

Universidade Federal de Santa Catarina
Centro Tecnológico
Departamento de Informática e Estatística
Graduação em Sistemas de Informação
Disciplina: INE5632 - Projetos II

**Anotações Semânticas de Fontes de Dados
Heterogêneas
Um Estudo de Caso com a Ferramenta Smore**

Autor: Markus Pereira Eller
Orientador: Renato Fileto

Florianópolis, 22 de Novembro de 2008.

UNIVERSIDADE FEDERAL DE SANTA CATARINA
DEPARTAMENTO DE INFORMÁTICA E ESTATÍSTICA
CURSO DE BACHARELADO EM SISTEMAS DE INFORMAÇÃO

**Anotações Semânticas de Fontes de Dados
Heterogêneas
Um Estudo de Caso com a Ferramenta Smore**

Markus Pereira Eller

Trabalho de conclusão de curso
apresentado como parte dos
requisitos para obtenção do grau
de Bacharel em Sistemas de
Informação.
Orientador: Prof.Dr. Renato Fileto

Florianópolis - SC
2008/2

**Anotações Semânticas de Fontes de Dados
Heterogêneas
Um Estudo de Caso com a Ferramenta Smore**

Markus Pereira Eller

Trabalho de conclusão de curso apresentado como parte dos requisitos para obtenção do grau de Bacharel em Sistemas de Informação.

Orientador: Renato Fileto
Universidade Federal de Santa Catarina
fileto@inf.ufsc.br

Banca examinadora

Prof. Fernando Álvaro Ostuni Gauthier
Universidade Federal de Santa Catarina
gauthier@inf.ufsc.br

Prof. Leandro José Komosinski
Universidade Federal de Santa Catarina
leandro@inf.ufsc.br

*“Podem até dizer que tive sorte...
Pois eu digo que sou um
reflexo de minhas escolhas.
Que Deus me permita chegar
até onde meus sonhos alcançarem.”*

Agradecimentos

Primeiro a Deus, por ter sempre iluminado o meu caminho e me dado força para fazer as mais difíceis escolhas.

A todas as pessoas que contribuíram para eu ter conseguido chegar neste momento. Começando pelos meus pais, João Batista e Edite, por estarem ao meu lado em todos os momentos.

A minha família, que sempre soube me mostrar o quanto é importante se esforçar para conseguir alcançar os nossos objetivos.

Aos colegas, alguns desde lá do Anísio, em Santo Amaro, outros da Escola Técnica de São José e outros da própria UFSC, pela ajuda técnica, incentivo, colaboração e muita alegria ao meu lado.

Agradeço a todos os professores que tive na vida, com quem consegui aprender não só teorias ou funções matemáticas, mas também a dar valor às coisas e respeitar as pessoas. Alguns que foram muito além de professores e viraram referência na minha vida, realmente amigos e companheiros.

Claro, todos os amigos que de alguma forma se preocuparam comigo e sei que, mesmo de longe, torceram por mim. Assim como eu desejo o melhor pra vocês!

A pessoa que, pela primeira vez na minha vida, me disse um "Eu te amo" e consegue fazer o meu coração transbordar de felicidade toda vez que me dá um abraço!

Resumo

A tarefa de recuperação de informação vem se tornando cada vez mais trabalhosa devido à quantidade de documentos existentes e à falta de padronização para descrever e recuperar tais documentos. Várias técnicas tentam acelerar e fazer a correta recuperação de documentos. A maioria dessas técnicas é baseada em análise quantitativa dos componentes léxicos e sintáticos do conteúdo dos documentos e dos metadados que os descrevem. A Web semântica propõe uma nova alternativa para a recuperação de informação, mediante o uso de conhecimento formalizado em ontologias processáveis por computadores para anotação e recuperação dos documentos digitais. Esta abordagem tem potencial para aumentar os níveis de precisão e revocação da recuperação de informação. O presente trabalho objetiva a pesquisa de ferramentas para fazer a anotação semântica automática e semi-automática de documentos. Essas anotações serão utilizadas posteriormente para efetuar a recuperação de informação de grandes volumes de documentos.

Palavras-chave: Web Semântica; Anotação Semântica; Recuperação de Informação; Ontologia.

Abstract

The task of recovering information is becoming increasingly laborious because of the number of existing documents and the lack of standardization to describe and recover these documents. Several techniques try to speed up and improve the quality of documents recovering. Most of these techniques are based on quantitative analysis of the lexical and syntactic components of the contents of the documents and their descriptive metadata. The Semantic Web proposes a new alternative for the information retrieval, that uses knowledge formalized in computer processable ontologies for annotating and retrieving digital documents. This approach has the potential to increase the levels of precision and recall of information retrieval. The goal of this study is to search for and evaluate tools for automatic and semi-automatic semantic annotation of documents. These annotations will be used later for information retrieval from large volumes of documents.

Keywords: Semantic Web; Semantic Annotation; Recovery Information; Ontology.

Sumário

Agradecimentos	5
Resumo	6
Abstract	7
Lista de Figuras	10
Lista de Tabelas	11
Lista de Siglas	12
1 - Introdução	13
1.1 - Considerações Iniciais	13
1.2 - Justificativa	14
1.3 - Objetivos	15
1.3.1 - Objetivo Geral	15
1.3.2 - Objetivos Específicos	15
1.4 - Técnicas e ferramentas	15
1.5 - Organização do trabalho	15
2 – Fundamentos	17
2.1 – Recuperação de Informação	17
2.1.1 - Técnicas de Recuperação da Informação	18
2.1.1.1 - Sistema de indexação baseado em palavras-chave	18
2.1.1.2 - Sistema baseado em ontologias	19
2.2 – Web Semântica	20
2.2.1 - As camadas da Web Semântica propostas pelo W3C	22
2.1.1.1 - URI + Unicode	22
2.1.1.2 - XML	23
2.1.1.3 - RDF + rdfschema	26
2.1.1.3 - A camada lógica	29
2.3 Ontologias	29
2.3.1 – Tipos de Ontologias	30
2.3.2 – Profundidade Ontologica	31
2.3.3 – Componentes de um ontologia	31
2.3.4 – Aplicações de Ontologias	32
2.3.5– Linguagens para representação de ontologias	34
2.3.5.1 – DAML+OIL	34

2.3.5.2 – OWL	35
3 – Anotação Semântica	38
3.1 – Considerações sobre as anotações semânticas	38
3.2 – Tipos de ferramentas para anotação semântica	41
3.3 – Características relevantes	43
3.4 – Ferramentas para anotação semântica	45
3.4.1 – Ont-O-Mat	46
3.4.2 – MnM	47
3.4.3 – KIM	48
3.4.4 – Smore	49
3.4.5 - Annotea	50
3.5 – Comparação das Ferramentas	51
4 – Estudo de Caso	53
4.1 – O documento anotado	53
4.2 – A ontologia Ontojuris	54
4.3 – O processo de anotação semântica	57
4.3.1 – Etapas do processo de anotação	60
4.3.2 – Código RDF gerado	66
4.4 – Considerações finais	68
5 – Conclusões e Trabalhos Futuros	69
5.1 – Contribuições	70
5.2 – Trabalhos futuros	70
6 – Referências	72
Anexo 1: Exemplo de acórdão	77
Anexo 2: Padrão do acórdão	81
Apêndice 1 – ARTIGO	82

Lista de Figuras

Figura 1: As camadas da Web semântica	22
Figura 2: Exemplo recursos da XML	24
Figura 3: Exemplo de estrutura XML em árvore	25
Figura 4: Exemplo de DTD	25
Figura 5: Modelo gráfico de representação RDF	28
Figura 6: Sentença RDF escrita em XML	28
Figura 7: Sentença RDF na forma de triplas	28
Figura 8: Evolução da linguagem OWL	36
Figura 9: Exemplo de marcação semântica	40
Figura 10: Classes de geradores de anotação semântica	43
Figura 11: Rede Semântica representando a ONTOJURIS	55
Figura 12: Atributos da classe oj_ClasseDocumento	55
Figura 13: Declaração de <i>Namespaces</i> da ontologia	56
Figura 14: Declaração de propriedades de classes da ontologia	56
Figura 15: Declaração de classe e subclasse	57
Figura 16: Interface da ferramenta Smore	58
Figura 17: Exemplo de anotação gerada pela ferramenta Smore	60
Figura 18: Localização da ontologia com a ferramenta Smore	61
Figura 19: Documento que será anotado carregado no Smore	61
Figura 20: Seleção de termos do documento	63
Figura 21: Criação de uma nova instancia	64
Figura 22: Geração das anotações em RDF	65
Figura 23: <i>Namespaces</i>	66
Figura 24: Declaração de <i>headers</i>	66
Figura 25: Associação dos termos do texto com as Classes da ontologia	67

Lista de Tabelas

Tabela 1: Comparação de ferramentas de anotação semântica

52

Lista de Siglas

API	<i>Application Programming Interface</i>
DAML	<i>DARPA Agent Markup Language</i>
DTD	<i>Document Type Definition</i>
HTML	<i>Hipertext Markup Language</i>
HTTP	<i>Hypertext Transfer Protocol</i>
IE	<i>Information Extraction</i>
OIL	<i>Ontology Inference Layer</i>
OKBC	<i>Open Knowledge Base Connectivity</i>
OWL	<i>Web Ontology Language</i>
PLN	<i>Processamento de Linguagem Natural</i>
RDF	<i>Resource Description Framework</i>
RDFS	<i>RDF Schema</i>
SAP	<i>Semantic Annotation Plataforms</i>
SHOE	<i>Simple HTML Ontology Extensions</i>
SMA	<i>Sistemas Multiagentes</i>
UNICODE	<i>Universal Character Encoding</i>
URI	<i>Uniform Resource Indicator</i>
URL	<i>Uniform Resource Locator</i>
W3C	<i>World Wide Web Consortium</i>
WWW	<i>World Wide Web</i>
XML	<i>eXtensible Markup Language</i>
XMLS	<i>XML Schema</i>
XOL	<i>XML-based Ontology Exchange Language</i>
XPointer	<i>XML Pointer Language</i>

1 – Introdução

1.1 - Considerações Iniciais

A Internet, em particular a *World Wide Web* (Web), é atualmente um grande meio de disseminação de informação. O seu livre crescimento sem meios para representar ou padronizar a semântica das informações nela publicados fez com que os documentos contendo essas informações ficassem dispersos de forma desorganizada e sem padronização. Isso, apesar de contribuir para a sua grande popularização, está se tornando um desafio para gestores que buscam o gerenciamento e a integração dos recursos de informação.

O que se tinha antes da Internet eram compartilhamentos de documentos apenas de uma organização e usando um mesmo tipo ferramenta. Hoje as pessoas podem acessar documentos dispersos geograficamente e sem padronização de conteúdo, formatação e descrição. Os recursos de informação podem ser geridos com o auxílio de ferramentas diversas, como sistemas de bancos de dados, *data warehouses*, gerenciadores de correio eletrônico, navegadores Web e mecanismos de busca variados. Os documentos compartilhados através da linguagem HTML (*Hyper Text Markup Language*) necessitam que pessoas façam o trabalho de formatação, ligação e interpretação dos documentos publicados. Aos computadores cabe a tarefa de exibí-los de acordo com a formatação definida. Os mecanismos de busca de informação na Web, baseados principalmente em buscas por palavras-chaves, apresentam problemas de precisão e revocação. A precisão, ou relevância, refere-se à aderência do conjunto de documentos recuperados pelo sistema à consulta submetida por um usuário. Ela é matematicamente definida pela relação entre o número de documentos relevantes recuperados e o número total de documentos recuperados pelo sistema em atendimento a uma consulta. A revocação, por sua vez, é a relação entre o número de documentos relevantes para o usuário recuperados pelo sistema, e o número total de documentos relevantes para o usuário existentes nos registros do mesmo sistema. Seguem abaixo as definições matemáticas de precisão e revocação:

$$\text{Precisão} = \frac{\text{N.º de documentos relevantes recuperados}}{\text{N.º total de documentos recuperados}}$$

$$\text{Revocação} = \frac{\text{N}^\circ. \text{ de documentos relevantes recuperados}}{\text{N}^\circ. \text{ de documentos relevantes existentes}}$$

A idéia de associar anotações semânticas aos documentos, proposta pela Web semântica, é uma maneira de organizar o processo de publicação e recuperação de informação. Uma anotação semântica, isto é, uma associação entre expressões relevantes de trechos do texto ou dos metadados descrevendo um documento a conceitos e instâncias descritos em uma ontologia, pode permitir a recuperação de informação contida nos documentos com maior grau de precisão e revocação. No entanto, vários problemas precisam ser solucionados para alcançar este objetivo, incluindo a definição de métodos e ferramentas para automatizar o processo de anotação semântica.

O presente trabalho tem como objetivo fazer uma análise comparativa de ferramentas desenvolvidas para se fazer a anotação semântica de documentos de maneira automática e semi-automática. Tais anotações devem contribuir para a extração de informação de grandes volumes de documentos publicados na Web.

1.2 - Justificativa

A alta heterogeneidade, a autonomia e a ampla distribuição dos dados na Web sem uma padronização adequada tornam as consultas difíceis de serem processadas e os resultados pouco precisos. As consultas geralmente são efetuadas por navegação exaustiva ou através de mecanismos de busca por palavras-chaves. Esses métodos dificilmente resultam em sucesso sem dispendir muito esforço, devido ao grande volume de dados a serem consultados e a limitações dos mecanismos de busca por palavras-chave baseados em processamento puramente sintático.

Uma situação semelhante ocorre dentro de algumas instituições, devido ao excessivo volume de dados e documentos heterogêneos gerados em períodos de tempo cada vez menores. O acesso às informações neles contidas depende de uma classificação robusta e de mecanismos de busca que localizem a informação necessária rapidamente. A Web Semântica pode ajudar neste processo, pois propõe mecanismos que levam em consideração a semântica (significado) da informação dentro de certo contexto.

1.3 - Objetivos

1.3.1 - Objetivo Geral

O objetivo geral deste trabalho é pesquisar ferramentas para fazer a anotação semântica automática e semi-automática de documentos, usando termos (conceitos e instâncias) de uma ontologia de domínio pré-definida. Essas anotações serão utilizadas posteriormente para efetuar a recuperação de informação de grandes volumes de documentos.

1.3.2 - Objetivos Específicos

- Identificação das técnicas, conceitos e definições referentes à Web Semântica;
- Definir um esboço de ontologia base para a realização do trabalho;
- Buscar e selecionar ferramentas para a anotação semântica de documentos;
- Fazer a anotação semântica de documentos, dentro de um escopo definido e utilizando as ferramentas de anotação selecionadas;
- Descrever como essas anotações poderão ser utilizadas na recuperação de informação dos documentos.

1.4 - Técnicas e ferramentas

Serão estudadas técnicas, ferramentas e metodologias para fazer a anotação semântica de documentos. O trabalho envolve ainda o estudo de uma ontologia que definirá o escopo dos documentos que passarão pelo processo de anotação. Serão utilizadas as linguagens XML, RDF e OWL, assim como ferramentas para a manipulação da ontologia para posterior anotação semântica dos documentos.

1.5 - Organização do trabalho

O presente trabalho está dividido em cinco capítulos principais, incluindo este primeiro capítulo de introdução. O segundo capítulo apresenta a fundamentação teórica, com noções de recuperação de informação e apresentação de conceitos e definições que envolvem a Web semântica. O terceiro capítulo apresenta características

das anotações semânticas, fundamentais para a boa compreensão do trabalho. Ele também faz uma síntese das ferramentas e métodos de anotação semântica estudados, apresentando suas características próprias, pontos fortes e fracos, colhidos nas documentações associadas e avaliados em experimentos. O quarto capítulo é um estudo de caso, onde é descrito o processo de anotação semântica com uma ferramenta específica. O quinto capítulo apresenta as conclusões e os trabalhos futuros.

2 – Fundamentos

2.1 – Recuperação de Informação

Recuperação da informação é uma subárea da Ciência da Computação que estuda o processo de armazenamento e recuperação automática de documentos, que são objetos de dados, geralmente textos [CARDOSO, 2003]. O principal objetivo é encontrar, de forma eficiente, os documentos que satisfaçam a necessidade do usuário.

A tarefa de recuperação de informação se torna mais trabalhosa a cada dia devido ao crescimento do número de documentos e à desordem em que estes são mantidos. Documentos são disponibilizados de diversas maneiras, como manuais, formulários, cartas, informativos, entre muitas outras formas, sem padronização estrutural ou semântica de seus conteúdos.

A recuperação da informação disponível é de extrema importância para qualquer tipo de organização. A crescente complexidade dos objetos armazenados e o grande volume de dados exigem processos de recuperação cada vez mais sofisticados. Diante deste quadro, a recuperação de informação apresenta novos desafios e novas alternativas têm sido propostas.

