocorrências da mesma palavra no documento e o pressuposto de que uma palavra tem um significado dominante através do documento.

Para compor o vetor de contexto é necessário que o especialista identifique uma ou mais palavras alvo relacionadas com o domínio da aplicação. A partir dessas palavras são calculadas as distâncias semânticas entre duas palavras em uma janela<sup>1</sup>, utilizando a fórmula a seguir:

$$Peso = \frac{fr(I|T)}{fr(T)} \quad \text{Onde:}$$

A variável *I* é a palavra da janela em questão e *T* é a palavra alvo.

A fórmula calcula um peso para os termos do documento levando em consideração a palavra alvo (basicamente o peso equivale à frequência em que o termo em questão ocorre juntamente com a palavra-alvo dividida pela frequência total da palavra-alvo).

Após atribuir o peso aos termos do documento é construído um vetor ordenado de maneira decrescente, onde segundo os autores, devem-se selecionar os 10 primeiros termos relacionados com as palavras alvo eleitas pelo especialista para representar o domínio. O próximo passo é verificar a frequência conjunta dos termos do vetor com relação a todos os documentos (aqui se verifica se a relação entre os termos faz sentido ao domínio como um todo, diferentemente da etapa anterior na qual era levando em consideração apenas o documento em questão), conforme a seguinte fórmula:

$$R(I1, I2) = \frac{DF(I1 I2)}{DF(I1) + DF(I2) - DF(I1 I2)}$$

A relação conjunta entre os termos é calculada dividindo-se a frequência conjunta em que ocorrem dois termos pela soma que ocorre os termos de maneira isolada, menos o valor da frequência conjunta. Ao final deste processo é gerado um vetor com os termos mais representativos do documento em relação ao domínio.

Conforme mencionado anteriormente, o processo de anotação semântica recebe como entrada uma ontologia de domínio. Assim, para cada classe da ontologia é criado um vetor de *labels*. Esse processo é realizado durante a criação da base de conhecimento, que é composta pela ontologia, instâncias e *labels*, podendo ser feito de forma manual por um especialista do domínio ou com o auxílio de ferramentas que elegem os termos candidatos para a ontologia a partir do próprio corpus. O processo “Acesso à base de conhecimento” ilustrado na Figura 2 recupera os *labels* e disponibiliza-os para o “Analisador de distância entre vetores”.

De posse do vetor de contexto e do vetor de labels, o “Analisador de distância entre vetores” cria a “Estrutura semântica estendida”: para cada documento da coleção, são percorridos os vetores de labels de todas as classes da ontologia bem como o vetor de contexto para calcular o peso do documento relativo à cada classe. O cálculo de distância entre os vetores é representado pela seguinte equação:



$$dist(vc, vtc) = \sum_{i=1}^{10} (R(A_i, B_{m(i)}) \cdot w_{1,i} \cdot w_{2,m(i)})$$

Onde:

$$vc = \{A1(w_{1,1}), A2(w_{1,2}), \dots, A10(w_{1,10})\}$$

$$vtc = \{B1(w_{2,1}), B2(w_{2,2}), \dots, B10(w_{2,10})\}$$

O vetor  $vc$  representa o vetor de contexto de um documento e o  $vtc$  representa o vetor de termos da classe (*labels*). Para calcular a distância entre esses vetores é realizado o somatório da relação conjunta ( $R(A_i, B_{m(i)})$ ) de cada par de termos (um do  $vc$  e outro do  $vtc$ ) multiplicada pelo peso das duas palavras de maneira isolada ( $w_{1,i}$  e  $w_{2,m(i)}$ ). O resultado deve ser normalizado e apresentar um grau entre 0 e 1. Se o peso for superior a 0,6 o documento é relacionado à classe em questão.

Os pesos extraídos do relacionamento dos documentos com as classes são armazenados na “Estrutura semântica estendida” e estão diretamente relacionados com o identificador do documento no índice invertido

### 3.2 Ferramentas Utilizadas

O modelo concebido neste trabalho foi desenvolvido a partir de uma série de estudos isolados. Para a implementação do processo de indexação foi utilizada a linguagem Java e o framework de busca e indexação Lucene (<http://lucene.apache.org>). A modelagem da ontologia de domínio foi realizada utilizando a ferramenta Protegé (<http://protege.stanford.edu/>). Para a persistência da base de dados e conhecimento foi utilizado o SGBD PostgreSQL (<http://www.postgresql.org/>).

## 4 ESTUDO DE CASO

Para atestar a viabilidade do modelo proposto, foi desenvolvido um esse estudo de caso que utiliza como base 89 monografias dos cursos de Ciência da Computação e Sistema de Informação de uma universidade do estado de Santa Catarina. Estas monografias tiveram seu

conteúdo textual indexado completamente, além de armazenados o nome do documento e o seu título.