A ferramenta mais importante para auxiliar o processo de recuperação é denominada índice, que é uma coleção de termos com indicação do local (documento ou trecho de documento) onde informação associada a esses termos pode ser localizada [FRAKES, 1992]. Estes termos devem ser organizados de forma a facilitar a busca de informações no conjunto de documentos.

O processo de recuperação de informação inicia-se com a especificação dos termos para a consulta que será realizada sobre os documentos. Geralmente, a especificação da consulta é uma tarefa difícil. Há frequentemente uma distância semântica entre a real necessidade do usuário e o que ele expressa na consulta formulada. Essa distância pode ser gerada pelo limitado conhecimento do usuário sobre o universo de discurso ou pelo formalismo que a linguagem de consulta exige.

O processo de recuperação de informação consiste na geração de uma lista de documentos recuperados para responder a consulta inicialmente formulada pelo usuário. Os índices construídos para uma coleção de documentos são usados para acelerar esta tarefa.

2.1.1 - Técnicas de Recuperação da Informação

Como a recuperação da informação se tornou algo de extrema importância para qualquer tipo de aplicação que lida com dados, algumas técnicas foram estudadas, planejadas e desenvolvidas para tentar realizar esta tarefa de forma mais precisa e rápida. Atualmente, existem várias técnicas para fazer a recuperação de informação. A recomendação de uma ou de outra depende do tipo de documento a ser manipulado, do tipo de consulta e da amplitude e precisão desejados para o resultado das buscas.

Os modelos clássicos utilizados no processo de recuperação de informação são o booleano, o vetorial e o probabilístico [CARDOSO, 2003]. Estes modelos consideram que cada documento é descrito por um conjunto de palavras-chaves, chamadas termos de indexação. Associa-se a cada termo de indexação ti em um documento dj um peso $wij \geq 0$. Os pesos quantificam a correlação entre os termos e o documentos. Além dos três modelos clássicos, modelos muito mais avançados de recuperação de informação têm sido propostos ao longo dos anos. Dentre eles, destacam-se modelos baseados em bases de conhecimento [BIWAS, 1987], lógica *fuzzi* [BOOKSTEIN, 1980], redes neurais [KWOK, 1995] e os modelos de recuperação baseados em ontologias e Web Semântica.

2.1.1.1 - Sistema de indexação baseado em palavras-chave

Os três modelos clássicos utilizados no processo de recuperação de informação (booleano, vetorial e probabilístico), foram desenvolvidos para funcionarem a partir de pesquisas em que o usuário realiza suas consultas informando palavras-chaves como ponto de partida. Com as palavras-chaves, cada um dos modelos utiliza métodos e algoritmos diferentes para retornarem o resultado para o usuário final.

O sistema de indexação baseado em palavras-chave é o mecanismo de busca mais utilizado atualmente. Nele, o usuário insere uma lista de palavras-chaves ou expressões booleanas no campo de busca, e o sistema procura por estas palavras ao longo do texto. É um tipo de pesquisa binária, isto é, que retorna ou não um documento, dependendo da ocorrência ou não de alguma das palavras-chaves no seu conteúdo. Os documentos recuperados serão aqueles que contiverem os termos que satisfazem a expressão lógica da consulta, que podem conter conectores como AND, OR e NOT entre as palavras-chaves, para indicar que se deseja recuperar os documentos contendo todas as palavras-chaves, alguma delas ou que não contenha uma ou mais palavras. As

vantagens desse modelo são facilidade de implementação e certa expressividade das expressões de consulta. Porém, ele apresenta várias limitações e desvantagens [BAEZA-YATES & RIBEIRO-NETO, 99]:

- A busca geralmente retorna uma grande quantidade de documentos irrelevantes;
- A classificação das palavras é estática, quer ela seja feita de forma manual ou automática;
- Dependendo do domínio, o sistema tem baixo ou nenhum retorno para determinadas palavras pesquisadas;
- Os resultados são extremamente dependentes do vocabulário utilizado. O sistema só retorna documentos que contenham exatamente o que o usuário digitou.

Alguns mecanismos de busca deste tipo que são amplamente conhecidos e muito utilizados: Yahoo!, AltaVista, HotBot, Lycos, Infoseek, Cadê.

2.1.1.2 - Sistema baseado em ontologias:

Visto que o sistema de recuperação de informação através de indexação de palavras-chaves apresenta limitações, os pesquisadores da área começaram a buscar propostas para solucionar esta lacuna.

Uma ontologia é uma definição formal de entidades e das relações entre elas, relevantes a um domínio, um entendimento compartilhado explícito em uma linguagem [GRUBER, 1995]. O poder das ontologias está no fato de que elas fornecem um entendimento comum, padronizado e compartilhado de um domínio, conhecimento este que pode ser comunicado ou compartilhado entre pessoas e sistemas aplicativos [HORROCKS *et al.*, 2000]. Com o sistema baseado em ontologias, procura-se aumentar o grau de precisão e a cobertura das buscas, mantendo a transparência dos recursos complexos de processamento semântico para o usuário que está efetuando a pesquisa, na medida do possível. Um sistema de busca baseado em ontologias apresenta algumas características importantes:

- possui ontologias explícitas;
- essas ontologias têm um papel proeminente, com vários componentes do sistema utilizando-as em suas tarefas.

2.2 – Web Semântica

A Web Semântica vem se apresentando como solução para ordenar o caos informacional existente na Web. É considerada uma visão para o futuro em que informação recebe significado explícito, tornando mais fácil para máquinas processar e integrar automaticamente a informação disponível na Web.

Segundo [BERNERS-LEE, 2001], a Web semântica será uma extensão da Web atual; porém apresentará estrutura que possibilitará a compreensão e o gerenciamento dos conteúdos armazenados na Web independente da forma em que estes se apresentem (como texto, som, imagem ou gráfico) a partir da valoração e formalização das descrições semânticas desses conteúdos. Agentes serão então desenvolvidos para usar a semântica para coletar conteúdos advindos de fontes diversas, processarem as informações e interoperar resultados com outros programas.

Para [GUHA, 2003], a Web Semântica conterà recursos que não somente corresponda a objetos digitais (páginas da web, imagens ou vídeos) como a web atual, mas também objetos do mundo real, como pessoas, lugares ou eventos.

Na Web Semântica, os documentos são anotados com meta-informação, que define quais informações ele contém [DAVIES, 2003]. Esta meta-informação, acompanhada de alguma teoria de domínio, na forma de ontologias, por exemplo, possibilitará uma Web que fornecerá um novo nível de serviços.

O par formado entre a meta-informação e uma ontologia faz com que os recursos sejam dispostos na web de forma mais abrangente. Isto faz com que mecanismos de recuperação de conteúdo atuem de forma mais precisa e com maior qualidade em suas tarefas [MOURA, 2001].

Segundo [DACONTA, 2003], a Web Semântica propõe solucionar vários problemas chaves das atuais arquiteturas de tecnologia da informação, tais como:

- Sobrecarga de informação: as ferramentas de busca enfrentam a dificuldade de executar pesquisas em documentos que não estão diferenciados em termos de assunto, qualidade e relevância. A tecnologia atual não consegue diferenciar a informação entre diferentes assuntos. É necessário haver informações de qualificação da própria informação para que seja possível classificá-las e tornar os processos de recuperação de informações mais eficazes.

- Integração de informações: a falta de padronização sintática, semântica e estrutural entre os documentos é muito grande, tornando o processo de integração, compartilhamento e resolução de conflitos entre informações em um problema muito difícil de ser solucionado. A heterogeneidade estrutural e semântica da Web atualmente é imensa e a maioria das propostas de integração ainda adota soluções com alto índice de centralização, tornando seu uso na Web inviável.
- Conteúdo não estruturado: um dos motivos do grande sucesso da Web atual é a sua liberdade de publicação de informações. Essa liberdade proporcionou uma enorme quantidade de documentos e recursos de todo tipo, tais como: banco de dados, artigos, programas, arquivos, etc. Devido a falta de padronização, essas informações são difíceis de serem abrangidas pelos mecanismos de pesquisa, ocasionando demora e ineficácia na localização de informações. A efetividade dos mecanismos de busca depende principalmente da maneira pela qual as informações foram estruturadas e catalogadas na Web.

Atualmente a web semântica se apresenta como uma área de pesquisa muito ativa que visa estender o papel dos computadores no suporte a diversas atividades humanas, através do uso de ontologias e anotações semânticas para catalogar, recuperar e compor dados e serviços de processamento disponíveis em rede.

Ontologias são desenvolvidas e aprimoradas constantemente para os mais variados domínios, enquanto as anotações semânticas ainda enfrentam dificuldades para cumprirem o seu papel fundamental dentro da Web semântica. As técnicas de anotação semânticas automáticas recomendadas para o grande volume de documentos existentes na Web são muito estudadas, porém ainda não se tornaram totalmente confiáveis quando aplicadas sobre um grande volume de dados.

2.2.1 - As camadas da Web Semântica propostas pelo W3C

O Consórcio da Web ou W3C (*World-Wide Web Consortium*) tem se dedicado a padronizar novas linguagens para definição de páginas e seus respectivos padrões, baseadas em conhecimento estruturado em ontologias. Na proposta de desenvolvimento da Web Semântica [BERNERS, 2003] é sugerida uma nova arquitetura de três camadas:

- **Lógica:** que define mecanismos para fazer inferências sobre os dados.
- **Ontologia:** que define as relações entre os dados;
- **Esquema:** que estrutura os dados e define seu significado;

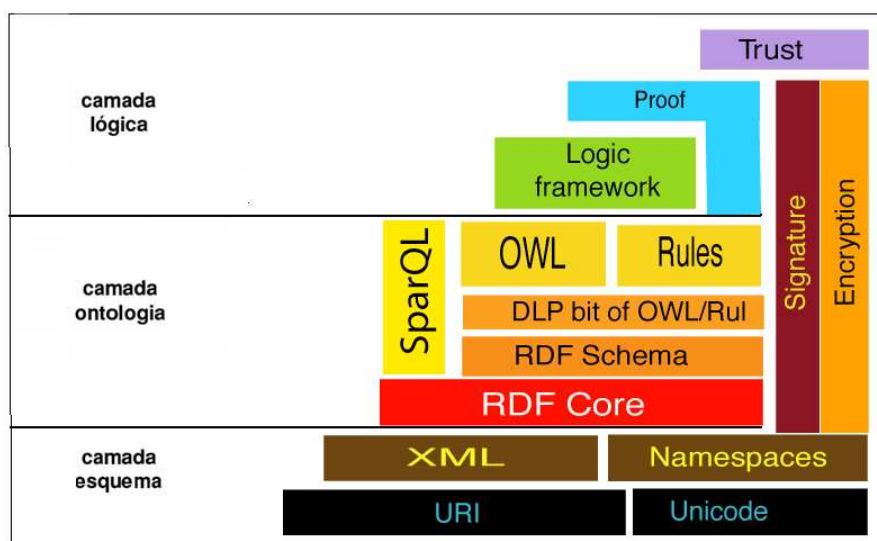


Figura 1: As camadas da Web semântica [BERNERS, 2003] + <http://www.w3.org/2005/Talks/0511-keynote-tbl>.

2.1.1.1 – URI + Unicode

Os componentes da base da arquitetura que representam a Web semântica garantem o uso padronizado do mesmo conjunto de caracteres (Unicode) e a identificação e localização de páginas de uma forma única (URI – *Uniform Resource Indicator*, indicador uniforme de recursos).

Segundo [LEE, 1998], são caracterizados por:

- **Identificador:** Um identificador é uma seqüência de caracteres com sintaxe restrita que faz referência a alguma coisa que tenha identidade.
- **Uniforme:** A uniformidade fornece vários benefícios, como permitir que diferentes tipos de identificadores de recursos sejam usados no mesmo contexto, permitir a introdução de novos identificadores de recursos sem

interferir nos já existentes e permitir a interpretação semântica uniforme de convenções sintáticas nos diferentes tipos de identificadores.

- Recurso: Um recurso pode ser observado como qualquer documento eletrônico, imagem, serviço ou coleção de outros recursos que possua identidade.

Tipos mais específicos de URI podem ser classificados como um localizador, um nome ou ambos. O tipo de URI mais conhecido é o URL (Uniform Resource Locator), que identifica o recurso através da representação de seu mecanismo primário de acesso, ou seja, sua localização na Internet.

2.1.1.2 - XML

As linguagens de marcação (markup languages) evoluíram desde o SGML (Standard Generalized Markup Language), para o HTML (Hypertext Markup Language) em 1980 e XML (Extensible Markup Language) em 1996.

Ao contrário da HTML, que através das marcas pré-definidas gerenciam os textos marcados e controlam sua representação estabelecendo ligações entre os documentos, a linguagem XML marca semanticamente um documento.

Segundo [DEITEL, 2003], “XML é uma tecnologia para criar linguagens de marcação para descrever dados de virtualmente qualquer tipo de uma maneira estruturada”. Sua sintaxe é semelhante ao HTML, que também é uma linguagem de marcação, mas o objetivo é outro. Enquanto HTML trata da formatação dos dados para sua exibição em navegadores, a XML fornece estrutura a dados de diversas naturezas.

XML é um padrão utilizado para marcação de documentos que contém informações estruturadas, ou seja, documentos que contém uma estrutura clara e precisa da informação armazenada. Esta estruturação define e separa claramente conteúdo, significado e apresentação. Assim os documentos em XML podem ser indexados com maior precisão que as páginas escritas em HTML.

Os dados contidos nos documentos XML podem ser exibidos em uma infinidade de maneiras, dependendo do dispositivo em que serão manuseados (telas de computador, celulares, PDAs etc.). Os documentos XML não contém, em si, as diretivas para exibição dos dados, e, para cada dispositivo-destino específico, podemos realizar uma transformação do documento originalmente em XML para um documento

passível de ser exibido ao usuário ou entendido e utilizado por outro dispositivo tecnológico. Esta transformação é realizada utilizando-se a linguagem XSL (eXtensible Stylesheet Language), e cada arquivo XSL contém as definições de exibição ou leitura de um ou vários dispositivos específicos (tela do computador, tela do celular, impressora, coletores de dados, outros sistemas de informação etc.), no formato que melhor convier (tabelas, gráficos, seqüência de caracteres etc.). O arquivo XML passa por uma transformação definida pelo XSL, e o resultado é um arquivo muito semelhante a um documento HTML comum. Desta forma, o trio XML, seu DTD específico e o XSL se apresentam como um conjunto de padrões que possibilitam o armazenamento, descrição significativa, intercâmbio e exibição dos dados de forma personalizada [ALVARENGA, SOUZA, 2004].

Um documento escrito em XML pode conter um conjunto infinito de *tags*, enquanto na linguagem HTML este conjunto de marcações é limitado. Isto é possível devido à linguagem XML permitir que os programadores criem suas próprias *tags* [LEE, 2001]. As *tags* são rótulos ocultos que demarcam seções de texto num documento.

O padrão XML permite troca de informações entre diversas plataformas e possibilita a descrição de dados em arquivos texto. A linguagem XML torna-se poderosa ferramenta para a publicação de informações na web [OLIVEIRA, 2002].

Os blocos de construção básicos em XML são os *elementos*, que são determinados por uma *tag* inicial e outra final, que consiste no nome do tipo do referido elemento. Um elemento pode conter outros elementos. Um elemento também pode possuir *atributos*, que são informações anexadas na *tag* inicial do mesmo. A Figura 2 abaixo representa um exemplo de alguns recursos de XML:

```
<library>
  <book isbn="25-8563-456-8">
    <author>Charles M. Schulz</author>
    <title>Being a Dog Is a Full-Time Job</title>
  </book>
</library>
```

Figura 2: Exemplo recursos da XML

A estrutura lógica de um documento XML é semelhante a uma árvore, onde seus elementos são estruturados em hierarquia. Segue a árvore que representa o exemplo anterior:

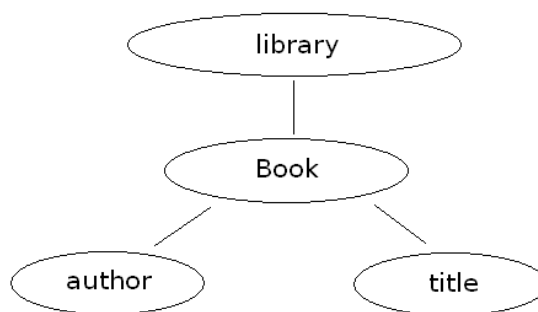


Figura 3: Exemplo de estrutura de documento XML em árvore.

No exemplo, os dados estão descritos por *elementos*, o que facilita seu tratamento, se o software que trata o conteúdo conhece este formato. Este formato pode ser especificado principalmente por duas linguagens para esquemas: DTD ou por esquemas XML (XMLS).

a) DTD:

Os DTDs determinam as regras, hierarquias e marcações criadas para caracterizar as informações do documento [BAX, 2001]. DTDs definem a estrutura e sintaxe de um documento, ajudando a validar se ele está ou não em conformidade com a estrutura definida. Não é necessário que um documento tenha uma DTD associada a ele, mas se tiver, deverá seguir a sua estrutura.

A vantagem na utilização da DTD é que o vocabulário fica documentado oficialmente e de modo preciso, interna ou externamente ao documento. Todas as regras de vocabulário estão na DTD, possibilitando que o autor adicione somente elementos ou atributos definidos na DTD e de acordo com a estrutura pré-definida [MARTIN, et al, 2001].

Abaixo, temos um exemplo de DTD:

```

<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE book (author, title)[
<!ELEMENT author id ID #IMPLIED>
<!ELEMENT title id ID #IMPLIED>
]>
<book isbn="25-8563-456-8">
  <author>Charles M. Schulz</author>
  <title>Being a Dog Is a Full-Time Job</title>
</book>
  
```

Figura 4: Exemplo de DTD

Segundo [CASTRO, 2001], a DTD possui algumas limitações:

*são escritas em uma sintaxe que possui pouca relação com XML e que um *parser* não consegue analisá-la;

*as declarações são globais, ou seja, não é possível definir dois elementos diferentes com o mesmo nome, mesmo em contextos separados;

*não é possível declarar para cada elemento os tipos de dados, como por exemplo, um inteiro, um real, uma data, etc.

b) XML Schema:

O XMLS foi proposto pela W3C para servir como uma linguagem de definição de tipos de documentos XML. São escritos através da uma sintaxe XML, podendo ser analisados por *parsers* XML.

Esquemas XML têm a mesma função das DTDs, mas são mais ricos. Pode-se definir tipo e formato exato dos atributos, número exato de instâncias de um aninhamento e há mecanismos de inclusão e derivação que proporcionam o seu reuso.

O XML Schema consegue suprir algumas deficiências das DTDs, o que torna o seu uso mais vantajoso do que a utilização das DTDs [CASTRO, 2001].

2.1.1.3 - RDF + rdfschema

O RDF(*Resource Description Framework*, ou modelo de descrição de recursos) é um *framework* recomendado pelo W3C para representação ou definição de metadados e informações na *Web*[KLYNE & CARROLL, 2004], promovendo interoperabilidade entre aplicações, que trocam informações compreensíveis por máquina, na *Web*. Com o RDF é possível referenciar objetos com URIs, assim como permite definir vocabulários para serem referenciados por URIs.

RDF adiciona mais semântica a um documento, com a vantagem de não precisar referir-se à sua estrutura [FENSEL 2001]. “RDF é uma linguagem baseada em XML para descrever a informação contida em um recurso” [DEITEL, 2003]. Um recurso pode ser uma página, um site inteiro ou qualquer item na *Web* que contém informação em algum formato. Com o RDF, os recursos na *Web* são descritos de uma forma neutra, sem descrever uma área de aplicação específica ou domínio de conhecimento. Em princípio, o RDF não define a semântica de nenhum domínio.

A sintaxe do RDF utiliza a linguagem XML para expressar o significado da informação. A utilização do XML para a construção dos modelos permite que as descrições sejam analisadas sintaticamente e interpretadas por aplicativos capazes de

processar documentos XML. Consequentemente, a XML e o RDF se tornam complementares. Enquanto a XML define a estrutura, o RDF permite expressar o significado associado aos dados.

A flexibilidade da XML permite ao RDF expressar a semântica através da representação padronizada de triplas, seguindo a formatação <objeto, predicado, sujeito>, escritas sob a estrutura sintática do XML.

O modelo de dados RDF é definido como:

- Recursos;
- Literais;
- Propriedades;
- Sentenças.

As sentenças são formadas por <objeto, predicado, sujeito>, onde:

- Objeto é um recurso;
- Predicado é uma propriedade;
- Sujeito é um recurso ou literal.

O relacionamento entre um recurso e um literal é chamado de sentença. A sentença relaciona um objeto a um sujeito através de um predicado. Exemplo:

- Sentença: Bill Gates é dono da <http://www.microsoft.com>.
 - Predicado: dono;
 - Objeto: <http://www.microsoft.com>;
 - Sujeito: Bill Gates.

Os recursos que representam os objetos das sentenças devem utilizar identificadores no padrão URI, pois representam um endereço único para cada recurso na Web.

As sentenças RDF podem ser representadas de três formas:

- Grafo: [KLYNE, 2004], enumerou diversos conceitos sobre RDF. Um destes conceitos é o Graph Data Model, onde as triplas são representadas como ligações nó-arco-nó, com o predicado (também

conhecido por propriedade) definindo o relacionamento entre sujeito e o objeto.

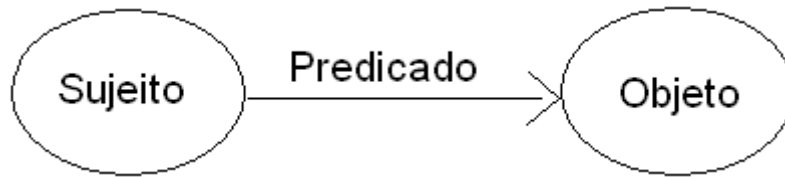


Figura 5: Modelo gráfico de representação RDF [KLYNE, 2004].

- XML: As sentenças RDF escritas em XML habilitam o intercâmbio entre máquinas, sem interferência humana através de várias aplicações e serviços.

```
<rdf:RDF xmlns:dc="http://purl.org/metadata/dublin_core#">
  <rdf:Description about=http://www.microsoft.com>
    <dc:Creator>Bill Gates</dc:Creator>
  </rdf:Description>
</rdf:RDF>
```

Figura 6: Sentença RDF escrita em XML.

- Tiplas: As triplas são acessíveis às aplicações que irão utilizá-las como entradas para suas operações conforme demonstrado na figura abaixo:

Objeto	Predicado	Sujeito
http://www.microsoft.com	dono	Bill Gates

Figura 7: Sentença RDF na forma de triplas.

A linguagem RDF satisfaz as três condições para que haja a representação de conhecimento:

- Interoperabilidade sintática: através do RDF Schema, ou seja, provê regras precisas, utilizando uma gramática, que estabelece uma sintaxe;
- Interoperabilidade estrutural: provendo uma representação aos dados especificando tipos e possíveis valores para cada forma de representação, através do mecanismo de <objeto, predicado, sujeito>;
- Interoperabilidade semântica: através do significado que é atribuído aos dados, através das triplas <objeto, predicado, sujeito>.

2.1.1.4– A camada lógica

A camada mais alta da proposta de Web semântica pelo W3C, composta por lógica, prova e confiança, ainda não tomou corpo. A lógica permite a especificação de regras que atuam sobre instâncias e recursos, enquanto a prova as executa, e a confiança avalia se a prova está correta ou não [KOIVUNEN & MILLER 2001]. Para que esta camada entre em operação, as camadas inferiores devem estar bem sedimentadas, o que ainda está acontecendo. Além do mais, sob o ponto de vista ontológico, não é interessante antecipar o uso de ontologias com regras, pois isto pode restringir a sua aplicabilidade. Porém regras podem ter utilidade para restringir atributos e exprimir axiomas.

2.3 - Ontologias

Esta é a camada mais importante e pesquisada da Web semântica. Ela é responsável por oferecer a expressividade necessária à representação de ontologias. Isso é feito aproveitando a extensibilidade de RDFs para definir restrições complexas e outras construções que implementam características de *frames* e lógica de descrições.

Nos últimos anos as ontologias vêm ganhando grande ênfase no campo da Ciência da Computação e Inteligência Artificial como meio de representar, compartilhar e reusar o conhecimento de forma legível para um computador.

Antes disso, ontologias era um tema relacionado apenas à Filosofia e pouco estudado pela comunidade da Ciência da Computação.

No contexto filosófico, ontologia é parte da filosofia que estuda o ser e seus relacionamentos. Esta definição é bastante ampla e permite diversas interpretações mais específicas de acordo com a área de aplicação, seja sistemas de informação, lingüística ou ciência da informação.

Outras definições de ontologias:

“Ontologia é um ‘catálogo de tipos de coisas’ em que se supõe existir um domínio, na perspectiva de uma pessoa que usa uma determinada linguagem.” [SOWA, 1999].

“Uma ontologia é uma especificação explícita dos objetos, conceitos e outras entidades que assumimos existirem em uma área de interesse, além das relações entre esses conceitos e restrições expressados através de axiomas.” [GRUBER,1995]

“As Ontologias são especificações das relações entre as entidades mais um conjunto de regras automáticas de inferência e ações associadas. É uma descrição formal dos conceitos e relacionamentos que existem dentro de um domínio. Isso significa que uma ontologia se relaciona com um vocabulário específico, uma linguagem específica e a conceitualização de determinado domínio.” [NETIC, 2003]

2.3.1 – Tipos de Ontologias

Diferentes tipos de ontologias, de acordo com seu assunto de conceituação podem ser delineados [GOMES-PEREZ, 1999]:

- Ontologias de Domínio: expressam conceituações que são particulares a um tipo de domínio, como por exemplo, eletrônica, medicina, mecânica, etc.
- Ontologias Genéricas: são similares às Ontologias de Domínio, mas seus conceitos são aplicáveis a vários campos, por exemplo, estados e processos. Geralmente os conceitos contidos nas ontologias de domínio são especializações dos conceitos das ontologias genéricas;
- Ontologias de alto-nível: descrevem conceitos muito gerais como espaço, tempo, evento, etc. Esses conceitos tipicamente são independentes de um problema particular ou domínio. Sendo assim, é bem razoável ter-se uma ontologia de alto-nível compartilhada por grandes comunidades de usuários.
- Ontologias de Aplicação: contêm todas as definições necessárias para modelar o conhecimento específico de uma aplicação. Geralmente são compostas por conceitos contidos nas ontologias de domínio e genéricas. Por terem um propósito tão específico, ontologias de aplicação geralmente não são reusáveis;
- Ontologias de Representação: dão suporte aos formalismos de representação. Estas ontologias não influenciam no domínio modelado. Elas provêm uma estrutura (*framework*) representacional para descrever as ontologias genéricas, de domínio e de aplicação.

2.3.2 – Profundidade Ontológica

Segundo [GUARINO & WELTY, 1998], a profundidade ontológica pode ser classificada em quatro níveis:

- Vocabulário: define a ontologia em sua forma mais simples, onde uma ontologia é apenas um vocabulário, definida por uma DTD ou um XML-Schema;
- Taxonomia: o significado dos termos é estabelecido pela definição de relacionamentos entre objetos e classes, subclasses e classes-pai. Esse tipo de ontologia normalmente é estabelecido por sistemas orientados a objetos;
- Sistema relacional: as ontologias também podem incluir relacionamentos não hierárquicos como nos diagramas de relacionamento de entidades e nos bancos de dados relacionais;
- Teoria axiomática: um axioma é uma afirmação lógica que não pode ser provada a partir de outras afirmações, mas podendo derivar outras afirmações. Axiomas podem ser usados para restringir valores das classes e suas instâncias, ou ainda para incluir regras mais genéricas.

2.3.3 – Componentes de uma Ontologia

Os principais componentes de uma ontologia são entidades, atributos, relações e restrições [CHANDRASEKARAN et al, 1999]:

- Entidades: representam classes ou conceitos em um domínio a ser modelado. Por exemplo, computador é um conceito dentro do domínio ciência da computação;
- Atributos: item de informação que descreve as propriedades das entidades. Por exemplo, cor preta é um atributo da entidade computador;
- Relações: relações podem ser hierárquicas ou não hierárquicas. Elas descrevem interações, ligações ou associações, entre as entidades e

atributos. A hierarquia de conceitos de uma ontologia é definida por relações hierárquicas (é-um, parte-de). Exemplo de relação hierárquica: note book é um tipo de computador;

- Restrições: são condições que podem ser impostas ao modelo, restringindo as entidades, atributos ou relações.

2.3.4 – Aplicações de Ontologias

[NOVELLO, 2003] indica que com a utilização de ontologias é possível definir uma infra-estrutura para integrar sistemas inteligentes no nível do conhecimento, trazendo grandes vantagens, como:

- Colaboração: possibilitam o compartilhamento do conhecimento entre os membros interdisciplinares de uma equipe;
- Interoperação: facilitam a integração da informação, especialmente em aplicações distribuídas;
- Informação: podem ser usadas como fonte de consulta e de referência do domínio;
- Modelagem: as ontologias são representadas por blocos estruturados que podem ser reusáveis na modelagem de sistemas no nível de conhecimento.
- Busca baseada em ontologia: recuperar recursos desejados em bases de informação estruturadas por meio de ontologias. Desta forma, a busca torna-se mais precisa e mais rápida, pois quando não é encontrada uma resposta exata à consulta, a estrutura semântica da ontologia possibilita, ao sistema, retornar respostas próximas à especificação da consulta.

Em conjunto com ao conceito de Web Semântica, as ontologias passaram a serem empregadas na Ciência da Computação por oferecer muitas outras utilidades, específicas para este domínio da ciência:

- Fornece um vocabulário para criar anotações semânticas;
- Permite criar novos termos, combinando os já existentes;
- Especifica formalmente o significado dos termos;
- Permite definir relacionamentos com termos de outras ontologias.

As ontologias podem ser aplicadas em várias aplicações que envolvem Sistemas Multiagentes (SMA), Sistemas de Apoio à Decisão, análise de imagens, manipulação das informações contidas na internet, etc.

Podemos classificar o emprego de ontologias em três grupos segundo sua missão: troca de informação, estruturação da informação e busca da informação.

- Troca de Informação: A troca de informação é bastante utilizada por SMA em suas interações. Para que isto seja possível, os agentes de uma sociedade devem compartilhar do mesmo conhecimento, o que pode ser obtido através de ontologias. A vantagem oferecida pelas ontologias de permitir a especialização de um conhecimento também é aplicada em agentes. Cada agente detém o conhecimento geral sobre o domínio mais o conhecimento sobre o seu papel a ser desempenhado na sociedade.

- Estruturação da Informação: A informação disponibilizada na internet tinha como único objetivo alcançar as pessoas. Esta foi tida como a primeira geração da Web. Não havia preocupação em estruturar a informação contida nas páginas já que ela era legível para as pessoas. A segunda geração foi a Web dinâmica que agregou sistemas de acesso à bases de dados tornando a informação flexível ao usuário. Com o surgimento dos agentes (termo usado na web para definir aplicações que interagem com pessoas ou outras aplicações na Web) percebeu-se que, além da informação léxica, era necessário compreender o contexto da informação. A terceira geração da Web visa adicionar semântica às informações – a Web Semântica. Ela tem como motivação a estruturação das informações segundo um contexto legível para agentes inteligentes, facilitando a troca de informação entre eles.

- Recuperação de Informação: conforme mencionado anteriormente, a Web atual ainda apresenta muitos problemas para que o emprego das ontologias possa ser aplicado em sua totalidade. Os engenhos de busca tradicionais não empregam totalmente, ou simplesmente não empregam ontologias. O Yahoo é um exemplo de uso parcial de ontologias, com uma estrutura montada em diretórios e subdiretórios, porém não segue

necessariamente uma hierarquia de classes. Mas engenhos especializados em domínios como o *brint* (<http://www.brint.com>) adotam ontologias como representação do conhecimento. Nele, o conhecimento é definido em categorias bem definidas.

2.3.5 – Linguagens para representação de ontologias

Segundo [MOURA, 2001], “ontologias provêm o mecanismo formal capaz de viabilizar o processamento semântico da informação através de uma máquina”. Para representação das ontologias, foram criadas diversas linguagens. Uma característica importante dessas linguagens é a representação em RDF/RDFS.

Diversos padrões e linguagens para construção e compartilhamento de ontologias na *Web* estão sendo criados, todos baseados no XML, com algumas diferenças de sintaxe de marcação (*tags*). Alguns exemplos são o SHOE (<http://www.cs.umd.edu/projects/plus/SHOE/>), a Ontology Markup Language (OML e CKML(<http://www.ontologos.org/>)) e a Resource Description Framework Schema Language (RDFS)(<http://www.w3.org/TR/PR-rdf-schema/>). Existe uma proposta de extensão do RDF e o RDFS chamada OIL (Ontology Interchange Language)(<http://www.ontoknowledge.org/oil/>) e seu sucessor DAML+OIL (<http://www.daml.org/>). O DAML+OIL (DARPA Agent Markup Language – Ontology Interchange Language) é uma linguagem baseada no XML, desenhada para possuir muito mais capacidade que este na descrição de objetos e no seu relacionamento; para expressar semântica e criar um alto grau de interoperabilidade entre *sites Web*. O OWL é uma linguagem de marcação semântica para publicação e compartilhamento de ontologias na *Web* e do DAML+OIL. Um exemplo de editor que suporta a criação cooperativa de ontologias baseado na *Web* é o Webonto (<http://webonto.open.ac.uk/>).