Tendo em vista o domínio a ser modelado, foi construída uma ontologia da área da computação levando em conta suas principais subáreas e disciplinas, totalizando 20 classes (conceitos) e 64 *labels* distribuídos entre as mesmas. Para cada classe foram selecionados até 10 *labels* representativos para o mesmo. Para a extração desses *labels* foi utilizada uma estratégia baseada no trabalho de Ceci et al. (2012), que extraem os termos candidatos a partir dos próprios documentos. A base de conhecimento foi persistida num modelo relacional.

A partir da base de documentos indexada e da ontologia construída, as monografias foram submetidas ao processo de anotação, gerando a estrutura semântica estendida.

Para validar o modelo proposto neste trabalho, as anotações realizadas pela ferramenta implementada foram comparadas às anotações realizadas manualmente por especialista. Para isso, foi elaborada uma planilha com todos os títulos das monografias sendo submetida a 8 profissionais em nível de doutoramento (com formação na área da computação) para que indicassem qual conceito da ontologia melhor caracterizava cada monografia.

De posse das 8 planilhas preenchidas pelos profissionais, identificou-se, para cada monografia, o conceito mais utilizado para identificá-la. Os resultados do comparativo do processo manual e do processo automático são discutidos a seguir.

#### **4.1 Resultados**

Das 89 monografias submetidas ao processo de anotação, foi possível classificar 31 delas através do processo proposto contra 59 classificadas manualmente pelos especialistas. Assim, constatou-se que a ontologia necessitava incorporar mais conceitos para o domínio modelado. Foi verificada também uma grande divergência entre as classificações realizadas pelos entrevistados, o que caracteriza a dificuldade do processo mesmo para um especialista do domínio.

As 31 monografias classificadas pelo processo automático também foram classificadas pelo processo manual. Dentre as 31 monografias, 16 compartilharam a mesma classificação, ou seja, obteve-se um percentual de 51,6% de acerto. Esses números mostram que a presente proposta de modelo para classificação não supervisionada de documentos textuais é promissora, pois, levando em conta a pequena quantidade de conceitos contemplados pela ontologia modelada bem como a pequena quantidade de *labels* associados a esses conceitos, foi possível obter um resultado significativo.

Considerando todos os documentos que compõem o corpus, o percentual de acerto é de 18%; já levando em consideração o total de documentos classificados pelos especialistas, tem-se uma taxa de acerto de 27,1%.

Os dois conceitos da ontologia que mais tiveram acertos foram “Engenharia de Software” e “Sistemas de Apoio a Decisão”, ambas com 6 termos relacionados (as duas classes são também as que mais possuem *labels*). Esses dados demonstram que deve-se investir mais na definição dos conceitos e no reconhecimento de *labels*.

## **5 CONSIDERAÇÕES FINAIS**

Anotações semânticas em documentos geram metadados que podem ser analisados e explorados por ferramentas de recuperação de informação para obtenção de resultados mais precisos. As ontologias podem contribuir para este processo uma vez que possibilitam definir conceitos e suas relações representando o conhecimento de um documento em termos específicos de um domínio.

Este trabalho apresentou uma proposta de uso de anotações semânticas e ontologias para a classificação não supervisionada de documentos. No modelo proposto, o processo de anotação semântica estende o índice invertido incorporando conceitos e propriedades da ontologia de domínio. A estrutura semântica estendida utiliza como base o índice invertido, possibilitando que a estrutura em questão também possa ser aplicada para a categorização de documentos a partir de um processo de busca completa ao conteúdo dos documentos.

Foi desenvolvido um protótipo do modelo proposto para atestar a sua viabilidade, fazendo uso de um de um conjunto de monografias da área da computação como base de testes e construindo-se uma ontologia a partir das disciplinas dessa área.

Observou-se que a ontologia desenvolvida necessita da incorporação de novos conceitos de forma a enriquecer a representação do domínio e, conseqüentemente, obter melhores resultados no processo de classificação não supervisionada de documentos. Percebeu-se também que o processo de classificação de documentos é algo bastante complexo, até mesmo para especialistas de domínio.

Os resultados obtidos a partir deste estudo são iniciais e demonstram que o modelo proposto possui uma estratégia promissora cujo potencial deve ser explorado pela continuação do trabalho. Assim, como trabalhos futuros são sugeridos: i) a identificação de outros métodos para auxiliar na construção do vetor de contexto dos documentos; ii) aprimoramento da representação do conhecimento de domínio (ontologia) e utilização de suas propriedades para auxiliar o processo de classificação; e iii) desenvolvimento um módulo de recuperação de documentos utilizando o modelo de classificação aqui apresentado.