2.3.5.1 – DAML+OIL

A linguagem DAML+OIL surgiu da fusão de outras duas linguagens:

- DAML-ONT (*DARPA Agent Markup Language*, linguagem de anotação para agentes do Departamento de Defesa dos Estados Unidos);

- OIL (*Ontology Inference Layer*, camada de inferência para ontologias, ou ainda, *Ontology Interchange Language*, linguagem de intercâmbio em ontologias).

A versão inicial da linguagem DAML-ONT trazia consigo a integração de RDF, com XML embutido, com RDF Schema, o que tornava altamente compatível com os padrões da web e permitia a interoperabilidade com diversas ferramentas que vinham sendo desenvolvidas sob estes padrões [MCGUINNESS, 2002].

OIL foi a primeira destas linguagens, e teve como principal requisito a facilidade de adoção por parte dos desenvolvedores, servindo principalmente à comunidade ligada à Web semântica [HORROCKS et al 2000]. Ela foi resultado de diversas pesquisas na área de Lógicas de Descrição [MCGUINNESS, 2002], que são formalismos para representação de conhecimento onde interpretações fornecem significado, e estas interpretações fornecem a semântica formal da lógica [GRAU, 2004].

Dada a semelhança dos objetivos das duas linguagens, que é a geração de conteúdo compreensível por computadores, a fusão das duas foi um processo natural e obteve o aproveitamento do que cada uma delas possuía de melhor [MCGUINNESS, 2002].

2.3.5.2 – OWL

OWL (*Web Ontology Language*) é uma linguagem para definir e instanciar ontologias na Web. Uma ontologia OWL pode incluir descrições de classes e suas respectivas propriedades e seus relacionamentos. Foi projetada para o uso por aplicações que precisam processar o conteúdo da informação ao invés de apenas apresentá-la aos humanos. Ela facilita mais a possibilidade de interpretação por máquinas do conteúdo da Web do que XML, RDF e RDFS, por fornecer vocabulário adicional com uma semântica formal.

O OWL surgiu da necessidade de melhoria das linguagens para representação de ontologias existentes, pois o RDF não tinha todo o poder de expressividade necessário para a modelagem de ontologias. Por exemplo, RDFS não possui um mecanismo de restrição de cardinalidade.

As linguagens OIL e DAML foram criadas respectivamente em 1997 e 1999 como novas alternativas, mas também apresentaram algumas limitações. Em 2001 as duas linguagens foram unidas formando a linguagem DAML + OIL, que em 2002 foi

acrescida de requisitos de internacionalização (*Unicode*), apresentação e documentação, originando assim a linguagem OWL.

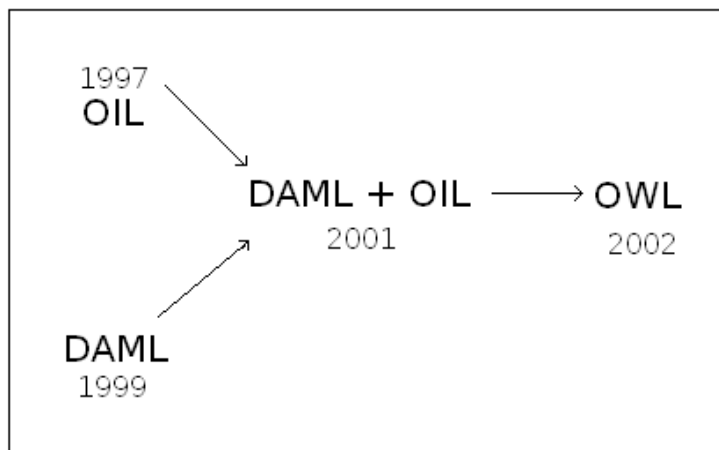


Figura 8: Evolução da linguagem OWL [HENDLER, 2005].

Segundo recomendações do Consórcio WWW [MCGUINNESS, 2005], existem três sublinguagens de OWL designadas para que comunidades específicas de implementadores possam utilizá-las de acordo com suas necessidades:

- **OWL Lite:** suporta aqueles usuários que necessitam principalmente de uma classificação hierárquica e restrições simples. É mais simples fornecer ferramentas que suportem OWL Lite que seus parentes mais expressivos. OWL Lite apresenta uma menor complexidade formal que OWL DL.
- **OWL DL:** foi definida para usuários que utilizem todos os recursos da linguagem OWL, mas que desejam assegurar-se da integridade computacional. Isto quer dizer que a utilização da linguagem deve obedecer a algumas restrições. OWL DL é assim chamada devido a sua correspondência com as lógicas de descrição, um campo de pesquisa que estudou a lógica que forma a base formal da OWL.
- **OWL Full:** é direcionada àqueles usuários que querem a máxima expressividade e a liberdade sintática do RDF sem nenhuma garantia computacional. OWL Full permite que uma ontologia aumente o vocabulário pré-definido de RDF ou OWL. Permite uma maior liberdade ao usuário com a restrição de não apresentar nenhuma garantia computacional.

As linguagens menos expressivas (OWL Lite e DL) estão contidas dentro das mais expressivas (OWL DL e Full), de maneira que uma ontologia definida numa linguagem menos expressiva é aceita por uma linguagem mais expressiva. Toda ontologia OWL Lite válida é uma ontologia DL válida, e toda OWL DL válida é uma ontologia OWL Full válida.

3 – Anotação Semântica

Este capítulo apresenta conceitos relacionados com a anotação semântica de documentos, incluindo os tipos de ferramentas, as técnicas possíveis para o processo de anotação e algumas ferramentas existentes que realizam este processo.

3.1 – Considerações sobre as anotações semânticas

A anotação semântica de documentos possui como objetivo facilitar a busca dos documentos no repositório digital. Com ela, é possível correlacionar termos (conceitos, instâncias ou propriedades) da ontologia a palavras, simples ou compostos, do texto que passou pelo processo de anotação semântica. Ela atribui às palavras que aparecem no documento ligações com suas descrições semânticas na ontologia.

As anotações semânticas possuem papel fundamental no desenvolvimento da Web Semântica, seja no sentido de criar novos documentos já com conteúdo semântico definido ou para prover semântica a documentos já existentes. A estrutura das anotações semânticas deve seguir as recomendações do W3C, na elaboração de sua estrutura e na utilização do modelo RDF como a linguagem para descrever os recursos.

O processo de anotação semântica é aplicável a qualquer tipo de texto (documentos HTML, documentos de texto comuns, campos de banco de dados, entre outros). Ele adiciona ao documento uma camada que descreve o seu conteúdo, permitindo que agentes de software possam recuperar informações de forma mais precisa, através da associação do documento a uma ontologia.

Na Web semântica, o documento dotado de uma estrutura semântica deve ser constituído de quatro componentes [GLONVEZYNSKI & GAUTHIER, 2005]:

- **Ontologia associada:** define conhecimento sobre o domínio de interesse ou sobre as informações descritas no documento;
- **Instâncias:** são chamados de instâncias de classe, representam as ocorrências individuais da ontologia.
- **Anotação:** definida como um bloco de texto que descreve relação do documento com uma instância da ontologia, o qual fica anexado ao conteúdo;
- **Conteúdo:** a informação contida no documento.

A anotação provê muitos benefícios quando é aplicada nos documentos: busca inteligente baseada em estruturas semânticas [BERNERS-LEE; HENDLER; LASSILA, 2001], geração de páginas adaptadas para melhorar a navegação de deficientes visuais [YESILADA, 2005], desenvolvimento de sistemas interativos enriquecidos com semântica, retornando respostas mais inteligentes ao usuário [KOGUT; HOLMES, 2001], etc.

A Web semântica precisa que o máximo de conteúdo relevante esteja anotado semanticamente para que se possam desenvolver sistemas que utilizem ao máximo este recurso e proporcionem novos serviços, ou melhorem os já existentes, sobre os recursos disponíveis atualmente. Pode-se citar, principalmente, sistemas de busca que utilizarão conhecimento de ontologias para efetuar as buscas e agentes inteligentes que possam executar tarefas mais complexas para o usuário de maneira automática, visto que as informações serão compreendidas com maior amplitude pelos agentes e aplicações na web [AGUIAR, 2007].

O processo de anotação semântica consiste de várias tarefas:

- Análise léxica, para separação dos tokens que formam as palavras simples ou compostas do texto. Análise sintática e classificação gramatical dos tokens. Identificação das palavras relevantes e eliminação de conectores.
- Extração dos radicais das palavras.
- Associação das palavras relevantes identificadas no texto as suas definições semânticas na ontologia.
- Armazenamento das anotações.

A representação das anotações podem ser intrusivas ou não intrusivas. Serão intrusivas quando são guardadas nos próprios documentos e não intrusivas quando são armazenadas em repositórios de anotações que apontam para os documentos que passaram pelo processo de anotação.

Para anotação de dados, em princípio, não é necessário possuir-se uma ontologia completa, definindo todo o domínio de aplicação. É necessário somente um vocabulário controlado, representando um esboço do domínio, já que o propósito principal é prover pontos de referência únicos e constantes.

Na figura abaixo, podem ser observadas as anotações semânticas referentes a um texto simples. As entidades presentes no texto são associadas à sua definição semântica [KIRYAKOV, 2003]:

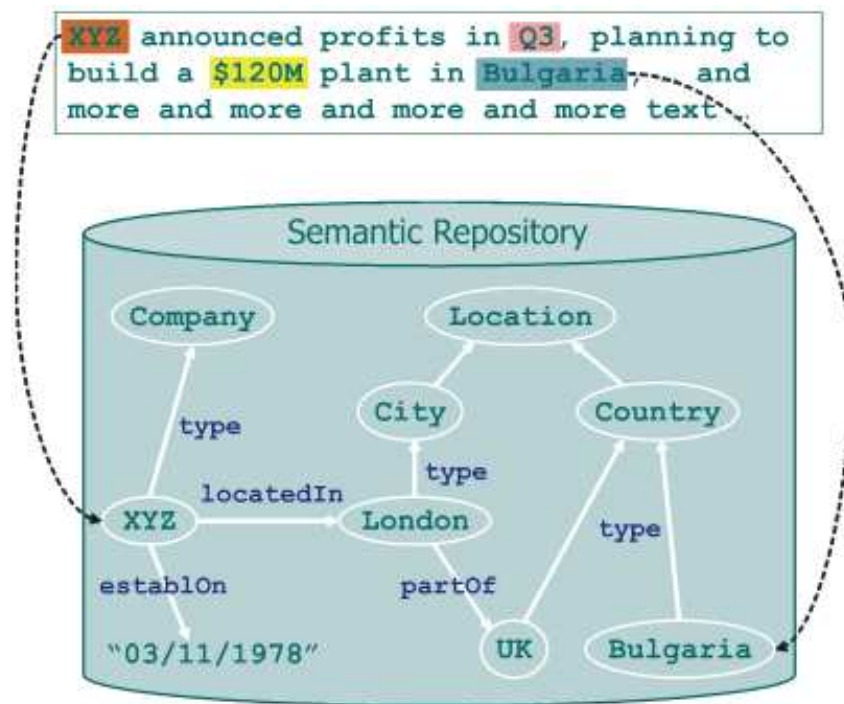


Figura 9: Exemplo de marcação semântica [KIRYAKOV, 2003].

As anotações devem ser feitas através da RDF, que é baseado em XML, pois é uma linguagem de marcação apropriada à representação de dados, cuja essência fundamenta-se na capacidade de agregar informações [BRAY, et al, 2004]. Como não possui *tags* predefinidas, permite definir novos elementos para criar vocabulários específicos, sendo assim possível incluir as marcações semânticas para a representação dos metadados no documento. A especificação XML define uma forma padrão de adicionar uma anotação a documentos [MENDONÇA, 2003].

O W3C recomenda que sejam usadas ontologias escritas em OWL para representação de conceitos sobre um determinado domínio de conhecimento a serem utilizados na Web Semântica e que as anotações semânticas, que descreverão a relação entre as instâncias das ontologias aos documentos, sejam escritos seguindo o modelo RDF [BECHHOFER, et al, 2004].

3.2 – Tipos de ferramentas para anotação semântica

Segundo [KOGUT, 2001], existem pelo menos três tipos de ferramentas que podem ser utilizadas para anotação semântica de documentos:

- Semi-automáticas: Associam palavras do texto a classes, instâncias e propriedades da ontologia, utilizando-se do julgamento humano. Esta associação geralmente é efetuada através de interfaces “arraste-e-solte”. A ferramenta *OntoMat* é um exemplo deste tipo de anotador.
- Automática: Aplica técnicas de processamento de linguagem natural (PLN), aprendizado de máquina e extração de informação, entre outras, para associar palavras à ontologia. Essas ferramentas podem utilizar ontologias padrão (por exemplo, *IEEE Standard Upper Ontology*) ou ontologias de domínios específicos (*Unified Medical Language System(UMLS)*).
- Híbrida: Utiliza as definições de anotação semântica semi-automática e automática para combiná-las em uma só ferramenta, ou seja, pode utilizar tanto o julgamento humano quanto técnicas de PLN para determinar as associações de palavras do texto com classes e propriedades.

Segundo [POPOV, 2003], apesar do método automático ainda não ser muito preciso no que diz respeito à identificação e classificação totalmente automática das entidades de um documento, esta tecnologia ainda está em desenvolvimento e já possui sistemas sendo desenvolvidos com um bom desempenho.

[BAYERI; GUT; LÜGEN, 2003] enfatizam que técnicas manuais possuem diversas limitações, pois tem que ser feita por pessoas e estas possuem limitações na utilização de esquemas ontológicos complexos e múltiplas ontologias. Além disso, os documentos e as ontologias podem mudar, exigindo assim modificações em suas marcações ou criação de novas [DINGLI; CIRAVEGNA; WILKS, 2003]. Anotações manuais tornam-se um processo exaustivo, considerando o volume de documentos existentes na web que precisam ser anotados para se adequarem à web semântica.

A criação automática de anotações semânticas ainda é um problema sem solução. Por isso, a melhor alternativa atualmente é investir em técnicas semi-automáticas para a geração de anotações. Ao contrário da geração automática, as técnicas semi-automáticas necessitam da intervenção humana em algum ponto do processo para identificar corretamente as relações entre os dados dos documentos e as entidades da ontologia. Outra vantagem é a possibilidade de utilizar múltiplas ontologias para anotar o mesmo documento com maior facilidade [REEVE; HAM, 2005].

Plataformas de anotação semântica semi-automáticas, também chamadas de SAP (Semantic Annotation Platforms), ainda podem ser classificadas com base no método de anotação utilizado. Há duas principais categorias, baseada em padrões e baseada em aprendizado de máquina. Além disso, plataformas podem utilizar métodos de ambos os tipos de categorias, chamada de Multiestratégia, a fim de obter vantagem com os pontos fortes de cada categoria, e compensar as deficiências dos métodos presentes em cada uma [REEVE; HAM, 2005].

Plataformas baseadas em padrões podem fazer a descoberta de novos padrões ou executar com padrões definidos manualmente. A maior parte dos métodos padrão-descoberta utiliza um conjunto inicial de entidades definido e o documento é varrido na busca dos padrões em que as entidades aparecem. Novas entidades são descobertas, juntamente com novos padrões. Esse processo continua recursivamente até que entidades não sejam mais descobertas ou quando o processo é interrompido pelo usuário. Anotações também podem ser geradas através de regras introduzidas manualmente para encontrar entidades no texto.

Plataformas baseadas em aprendizado de máquina utilizam dois métodos: Probabilístico e de Indução. O Probabilístico utiliza modelos estatísticos para prever os locais das entidades no texto.

O método Multiestratégia é capaz de combinar os dois métodos, baseado em padrões e baseado em aprendizado de máquina. Nenhuma plataforma atualmente suporta esta abordagem semântica de anotação, embora tenha sido executado em sistemas extração de ontologias, tais como On-To-Knowledge [KIETZ & VOLZ, 2000].

Segue abaixo o diagrama que representa as classes de plataformas de geração de anotação semântica e os métodos que podem ser utilizados.

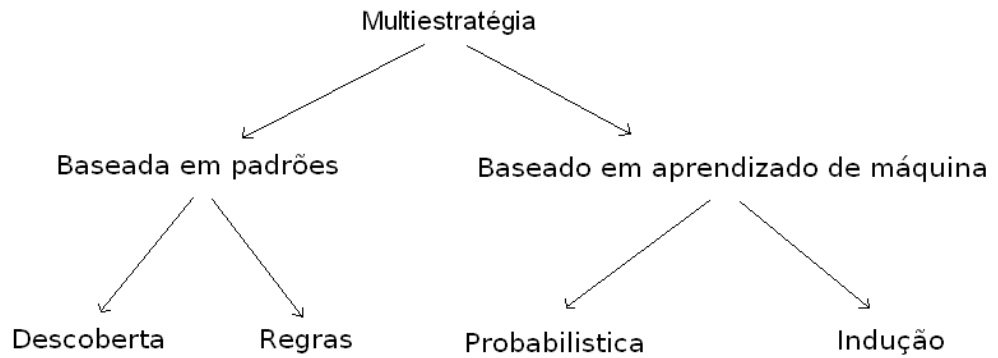


Figura 10: Classes de geradores de anotação semântica [REEVE; HAM, 2005].

3.3 – Características relevantes

As Plataformas de anotação semântica podem incluir um subconjunto de vários recursos, tais como APIs de acesso a ontologias, interfaces gráficas menos ou mais intuitivas aos usuários, editores para a base de conhecimento, armazenamento de recursos (por exemplo, repositórios RDF), etc. Nem sempre todas as características estarão presentes em todas as Plataformas.

Existem várias características desejáveis e que devem ser levadas em consideração para se fazer a seleção entre as ferramentas de anotação semântica existentes. As características estão relacionadas com seis propriedades principais:

- Tipo da ferramenta:
A ferramenta de anotação semântica pode ser Manual, Semi-automática, Automática ou Híbrida. A escolha de um tipo de ferramenta específica vai depender das necessidades do usuário, do seu conhecimento sobre o domínio de aplicação e sobre a(s) ontologia(s) disponíveis para este domínio.
- Quanto às ontologias:
Quatro características principais devem ser analisadas sobre este aspecto:
 - a) A primeira está relacionada com a possibilidade ou não de combinação de ontologias na mesma aplicação. Existem

ferramentas que não possibilitam a utilização de duas ontologias simultaneamente.

b) A segunda característica sobre este aspecto diz respeito à linguagem de representação da ontologia que a ferramenta suporta. Existem várias linguagens para fazer a representação de ontologias (RDFS, DAML+OIL, OWL,...). A maioria das ferramentas é limitada a uma ou duas dessas linguagens.

c) A terceira característica analisa se a ferramenta acessa a ontologia apenas localmente ou também possibilita o acesso remoto da mesma;

d) Por fim, é importante analisar os elementos da ontologia disponíveis para anotação (classes, instâncias, relações, triplas) e se a ferramenta possibilita a evolução da ontologia, ou seja, a definição de novas classes e subclasses para a mesma.

- Quanto às anotações:

É importante analisar se a ferramenta efetua as anotações de forma intrusiva, gravadas no próprio documento, ou não intrusiva, salvas em um servidor de anotações. Outra característica está relacionada com o formato da anotação gerada (XML, RDF , OWL , ...).

- Quanto aos documentos:

As ferramentas de anotação semântica suportam apenas alguns tipos de documento, como texto e imagens. Além disso, o formato em que estes documentos são salvos também deve ser levado em consideração (plain text, html, pdf, ...), pois cada ferramenta possibilita acesso a um número limitado de formatos. Outra característica deste aspecto deve avaliar se a ferramenta consegue acessar os documentos apenas localmente ou também possibilita acesso remoto.

- Arquitetura / Interface / Interoperabilidade:

É importante, principalmente para novos usuários, que a ferramenta apresente uma interface amigável, em conjunto com a facilidade de uso.

- Documentação:

Analisar se a ferramenta de anotação semântica possui documentação para consulta e se a mesma está atualizada. Também é importante verificar se existe algum tutorial explicando o seu funcionamento. Tutorial;

O próximo tópico descreve brevemente algumas Plataformas de anotação semântica disponíveis atualmente e mostra o seu desempenho testado empiricamente. As ferramentas apresentadas foram escolhidas a partir de uma revisão da literatura recente sobre o assunto. A idéia é realizar uma comparação, levando em consideração as propriedades apresentadas anteriormente, e ao final, a seleção de pelo menos uma delas para o Estudo de Caso.

3.4 – Ferramentas para anotação semântica

As pessoas que estudam formas de anotação semântica, seja manual ou automática, buscam diferentes pontos de vista sobre o problema. No processo de anotação Manual a investigação incide mais sobre a representação da anotação, compartilhamento e armazenamento, bem como mecanismos amigáveis de interface com o usuário para ajudar e facilitar as pessoas a escreverem as suas notas. Por outro lado, a investigação sobre ferramentas automatizadas de anotação incide mais sobre as formas de criar anotações, de acordo com determinados domínios de ontologias.

Uma ferramenta de anotação manual representativa é Annotea. Ela utiliza o RDF para descrever anotações e XPointer para localizar as anotações sobre os documentos [KAHAN, KOIVUNEN, PRUD'HOMMEAUX, SWICKD, 2001]. Além do Annotea, há algumas outras ferramentas manuais de anotação que podem potencialmente serem utilizadas para anotar semanticamente o conteúdo web, que incluem, mas não estão limitados a: CritLink, CoNote, Futplex, Annotator, ComMentor, e ThirdVoice. [HECK, LUEBKE, OERMARK, 1999].

Diversas ferramentas de anotação foram estudadas nos trabalhos de [DAVIES, STUDER & WARREN 2006]. A maioria das ferramentas analisadas usa em seu método de anotação o processo de extração de informação convencional (*information extraction - IE*) ou extração de informação baseada em ontologias (*ontology-based information*

extraction - OBIE). Estas duas formas de extração utilizam algoritmos que identificam automaticamente as entidades nomeadas nos textos, sendo que *OBIE* utiliza conhecimento da ontologia para efetuar a identificação das entidades, podendo fazer a ligação entre as entidades nomeadas encontradas nos documentos aos termos da ontologia mais direta e facilmente. Em sistemas que empregam técnicas de *IE* ou *OBIE*, quase sempre o usuário precisa anotar manualmente um conjunto de treinamento, para que o algoritmo aprenda a gerar regras de extração que possibilitarão mais tarde encontrar entidades nomeadas no texto automaticamente.

Existe também o método de anotação interativo, onde próprio usuário seleciona as entidades do texto a serem utilizadas na anotação semântica. Isso dispensa da ferramenta de anotação a tarefa de identificação automática das entidades nomeadas.

As ferramentas de anotação podem utilizar ontologias padrão (e.g., *IEEE Standard Upper Ontology*) ou ontologias de domínios específicos (e.g., *Ontologies of Professional Judicial Knowledge (OPJK)* [BENJAMINS, CASANOVAS, BREUKER & GANGAMI. 2005], e algumas até mais de uma ontologia. A maioria das ferramentas automáticas para anotação semântica adota um domínio de ontologia pré-selecionado para evitar que a geração automática seja um problema difícil. Todas as ferramentas têm de lidar com o problema de alinhar os conceitos das ontologias aos termos constantes nos documentos.

Abaixo, serão apresentadas algumas dessas ferramentas:

3.4.1 – Ont-O-Mat

O Ont-O-Mat [HANDSCHUH, STAAB, CIRAVOGNA, 2002] é uma implementação do S - CREAM, um framework que propõe anotações semânticas tanto manual quanto semi-automáticas das páginas web.

Ont-O-Mat adota uma técnica automática de extração de dados baseado no Amilcare, que é uma adaptação de IE (*Information Extraction*), um sistema designado para suportar anotações em documentos. O Amilcare baseia – se em um aprendizado (LP)², um algoritmo construído para aprendizado supervisionado baseado em regras de extração de informação em Lazy-NLP (*Natural Language Processing*). S - CREAM propõe um conjunto de heurísticas para pós - processamento e mapeamento dos resultados da extração de informação de acordo com uma ontologia. Com o Ont - O –

Mat. É possível acessar ontologias especificadas em mais de um tipo de marcação, como RDF e DAML + OIL. Mas cada ontologia somente pode ser acessada individualmente.

A ferramenta Ont - O - Mat pode armazenar páginas anotadas em DAML + OIL usando OntoBroker (*ontobroker.semanticweb.org/*), que é um servidor de anotações. Ele também fornece indexadores que podem pesquisar na Internet páginas anotadas para adicioná-las à sua base de conhecimento.

- Organização: AIFB (Institute of Applied Informatics and Formal Description Methods) na Universidade de Karlsruhe ;
- Tipo de ferramenta: semi-automática;
- Linguagem da Ontologia: DAML + OIL/RDF;
- Entrada de Ontologias: apenas uma por tarefa;
- Representação das Anotações: dentro de páginas da web;
- Evolução de Ontologia: Sim;
- Formato do documento: HTML ;
- Tipo de anotação: Intrusivo.

3.4.2 – MnM

O MnM [MOTTA, VARGAS-VERA, DOMINGUE, LANZONI, STUTT E CIRAVEGNA, 2002] é muito semelhante ao Ont - O - Mat. Fornece suporte para anotações automatizadas ou semi-automáticas.

MnM integra um navegador com um editor de ontologias. Além disso, fornece meios para acessar ontologias especificadas em diferentes formatos de marcação, da mesma maneira que Ont - O - Mat faz.

A ferramenta MnM também fornece API abertas, como a OKBC (*Open Knowledge Base Connectivity*), para fazer a ligação de servidores de ontologias e integração de ferramentas IE (*Information Extraction*). Além disso, ao contrário Ont - O - Mat, MnM pode gerenciar múltiplas ontologias ao mesmo tempo.

De acordo com os autores, a diferença entre os dois sistemas, Ont - O - Mat e MnM, é a sua filosofia. Embora Ont - O - Mat adota a filosofia que as marcações

devem ser incluídas como parte dos documentos, MnM armazena suas anotações tanto como marcações em documento como itens de uma base de conhecimento.

- Organização: KMI (*Knowledge Media Institute*) na Open University of United Kingdom;
- Tipo de ferramenta: híbrida;
- Linguagem da Ontologia: DAML + OIL/RDF;
- Entrada de Ontologias: um ou mais de um por tarefa;
- Representação das Anotações: dentro de páginas da Web e em uma base de conhecimento;
- Evolução de Ontologia: Sim;
- Formato do documento: HTML , texto;
- Tipo de anotação: Intrusivo.

3.4.3 – KIM

A plataforma KIM [POPOV, KIRYAKOV, KIRILOV, MANOV, OGNYANOFF & GORANOV, 2003] é uma parte do projeto SWAN (*Semantic Web ANnotator*). KIM consiste em uma ontologia KIM, uma base do conhecimento KIM, um KIM Server (com uma API para acesso remoto ou incorporação), e *front-ends* que fornecem acesso total às funcionalidades do KIM Server ao usuário.

A KIM ontologia (KIMO) representa uma ontologia de alto nível e define as classes, entidades e as relações de interesse. Ela possui cerca de 250 classes e 100 tipos de relações. A ontologia foi dividida em duas partes: PROTON (<http://proton.semanticweb.org/>), uma ontologia de alto nível genérica, e os módulos específicos da plataforma KIM, KIMSO e KIMLO.

A PROTON faz algumas distinções filosóficas, e descreve mais a fundo algumas entidades de importância geral (encontros, conflitos militares, governos e organizações, etc.). A intenção é que a ontologia seja utilizável para anotações de propósito geral, e também seja fácil de estender para domínios específicos.

Os autores optaram RDF (S) como a sua língua de representação para ontologias. O KIM básico contém uma entidade de descrição de informação para fins de anotação. Durante o processo anotação, KIM emprega uma técnica NLP para fazer a

extração da informação, que se baseia no GATE (General Architecture of Text Engineering) para extrair, indexar e anotar dados instanciados.

Na base de conhecimento são armazenados recursos com uso do Sesame (<http://www.openrdf.org/>), uma arquitetura para armazenamento de grandes quantidades de metadados, descritos em RDF e RDF Schema [BROEKSTRA; KAMPMAN; HARMELEN, 2002].

É possível utilizar várias ontologias no KIM, podendo ser instalado e configurado para atender um domínio específico. Possui um *plug-in* para o *Internet Explorer* que permite carregar a ontologia do servidor e efetuar anotação da página web carregado no navegador. Para que as pessoas possam visualizar essas informações anotadas, o texto aparece graficamente destacado através de diferentes cores.

- Organização: DERI (Digital Enterprise Research Institute) na Universidade Nacional da Irlanda em Galway, envolvendo também o GATE research team, bem como a OntoText laboratório de Sirma AI Ltd.;
- Tipo de ferramenta: automática;
- Linguagem da Ontologia: RDF/OWL;
- Entrada de Ontologias: KIM ontologia;
- Representação das Anotações: dentro de páginas da web;
- Evolução de Ontologia: Sim;
- Formato do documento: HTML;
- Tipo de anotação: Não-intrusivo.

3.4.4 – Smore

O SMORE [KALYANPUR, PARSIA, HENDLER & GOLDBECK, 2004] é uma ferramenta que permite aos usuários anotarem os seus documentos em RDF usando ontologias, fazendo a associação com termos específicos do documento com os elementos da ontologia.

O objetivo deste programa é proporcionar ao usuário um ambiente flexível, em que ele possa criar sua página da web sem esforços demasiados que envolvam a anotação dos documentos. O usuário deve ter um mínimo conhecimento sobre os termos em RDF e sua sintaxe. No entanto, ele deve ser capaz de classificar semanticamente os dados do documento, fazendo a correta associação sujeito-

predicado-objeto entre os elementos do documento e as suas referências nas ontologias existentes na Internet.

O usuário também pode criar sua própria ontologia a partir do zero, aproveitando fragmentos de ontologias existentes. O formato de linguagem para representação de ontologias suportado pela ferramenta é o OWL, conforme indicado pelo W3C.

- Organização: The Semantic Web Research Group na University of Maryland Institute for Advanced Computer Studies;
- Tipo de ferramenta: semi-automática;
- Linguagem da Ontologia: OWL;
- Entrada de Ontologias: uma ou mais de um por tarefa;
- Representação das Anotações: dentro de páginas da web;
- Evolução de Ontologia: Sim;
- Formato do documento: HTML , e-mail, imagens, texto;
- Tipo de anotação: Intrusivo.

3.4.5 – Annotea

O Annotea [KAHAN, KOIVUNEN, PRUD’HOMMEAUX & SWICK, 2001] faz parte dos esforços da W3C para firmar a Web semântica, sendo de código aberto e contribuindo para o avanço das normas W3C sempre quando possível.

A ferramenta permite a criação de anotações com metadados, gerando assim um ambiente compartilhado de anotações. As anotações são recursos Web e podem ser armazenados em um ou mais servidores de anotação. O servidor de anotação armazena as anotações realizadas em uma base de dados RDF e toda a comunicação entre a cliente e servidor utiliza métodos HTTP.

Com o Annotea, as anotações podem ser associadas a qualquer documento da Web ou apenas a uma parte selecionada dele, sem precisar mudar a sua estrutura. Quando o usuário recebe o documento, ele pode carregar as anotações referentes ao mesmo a partir de um ou vários servidores de anotação que fazem referência ao documento. Os metadados serão apresentados ao usuário por um cliente capaz de compreender os metadados e que seja capaz de interagir com a anotação do servidor remoto através de protocolo http.

O formato das anotações e descrição dos metadados é em RDF e o XPointer é utilizado para localizar as anotações nos documentos anotados.

As anotações podem ser privadas ou compartilhadas. Quando são armazenadas localmente e ficam restritas à máquina do usuário são privadas. As anotações compartilhadas são armazenadas na Web e podem ser lidas por todos os usuários registrados em um servidor remoto.

O Annotea foi desenvolvido para o editor e *browser* Amaya(<http://www.w3c.org/Amaya/>), mas como é de código aberto, nada impede que outros clientes possam implementar funcionalidades semelhantes.

- Organização: LEAD (Live early Adoption and Demonstration);
- Tipo de ferramenta: semi-automática;
- Linguagem da Ontologia: RDF;
- Entrada de Ontologias: uma ou mais de um por tarefa;
- Representação das Anotações: servidores de anotações;
- Evolução de Ontologia: Não;
- Formato do documento: HTML;
- Tipo de anotação: não-intrusivo.

3.5 – Comparação das Ferramentas

A Tabela 1 abaixo faz uma síntese das características relevantes apresentadas na sessão 3.3 sobre as ferramentas pesquisadas e apresentadas anteriormente. A primeira coluna apresenta o nome da ferramenta pesquisada. A segunda coluna informa o tipo da ferramenta, de acordo com o processo de anotação que a mesma é capaz de realizar, podendo ser automática, semi-automática ou híbrida. A terceira coluna representa as linguagens de representação das ontologias suportadas por cada ferramenta. A quarta coluna informa o número de ontologias que a ferramenta suporta simultaneamente. A quinta coluna informa a localização em que as anotações efetuadas serão salvas. A sexta coluna informa se a ferramenta analisada permite que novas classes da ontologia sejam definidas, permitindo a evolução da mesma. A sétima coluna apresenta o formato dos documentos suportados para a efetivação das anotações semânticas. Por fim, a

última coluna informa o tipo de anotação gerada por cada ferramenta, podendo ser intrusiva ou não-intrusiva.