Arquivo recebido em 18 de julho de 2013 e aceito para publicação em 16 de agosto de 2013.

## **THE USE OF SEMANTIC ANNOTATIONS AND ONTOLOGY FOR DOCUMENTS CLASSIFICATION**

### **Abstract**

*The problem of information retrieval can be viewed as a problem of documents classification. One of the challenges in the documents classification is to classify texts according to their semantic content. This paper presents a model for semantic classification of documents making use of ontologies and semantic annotations. The classification made by the proposed model was compared with manual classification performed by specialists, with 51.6% accuracy. Taking into account the limitations of the case study, this result confirms that ontologies and semantic annotations are promising for the classification of documents.*

**Keywords:** *Digital and social inclusion. Information Society. Citizenship. Autopoiesis.*

## REFERÊNCIAS

ARANTES, L. O. “Documentação Semântica no Apoio à Integração de Dados e Rastreabilidade”. Dissertação (Mestrado em Informática) - Universidade Federal do Espírito Santo, Vitória, 2010.

BOURAMOUL, A.; KHOLLADI, M.; DOAN, B. How Ontology can be used to Improve Semantic Information Retrieval: The AnimSe Finder Tool. **International Journal of Computer Applications**, Vol. 21, n. 9, 2011.

CECI, Flavio; PIETROBON, Ricardo; GONÇALVES Alexandre Leopoldo. Turning Text into Research Networks: Information Retrieval and Computational Ontologies in the Creation of Scientific Databases. **PloS One**, Vol. 7, n. 1, p. e27499, 2012.

DHARMADHIKARI, S. C.; INGLE, M.; KULKARNI, P. Empirical Studies on Machine Learning Based Text Classification Algorithms. **Advanced Computing: An International Journal**, Vol. 2, n. 6, p. 161-169, 2011.

HIMMA, K. The concept of information overload: A preliminary step in understanding the nature of a harmful information-related condition. **Ethics and Information Technology**, Vol. 9, n. 4, p. 259-272, 2007.

JING, H.; TZOUKERMANN, E. Information Retrieval Based on Context Distance and Morphology. **International Conference on Research and Development in Information Retrieval**, 1999.

KNUTH, D. E. **The Art of Computer Programming**. 3rd Edition. Massachusetts: Addison-Wesley, 1997.

LEWANDOWSKI, D. Web searching, search engines and Information Retrieval. **Information Services & Use**. Vol. 25, p. 137-147, 2005.

LI, Y. H.; JAIN, A. K. Classification of Text Document. **The Computer Journal**, Vol. 41, n. 8, p. 537-546, 1998.

MADUREIRA, F. A. P. “Classificação de Documentos”. Dissertação submetida para obtenção do grau de Mestre em Engenharia de Informática. Universidade de Nova Lisboa, Portugal, 2009.

MAMAKIS, G.; MALAMOS, A. G.; WARE, A.; KARELLI, I. Document Classification in Summarization. **Journal of Information and Computing Science**, Vol. 7, n. 1, p. 25-36, 2012.  
OREN, E.; MÖLLER, K. H.; SCERRI, S.; HANDSCHUH, S.; SINTEK, M. What are Semantic Annotations?, 2006. Disponível em: <<http://citeseer.ist.psu.edu/showciting?doi=10.1.1.97.7985>>. Acesso: 14/07/2012.

PARALIC, J.; KOSTIAL, I. Ontology-based Information Retrieval, 1999. Disponível em: <<http://people.tuke.sk/jan.paralic/papers/IIS03.pdf>>. Acesso: 14/07/2012.

POKORNÝ, J. Web Searching and Information Retrieval. **Computing in Science and Engineering**, Vol. 6, n. 4, p. 43-48, 2004.

POPOV, B.; KIRYAKOV, A.; OGNJANOFF, D.; MANOV, D.; KIRILOV, A.; GORANOV, M. Towards Semantic Web Information Extraction, Human Language Technologies. Workshop at the 2nd International Semantic Web Conference (ISWC2003), Florida, USA, 2003.

SÁNCHEZ-FERNÁNDEZ, L.; FERNÁNDEZ-GARCIA, N. The Semantic Web: Fundamentals and A Brief State-of-the-Art. **Upgrade Journal**. Vol. VI, n. 6 , 2005.

SEBASTIANI, F. Machine learning in automated text categorization. *ACM Computing Surveys*, Vol. 34, n. 1, p. 1-47, 2002.

STUDER, R.; BENJAMINS, R.; FENSEL, D. Knowledge engineering: Principles and methods. **Data & Knowledge Engineering**, Vol. 25, n. 1-2, p. 161-197, 1998.