Nome da Ferramenta	Tipo de Ferramenta	Linguagem da Ontologia	Entrada de Ontologias	Representação das Anotações	Evolução de Ontologia	Formato do Documento	Tipo de Anotação
Ont-O-Mat	Semi-automática	DAML + OIL/RDF	Apenas uma por tarefa	Dentro de páginas da web	Sim	HTML	Intrusivo
MnM	Híbrida	DAML + OIL/RDF	Uma ou mais de uma por tarefa	Dentro de páginas da Web e em uma base de conhecimento	Sim	HTML, texto	Intrusivo
KIM	Automática	RDF/OWL	Ontologia do KIM (Proton)	Dentro de páginas da web	Sim	HTML	Não-intrusivo
Smore	Semi-automática	OWL	Uma ou mais de um por tarefa	Dentro de páginas da web	Sim	HTML, e-mail, imagens, texto	Intrusivo
Annotea	Semi-automática	RDF	Uma ou mais de um por tarefa	Servidores de anotações	Não	HTML	Não-intrusivo

Tabela 1: Comparação de ferramentas de anotação semântica

Todas as ferramentas trabalham com documentos HTML sendo que muitas delas funcionam com navegadores Web associados em sua própria interface. Para representação das Ontologias, as linguagens DAML + OIL são usadas pelas ferramentas Ont-O-Mat e MnM, enquanto a linguagem RDF é utilizada nas ferramentas Annotea e KIM. Apenas as ferramentas Smore e KIM utilizam a OWL para representar as ontologias, conforme sugerido pelo W3C.

Na maioria das ferramentas pesquisadas, as anotações são intrusivas, ou seja, são guardadas junto ao documento anotado. As anotações são guardadas separadas do documento apenas no KIM e Annotea. Somente KIM e MnM implementam anotação automática. Todas as ferramentas pesquisadas suportam a evolução da ontologia (ao identificar entidades nomeadas ainda não descritas), exceto a ferramenta Annotea.

4 – Estudo de Caso

Depois de ter estudado conceitos sobre Web-semântica, sobre anotações semânticas e ter pesquisado algumas ferramentas que realizam a anotação semântica de documentos, a próxima etapa foi definir um domínio para efetuar anotações semânticas em documentos.

Buscando dar continuidade ao trabalho de [NUNES & FILETO, 2007] e analisando os trabalhos futuros do mesmo, o domínio selecionado foi a área jurídica. O principal objetivo é buscar atender algumas demandas do judiciário por ferramentas computacionais para gerenciar e facilitar a recuperação da informação produzida.

No Estado de Santa Catarina, além do Tribunal de Justiça, existem outras 112 comarcas espalhadas por seu território, produzindo cerca de 50.000 documentos anualmente. O sistema de automação do judiciário catarinense permite o acompanhamento de cada processo, a composição de sentenças, a troca de informações entre magistrados e operadores jurídicos. Esse sistema permite a busca de informações por palavras-chaves. As consultas são processadas de maneira meramente sintática, o que confere baixos índices de revocação e precisão nas respostas [NUNES & FILETO, 2007].

A proposta deste Estudo de Caso é associar alguns metadados baseados em uma ontologia aos documentos jurídicos para que seja possível estipular e processar consultas mais precisamente. Para isso, será definido um conjunto de documentos que serão anotados, uma ferramenta para auxiliar no processo de anotação semântica destes documentos e uma ontologia que represente o domínio escolhido.

4.1 – O documento anotado

Os documentos escolhidos para se realizar o processo de anotação semântica são acórdãos de processos do Tribunal de Justiça de Santa Catarina, conseguidos pelo site <http://tjsc6.tj.sc.gov.br/jurisprudencia/PesquisaAvancada.do>. Em Outubro de 2008, o sistema contava com aproximadamente 250.000 Acórdãos em sua base de dados.

Segundo o dicionário Aurélio, um acórdão é uma transcrição da decisão do órgão colegiado de um tribunal. Um acórdão é uma representação resumida da

conclusão a que se chegou, não abrangendo toda a extensão e discussão em que se pautou o julgado, mas tão-somente os principais pontos da discussão.

De acordo com o art. 165 do Código de Processo Civil brasileiro, os acórdãos devem ser proferidos em observância ao disposto no art. 458, ou seja, devem conter, obrigatoriamente, o relatório, a fundamentação e a parte dispositiva na qual se encontra a decisão propriamente dita. O acórdão também deve incluir uma ementa, conforme o art. 563 do Código Processo Civil, a qual constitui o resumo dos princípios fatos expostos ou o resumo do que se contém uma norma. O acórdão ou sentença deve levar a assinatura da autoridade a quem compete referendá-lo ou decretá-lo.

O acórdão, como as demais decisões judiciais, deve apresentar o nome de seu relator, dos membros componentes do órgão julgador (câmara, turma, seção, órgão especial, plenário etc.) e o resultado da votação. Caso a votação não seja unânime, o voto vencido, ou seja, o entendimento divergente, mesmo que de um membro apenas do órgão julgador deverá ser exposto no acórdão. Este registro é especialmente importante, pois as decisões não-unâнимes comportam embargos infringentes, por exemplo.

Todos os acórdãos devem seguir uma padronização, que está disponível no seguinte endereço:

http://tjsc25.tj.sc.gov.br/wiki/index.php/Formatação_de_Ementas_e_Acórdãos

O Anexo 1 deste trabalho apresenta um exemplo de acórdão utilizado no procedimento de anotação semântica realizado.

4.2 – A ontologia OntoJuris

A ontologia OntoJuris [NUNES & FILETO, 2007] é o princípio de uma ontologia jurídica que modela o conhecimento jurídico brasileiro para fins de recuperação de informação. Ela representa as principais características dos acórdãos, sentenças e outros tipos de decisões jurídicas. Este trabalho considera apenas acórdãos.

A ontologia OntoJuris está representada na linguagem OWL e se divide em duas classes principais:

- oj_documento;
- oj_FacetaDocumento.

Cada classe principal da Ontologia possui sub-entidades, que também possuem elementos filhos.

A classe `oj_FacetaDocumento` é a responsável por representar os componentes da ementa que consta no documento que será anotado. Seus atributos determinam as relações com os componentes do texto. A classe `oj_Documento` representa o documento acórdão, que possui como atributos os componentes que identificam cada acórdão.

Na figura abaixo se pode observar a estrutura base da ontologia OntoJuris:

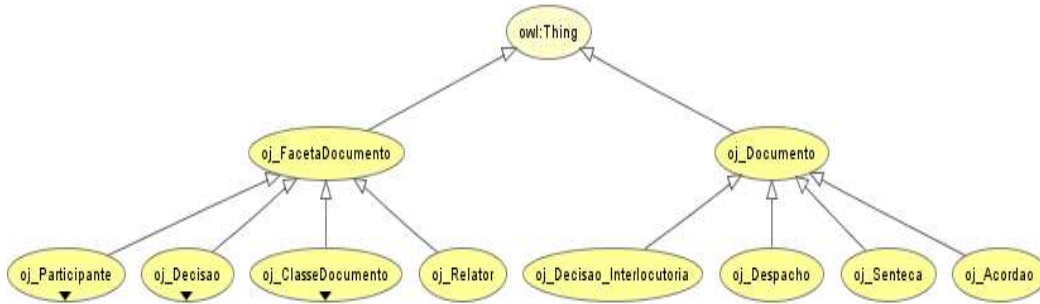


Figura 11: Rede Semântica representando a ONTOJURIS

Cada uma das classes principais pode possuir subclasses ou instâncias, formando uma estrutura hierárquica em forma de árvore. Segue abaixo a representação das subclasses de `oj_ClasseDocumento`:

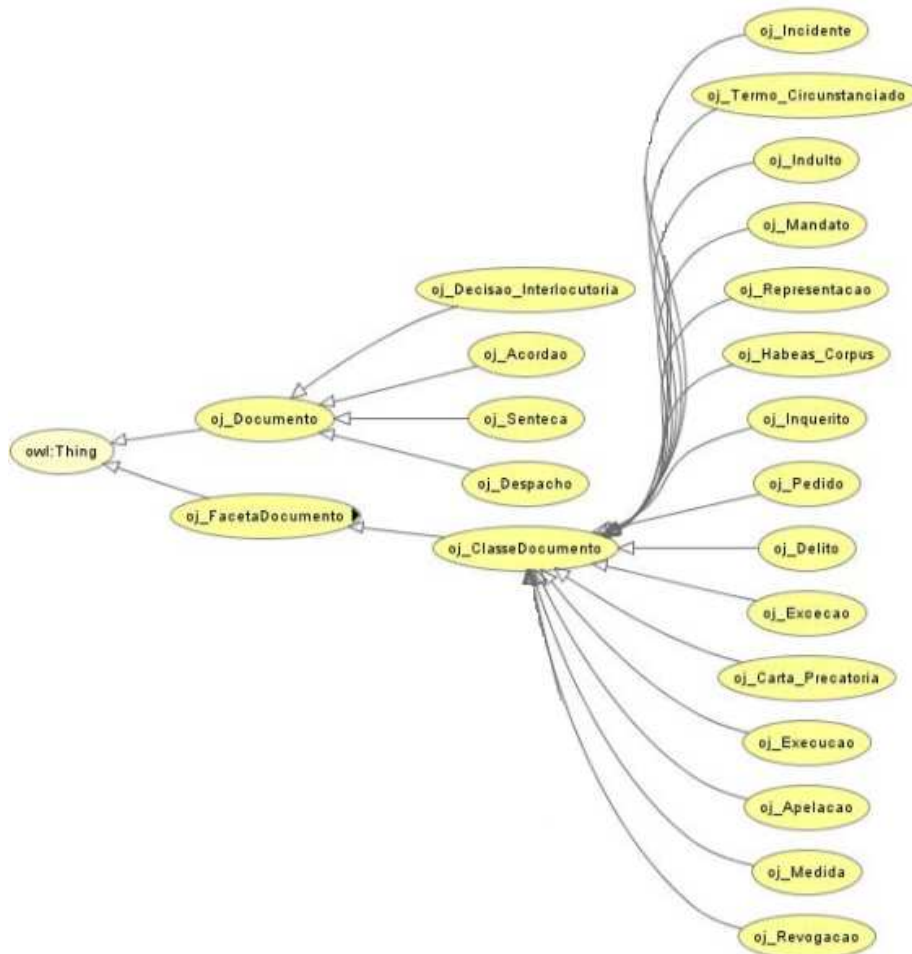


Figura 12: Subclasses da classe `oj_ClasseDocumento`

Seguem abaixo trechos do código OWL que representa a ontologia:

```
<?xml version="1.0"?>
<rdf:RDF
  xmlns="http://www.owl-ontologies.com/anotacaojuridica.owl#"
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:xsd="http://www.w3.org/2001/XMLSchema#"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
  xmlns:owl="http://www.w3.org/2002/07/owl#"
  xml:base="http://www.owl-ontologies.com/anotacaojuridica.owl">
<owl:Ontology rdf:about=""/>
```

...
Figura 13: Declaração de *Namespaces* da ontologia

A figura 13 apresenta o cabeçalho do documento OWL contendo a ontologia OntoJuris com as definições de *namespaces*. Antes de começar a definir os termos, é necessária uma indicação precisa da localização dos vocabulários específicos que serão utilizados na ontologia. Pode-se observar na figura 13 a declaração de namespaces referentes à RDF, XMLS, RDFS e OWL.

```
...
<rdfs:domain rdf:resource="#oj_ClasseDocumento"/>
</owl:DatatypeProperty>
<owl:DatatypeProperty rdf:about="#hasNumeroProcesso">
  <rdfs:range rdf:resource="http://www.w3.org/2001/XMLSchema#string"/>
  <rdfs:domain rdf:resource="#oj_Acordao"/>
</owl:DatatypeProperty>
<owl:DatatypeProperty rdf:about="#hasGrau">
  <rdfs:comment rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
  >Instâncias da justiça comum. Indica o grau do foro de
  tramitacao/julgamento do processo?</rdfs:comment>
```

...
Figura 14: Declaração de propriedades de classes da ontologia

A Figura 14 apresenta algumas propriedades definidas sobre a classe *oj_ClasseDocumento*. As propriedades descrevem características e relacionamentos entre classes OWL. Elas afirmam fatos gerais sobre os membros das classes e fatos específicas sobre os indivíduos [SMITH et al, 2004].

A tag *<rdfs:domain>*, presente na primeira linha da Figura 14, determina as propriedades que poderão ser aplicadas sobre os indivíduos da classe *oj_ClasseDocumento*. As propriedades definidas são *hasNumeroProcesso* e *hasGrau*, o que determina que quando um documento conter um número do processo ou um grau, definidos pela tag *<owl:DatatypeProperty>* ele será associado a classe documento. A tag *<rdfs:range>*, presente na quarta linha, limita a faixa de valores que os indivíduos de uma propriedade podem assumir.


```

...
<owl:Class rdf:ID="oj_Crime">
  <rdfs:subClassOf>
    <owl:Class rdf:about="#oj_ClasseDocumento"/>
  </rdfs:subClassOf>
...

```

Figura 15: Declaração de classe e subclasse

A Figura 15 apresenta como é a declaração de classes da ontologia utilizando a linguagem OWL. Todas as classes da ontologia são representadas desta forma. Cada objeto na linguagem OWL é um membro da classe *owl:thing*, que já vem definida por padrão da linguagem OWL, onde cada classe definida pelo usuário será implicitamente uma subclasse de *owl:thing*, que se torna a raiz de todas as classes definidas pelos usuários.

A tag *<owl:Class>*, presente na primeira linha da Figura 15, define o nome da classe que está sendo criada e a tag *<rdfs:subClassOf>*, da segunda linha da mesma figura, permite que se diga que uma classe mais específica esteja relacionada a uma classe mais geral [SMITH et al, 2004].

4.3 – O processo de anotação semântica

Depois da análise das Ferramentas de anotação semântica pesquisadas e apresentadas no Capítulo 3 deste trabalho, alguns critérios considerados relevantes foram escolhidos para se poder fazer a seleção entre as candidatas a realizar a anotação dos documentos. Um dos critérios mais importantes foi o formato de ontologia que a ferramenta suportava, já que a ontologia escolhida para o desenvolvimento do trabalho, OntoJuris, é representada através da linguagem de representação OWL. Facilidade de uso e documentação disponível também foram critérios relevantes para se fazer a seleção das ferramentas candidatas.

A ferramenta escolhida em um primeiro momento foi a ferramenta KIM, pois permite a geração automática de anotações em RDF, a anotação de documentos no formato HTML e a utilização de ontologias escritas em OWL. Porém a mesma apresentou vários problemas durante a instalação e incompatibilidades com o sistema operacional em execução. Tais problemas não são previstos na documentação da ferramenta, ocasionando esforço demasiado para tentar instalar e colocar a ferramenta em operação. Além disso, a ontologia representada em OWL deveria seguir a padronização da ontologia do KIM, Proton., para poder ser importada pela ferramenta.

Depois de várias tentativas e alterações nas configurações de instalação, a ferramenta finalmente foi instalada, porém apresentava erros em sua execução e mensagens de erro insatisfatórias. Além disso, a ferramenta se demonstrou extremamente lenta em sistemas com menos de 1Gb de memória RAM.

Esses e outras barreiras enfrentadas fizeram com que a ferramenta escolhida a priori fosse substituída por uma segunda opção. A ferramenta escolhida foi o Smore, pois é uma ferramenta para anotação semântica semi-automática que foi projetada para permitir aos usuários a marcação de documentos HTML em RDF utilizando Ontologias em OWL, conforme as indicações da W3C. Além disso, a ferramenta apresenta documentação e tutoriais que facilitaram o procedimento de instalação e operação da ferramenta.

Ao contrário do KIM, é uma ferramenta mais simples. Além disso, o processo de instalação do Smore foi muito mais simples, exigindo apenas que existisse uma versão igual ou superior da plataforma Java SE 1.4 instalada no sistema operacional. Segue abaixo a interface da ferramenta:

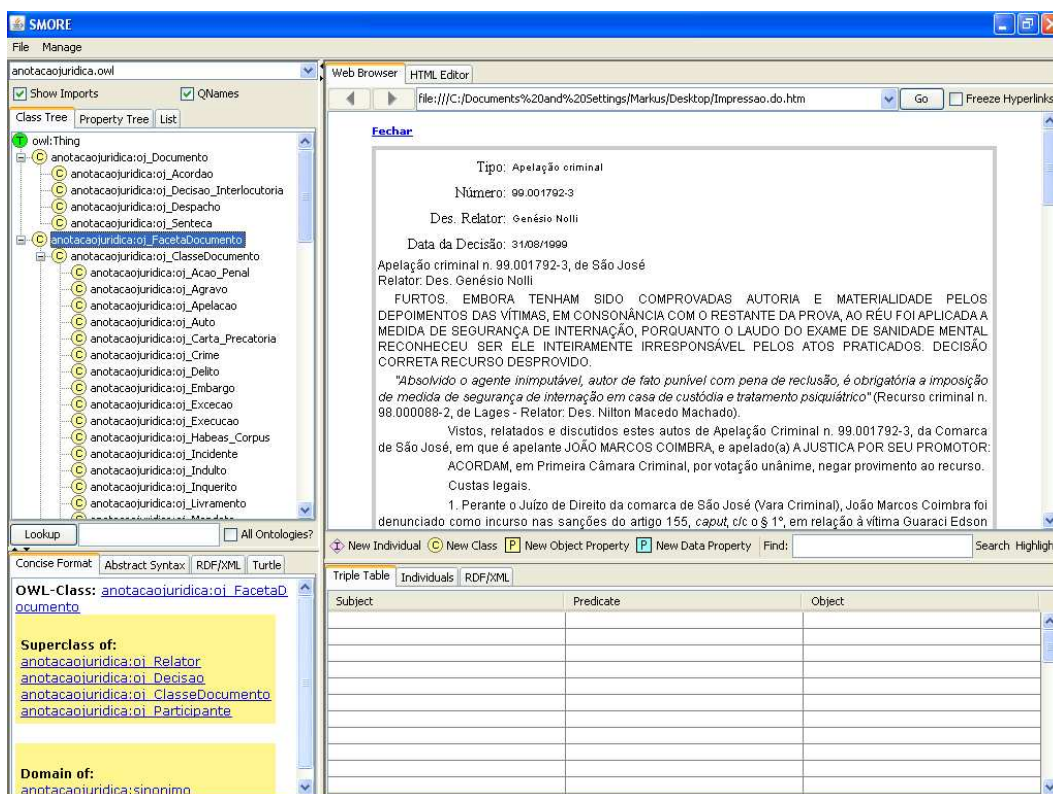


Figura 16: Interface da ferramenta Smore

Conforme mostra a figura 16, a ontologia é carregada na aba lateral esquerda. Nela pode-se navegar pela ontologia, observando suas propriedades e instâncias. Logo abaixo são apresentadas as informações do código OWL referentes à classe selecionada

na ontologia. A página HTML é carregada na aba à direita e suporta tanto documentos locais quanto remotos. As anotações são realizadas logo abaixo da aba com a página em HTML, seguindo o padrão de triplas RDF, sujeito – predicado – objeto.

No Smore, as anotações podem ser definidas de duas formas:

- a) **Seguir o padrão araste e solte:** identificando uma instância de uma classe da ontologia no texto do documento, o texto deve ser selecionado e arrastado até a aba abaixo, para se fazer a definição da tripla RDF que será inserida na anotação do documento. Na guia *Triple Table*, o texto selecionado deve ser arrastado até o campo *subject*. A classe da ontologia também deve ser selecionada e arrastada até a mesma guia e acomodada no campo *object*. No campo *predicate* deve-se definir qual é o tipo de relação entre o termo selecionado no texto e a classe na ontologia.

- b) **Definir diretamente no texto:** na página HTML, o texto que fará correspondência a alguma classe da ontologia deverá ser selecionado. Depois disso, o botão direito do mouse deverá ser apertado sobre o texto e a opção *Creat New Individual* deverá ser selecionada, abrindo assim uma nova janela na tela. Nesta nova janela, basta selecionar a classe da ontologia que fará associação com o texto especificado e clicar no botão *Creat New Individual* para confirmar a ação.

Também é possível criar novas classes ou subclasses para a ontologia seguindo os mesmos passos apresentados. A única alteração será escolher a opção *Create a New Class* depois de realizar a seleção do texto no documento que será anotado. Maiores detalhes serão apresentados na próxima sessão deste trabalho, onde o procedimento de anotação semântica efetuado será apresentado.

A ferramenta Smore utiliza o método intrusivo de anotação, onde as anotações deverão ser salvas no formato RDF no próprio documento que está sendo anotado. Por esta razão, amostras dos acórdãos provenientes da ferramenta de busca do site do Tribunal de Justiça de Santa Catarina foram salvos localmente, já que as anotações serão incluídas na estrutura do documento anotado.

A Figura 17 abaixo representa um esquema geral do processo de anotação realizado sobre um documento. Com o documento carregado e a ontologia disponível,

faz-se o relacionamento entre os termos presentes no documento com as classes definidas na ontologia através da interface do Smore, onde as triplas RDF podem ser definidas e o relacionamento efetivado. Depois disso, as anotações são geradas automaticamente pela ferramenta.

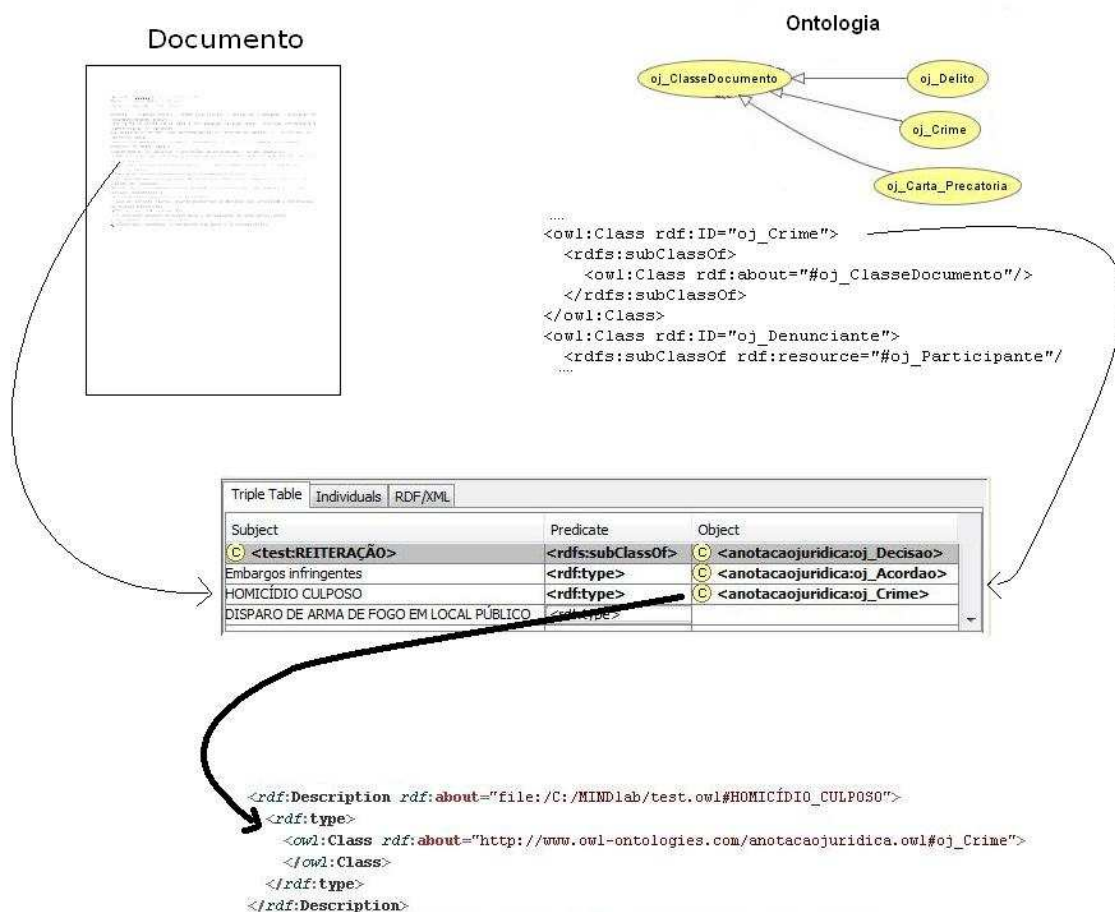


Figura 17: Exemplo de anotação gerada pela ferramenta Smore

4.3.1 – Etapas do processo de anotação

Esta seção apresenta a descrição do procedimento de anotação realizado sobre um documento com a ferramenta Smore. Tal ferramenta está disponível para *download* no endereço do link abaixo:

<http://www.mindswap.org/2005/SMORE/smores50.zip>

Para instalar o aplicativo, basta realizar o *download* do arquivo *smores50.zip*, descompactá-lo no sistema e executar o arquivo *runme*. Para que o arquivo seja executado corretamente, é necessário que uma versão igual ou superior da plataforma Java SE 1.4 esteja instalada no sistema operacional. Quando este arquivo é executado, a

interface da ferramenta, apresentada na Figura 14 deste trabalho, é exibida na tela. Seguem abaixo os passos que demonstram como é feita uma anotação:

***Passo 1:** abrir a ontologia que será utilizada para o processo de anotação.

O primeiro passo é abrir a ontologia que será utilizada para o processo de anotação. Ela poderá estar salva no próprio computador ou ser uma ontologia que está disponível na *web*. No caso, a ontologia OntoJuris está salva localmente. A Figura 18, ilustrando o procedimento:



Figura 18: Localização da ontologia com a ferramenta Smore

***Passo 2:** Carregar o documento que será anotado.

O acórdão que passará pelo processo de anotação semântica foi obtido através de pesquisas no site do Tribunal de Justiça de Santa Catarina e também está salvo localmente, no próprio computador. Para acessá-lo na ferramenta, basta digitar o seu endereço local na barra de endereços do *Web Browser* que o Smore disponibiliza em sua interface gráfica. A Figura 19 representa a ação a ação tomada:

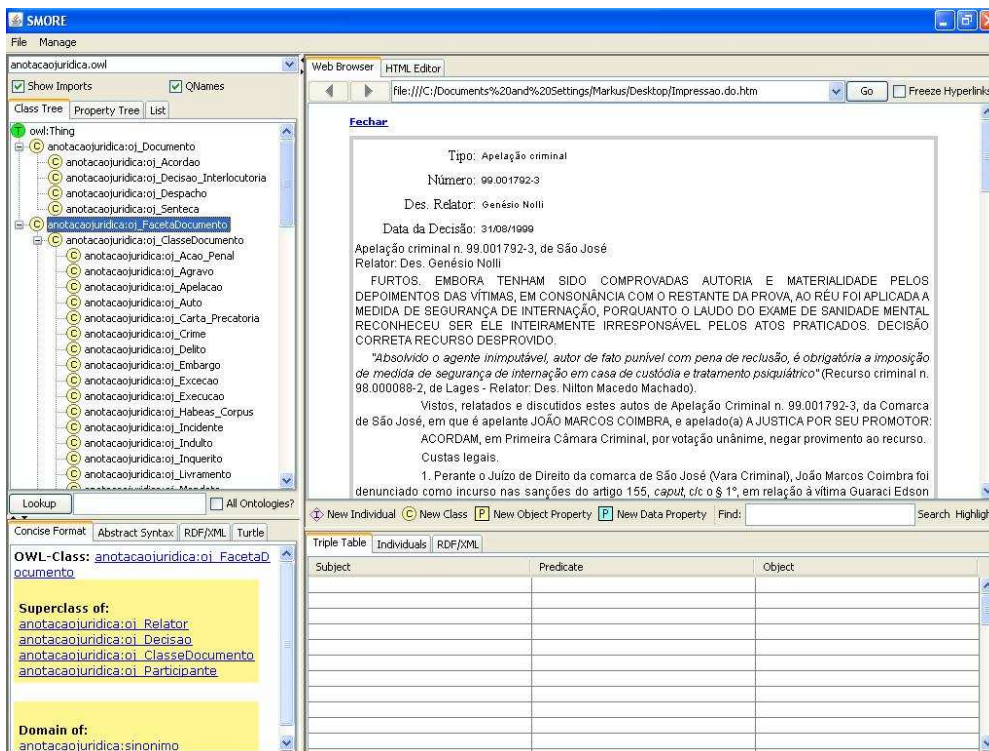


Figura 19: Documento que será anotado carregado no Smore

***Passo 3:** identificar no documento os termos que fazem referência a alguma classe da Ontologia OntoJuris.

Depois de estudar a ontologia desenvolvida no trabalho de [NUNES & FILETO, 2007] , a formatação definida para os acórdãos e observar alguns dos documentos, foi possível definir um padrão e identificar aonde os elementos do texto que fazem correspondência com os termos da ontologia estariam localizados. O Anexo 2 deste trabalho representa o padrão que os documentos apresentam, destacando entre colchetes os metadados da ontologia que possivelmente deverão ser referenciados no processo de anotação semântica.

O procedimento de anotação segue identificando no documento carregado os termos que fazem referência a alguma classe da Ontologia OntoJuris. Quando um termo do texto é identificado, deve-se fazer a seleção do mesmo e, com o botão direito do mouse, identificar qual a ação será tomada pela ferramenta com o texto selecionado. As opções possíveis são:

- *New Individual*: permite fazer a relação de termos do documento com classes da ontologia carregada, determinando que o termo selecionado seja associado a uma determinada classe da base de conhecimento. Um exemplo seria fazer a ligação do termo do texto “João Marcos Coimbra” como sendo do tipo *oj_Apelante*, referenciando a respectiva classe da OntoJuris;
- *New Class*: permite a evolução da ontologia base carregada no início do procedimento. Com esta opção, é possível definir novas classes para a ontologia. Com esta opção também é possível definir se a nova classe criada é uma subclasse ou superclasse de outra já existente na ontologia. Um exemplo seria selecionar o termo “RECURSO DESPROVIDO” no texto e defini-lo como sendo uma nova classe da ontologia, subclasse de *oj_Decisão*;
- *New Object Property*: permite a criação de uma nova propriedade sobre determinado termo do documento que referencia uma classe da ontologia. É possível fazer a definição das *tags domain* e *range*, que representam respectivamente a classe da ontologia e a faixa de valores que os indivíduos da propriedade definida poderão assumir. Um exemplo seria a definição de que a classe *oj_Decisão*, necessariamente precisa ter um *oj-Relator* associado e ele deve ser do tipo literal;

- **New Data Property:** permite a definição de um tipo de dado para determinada propriedade. Um exemplo seria especificar que o número do acórdão necessariamente deve ser um inteiro.

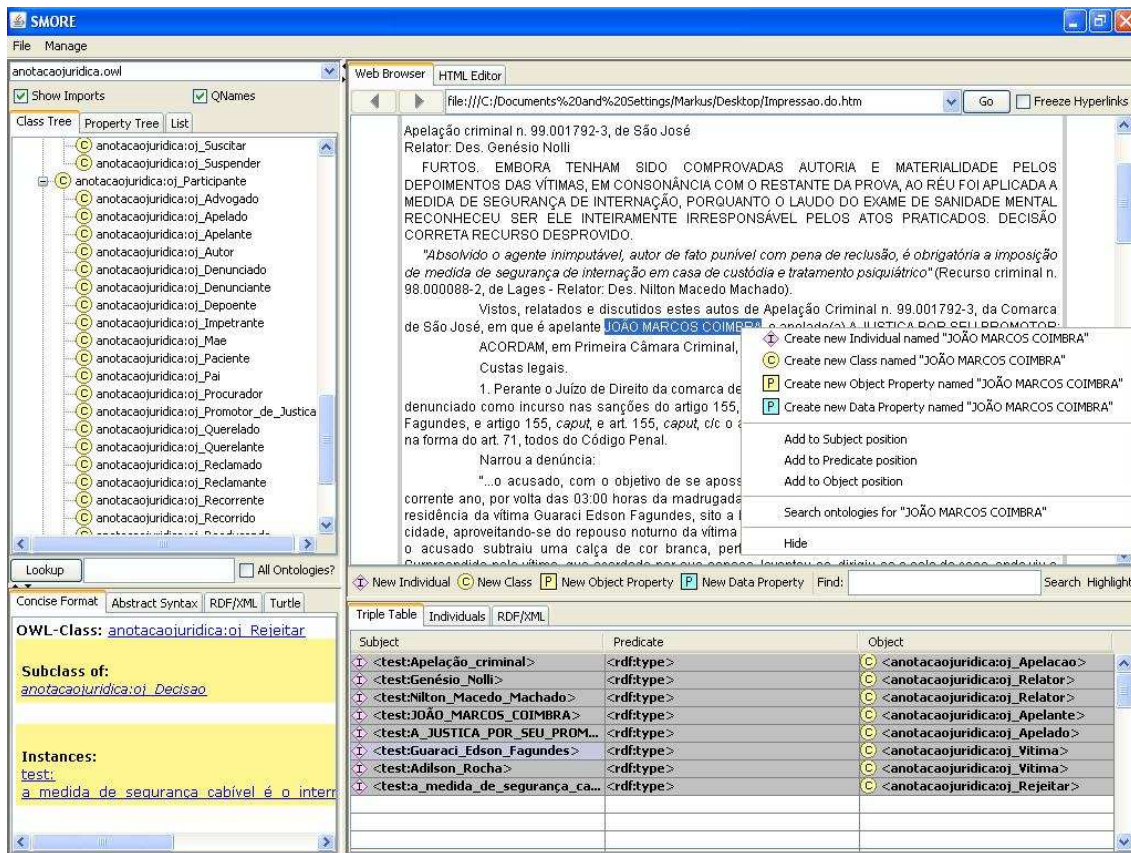


Figura 20: Seleção de termos do documento

***Passo 4:** Definição da tripla RDF utilizando a interface do Smore.

Com o texto selecionado, como demonstrado na Figura 18, escolhendo a opção *Create new Individual*, a ferramenta abre uma janela onde devem ser definidas as classes da ontologia as quais o termo presente no texto do documento será relacionado. Depois de escolher a classe da ontologia, deve-se clicar sobre o botão *Create Individual* na interface da ferramenta. A ação realizada pelo sistema será importar os dados cadastrados para a tabela de triplas, localizado abaixo do *Browser* onde o documento que está sendo anotado está carregado. Também é possível adicionar um termo diretamente na tabela que representa a tripla que irá gerar a anotação posteriormente, para isso, basta selecioná-lo e arrastá-lo até o campo desejado na tabela. O termo pode ser adicionado nas posições *subject*, *predicate* ou *object*.

A figura 21 representa a criação de uma nova instancia da classe *oi_Apelação*:

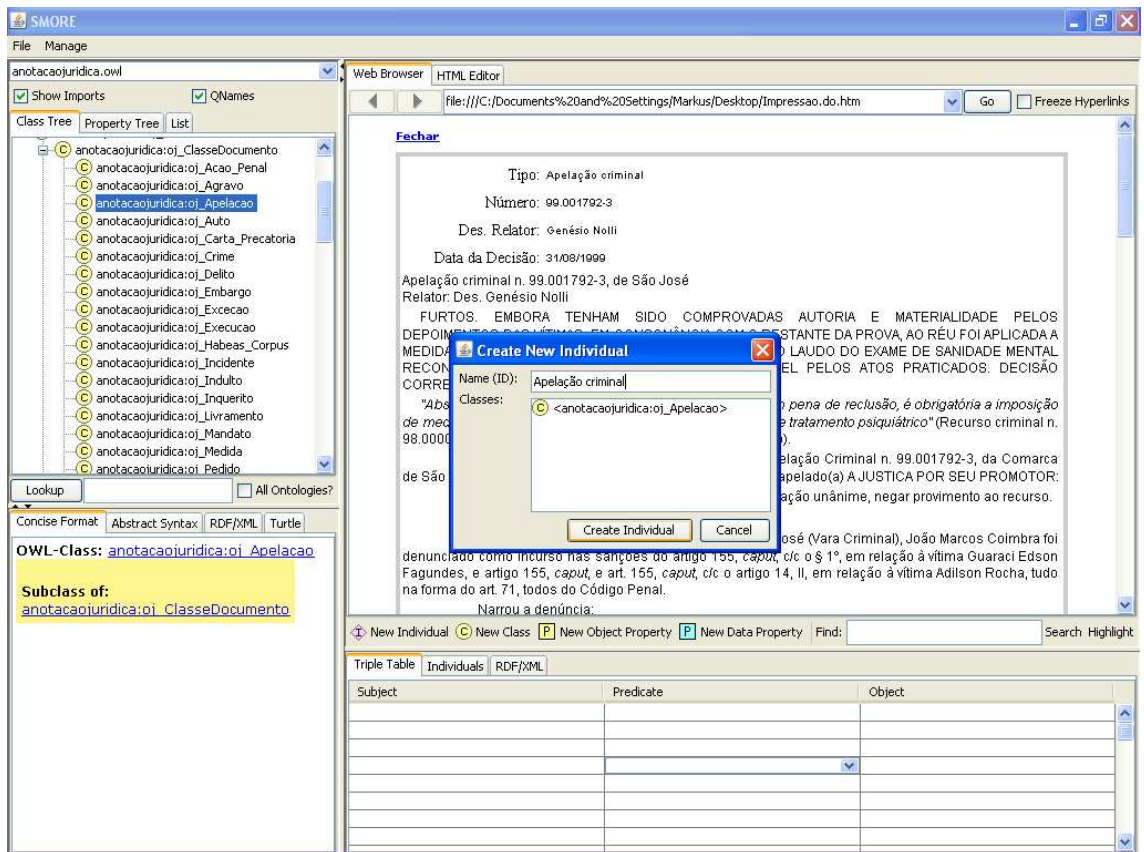


Figura 21: Criação de uma nova instancia

Os termos são carregados nos campos *subject*, *predicate* e *object*:

- *subject*: representa o texto selecionado no documento. É a ocorrência de uma classe da ontologia;
- *predicate*: representa o tipo de relação entre os termos presentes nos campos *subject* e *object*. As relações possíveis são *type*, *subClassOf*, *domain* e *range*.
- *Object*: representa a classe da ontologia a qual o termo selecionado no documento será relacionado.

Depois que a tripla é definida utilizando a interface do Smore, a ferramenta gera automaticamente o código RDF que representam as anotações. As anotações geradas pela ferramenta podem ser acessadas clicando sobre a guia RDF/XML, localizada ao lado direito da guia *Individuals*, que traz as informações referentes aos termos do texto que já foram referenciados pela anotação realizada.

A Figura 22 abaixo representa o procedimento realizado para gerar a anotação em RDF:

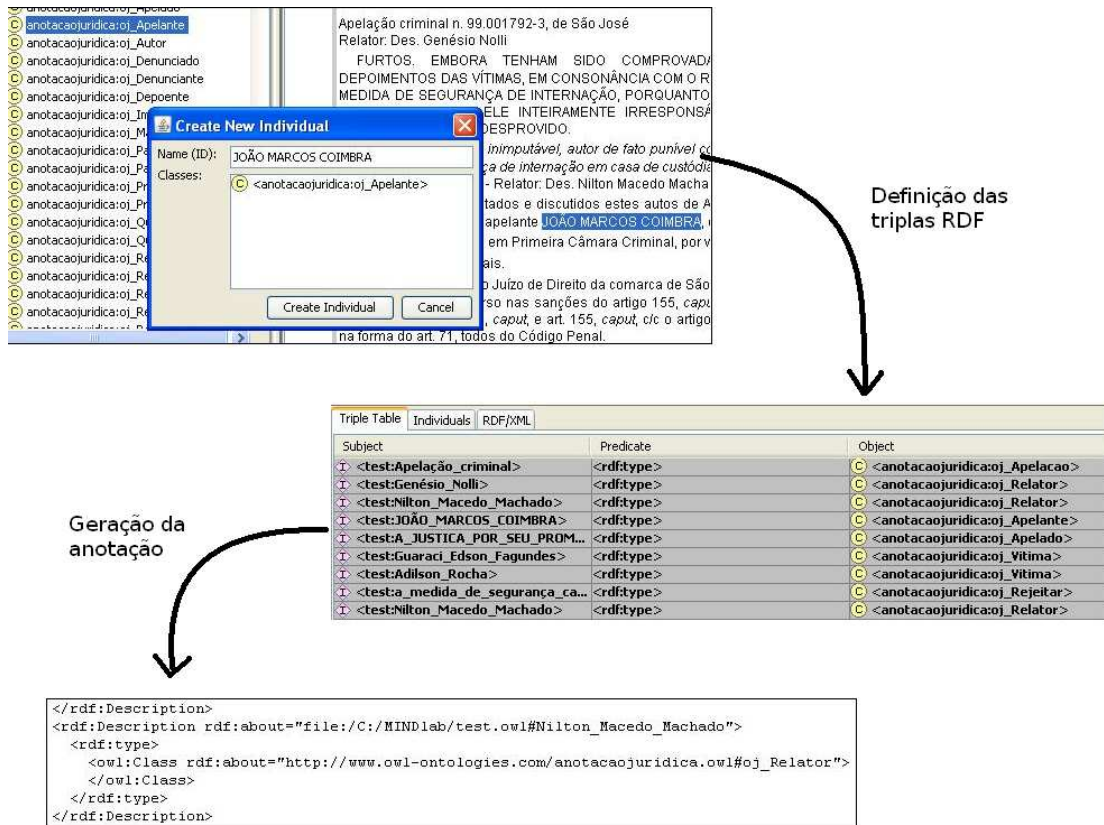


Figura 22: Geração das anotações em RDF

A definição de novas instâncias sempre é realizada desta forma: os termos do documento carregado no *web-browser* do Smore que farão referência às classes da ontologia devem ser identificados e selecionados. Depois disso, segue a definição da tripla RDF que irá efetivamente realizar esse relacionamento e gerar o código da anotação semântica resultante.

Depois que a tripla RDF é definida, o código RDF da anotação é gerado. A anotação faz com que o significado dos termos selecionados no texto deixem de ser analisados apenas sintaticamente, expressamente como são escritos, e passem a ter também um significado semântico, de acordo com as definições das classes da ontologia a quais foram relacionados através do Smore.

***Passo 5:** Salvar as anotações no documento.

As anotações deverão ser salvas no próprio documento que foi utilizado para se fazer as marcações. Para isso, deve-se copiar o código RDF gerado pela ferramenta, abrir o documento do acórdão em qualquer editor de textos e colar o código RDF antes da tag *<html>* da página e salvar o documento novamente. Com isso, o procedimento

de anotação termina e os metadados em RDF que representam as anotações estarão disponíveis para ferramentas de buscas semânticas.

4.3.2 – Código RDF gerado

Nesta seção é apresentado um exemplo de código RDF gerado pela ferramenta e explicação dos seus componentes.

```
<?xml version="1.0"?>
<!DOCTYPE owl [
<!ENTITY owl "http://www.w3.org/2002/07/owl#">
<!ENTITY xsd "http://www.w3.org/2001/XMLSchema#">
<!ENTITY rdf "http://www.w3.org/1999/02/22-rdf-syntax-ns#">
<!ENTITY rdfs "http://www.w3.org/2000/01/rdf-schema#">
]>
<rdf:RDF
xmlns:dc="http://purl.org/dc/elements/1.1/"
xmlns:xsd="http://www.w3.org/2001/XMLSchema#"
xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns:owl="http://www.w3.org/2002/07/owl#"
xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
xmlns:anotacaojuridica="http://www.owl-ontologies.com/anotacaojuridica.owl#"
xmlns:test="file:/C:/MINDlab/test.owl#"
>
```

Figura 23: Namespaces

A Figura 23 representa as declarações dos *namespaces* do documento que foi gerado. Os *namespaces* são declarados pelo atributo *xmlns*. Pode-se observar as definições primitivas de OWL, XMLS, RDF e RDFS. As duas últimas linhas identificam os *namespaces* associados à ontologia e ao documento anotado respectivamente.

Depois que os *namespaces* foram definidos, a ferramenta inclui uma coleção de afirmações sobre a ontologia agrupadas em *tags* `<owl:Ontology>`, conforme ilustrado na Figura 22:

```
<owl:Ontology rdf:about="file:/C:/MINDlab/test.owl">
<owl:imports rdf:resource="http://www.owl-ontologies.com/anotacaojuridica.owl"/>
</owl:Ontology>
<owl:Class rdf:about="http://www.owl-ontologies.com/anotacaojuridica.owl#oj_Apelado">
</owl:Class>
<owl:Class rdf:about="http://www.owl-ontologies.com/anotacaojuridica.owl#oj_Apelante">
</owl:Class>
<owl:Class rdf:about="http://www.owl-ontologies.com/anotacaojuridica.owl#oj_Relator">
</owl:Class>
<owl:Class rdf:about="http://www.owl-ontologies.com/anotacaojuridica.owl#oj_Vitima">
</owl:Class>
```

Figura 24: Declaração de headers

As tags `<owl:Ontology>` referenciam os dados OWL da ontologia empregada no processo de anotação. Estas tags também podem ser usadas para incluir comentários, controle de versões e inclusão de novas ontologias para a anotação [SMITH et al, 2004].

Por fim, são definidas as relações entre os termos do documento com as classes da ontologia, ambos definidos nos *namespaces* do documento RDF anteriormente na Figura 23.

A tag `<rdf:Description>` presente na Figura 24, determina que a descrição que se segue é do fragmento do texto presente no documento que será relacionado com a ontologia. A propriedade `rdf:about` dentro da tag `<rdf:Description>` determina qual é o fragmento do texto. As classes da ontologia as quais os termos declarados serão associados são declaradas pelo elemento `<owl:class>`. O nome das classes são definidos pelo atributo `rdf:about`, onde serve como identificador do recurso. O tipo de relação entre o termo do texto e a classe da ontologia é definido pela tag interna a tag `<rdf:Description>`, `<rdf:type>`. Segue abaixo o trecho final do código gerado:

```
<rdf:Description rdf:about="file:/C:/MINDlab/test.owl#A_JUSTICA_POR_SEU_PROMOTOR">
  <rdf:type>
    <owl:Class rdf:about="http://www.owl-ontologies.com/anotacaojuridica.owl#oj_Apelado">
      </owl:Class>
    </rdf:type>
  </rdf:Description>
<rdf:Description rdf:about="file:/C:/MINDlab/test.owl#Adilson_Rocha">
  <rdf:type>
    <owl:Class rdf:about="http://www.owl-ontologies.com/anotacaojuridica.owl#oj_Vitima">
      </owl:Class>
    </rdf:type>
  </rdf:Description>
<rdf:Description rdf:about="file:/C:/MINDlab/test.owl#Genésio_Nolli">
  <rdf:type>
    <owl:Class rdf:about="http://www.owl-ontologies.com/anotacaojuridica.owl#oj_Relator">
      </owl:Class>
    </rdf:type>
  </rdf:Description>
<rdf:Description rdf:about="file:/C:/MINDlab/test.owl#Guaraci_Edson_Fagundes">
  <rdf:type>
    <owl:Class rdf:about="http://www.owl-ontologies.com/anotacaojuridica.owl#oj_Vitima">
      </owl:Class>
    </rdf:type>
  </rdf:Description>
<rdf:Description rdf:about="file:/C:/MINDlab/test.owl#JOÃO_MARCOS_COIMBRA">
  <rdf:type>
    <owl:Class rdf:about="http://www.owl-ontologies.com/anotacaojuridica.owl#oj_Apelante">
      </owl:Class>
    </rdf:type>
  </rdf:Description>
<rdf:Description rdf:about="file:/C:/MINDlab/test.owl#Nilton_Macedo_Machado">
  <rdf:type>
    <owl:Class rdf:about="http://www.owl-ontologies.com/anotacaojuridica.owl#oj_Relator">
      </owl:Class>
    </rdf:type>
  </rdf:Description>
</rdf:RDF>
```

Figura 25: Associação dos termos do texto com as Classes da ontologia

4.4 – Considerações finais

O processo de anotação semântica proposto no Estudo de Caso foi bem sucedido. As anotações realizadas no documento são de grande importância, pois poderão ser utilizadas para melhorar o resultado das buscas sobre estes documentos, que apresentam atualmente um baixo índice de precisão e revocação. Desta maneira, as buscas poderão ser mais precisas e direcionadas, fazendo com que o usuário não necessite mais realizar demasiado esforço para fazer a filtragem dos resultados obtidos pela ferramenta de busca.

Atualmente, procurando pela palavra chave “relator”, por exemplo, o sistema retorna apenas os acórdãos onde a palavra relator está escrita, exatamente como foi digitada na busca. Efetuando uma busca semântica com as anotações realizadas no mesmo documento, digitando a mesma palavra, “relator”, o sistema retornará os documentos que possuem alguma anotação que aponte para a classe *ojRelator* da ontologia. O resultado da busca poderá ser a palavra “relator” ou o nome da pessoa que relatou o caso, tornando a busca muito mais rica.

Procurando pelo nome “João”, por exemplo, de acordo com as anotações realizadas sobre o documento, será possível especificar se o nome procurado é um acusado, relator, pai, réu, etc. Tornando a busca muito mais precisa.

Uma melhoria no atual sistema de buscas do Tribunal de Justiça de Santa Catarina poderia ser a anotação progressiva dos acórdãos existentes e a padronização de que os novos acórdãos gerados sejam obrigatoriamente anotados de acordo com uma ontologia pré-definida.

No site, uma alternativa seria manter o atual sistema de buscas e acrescentar uma nova funcionalidade, a busca semântica por acórdãos, que funcionaria baseado na mesma ontologia adotada para realizar as anotações nos documentos.

Uma nova interface teria que ser criada e adaptada para melhor atender aos requisitos das buscas semânticas. O objetivo da nova interface seria tornar o novo processo de busca em um procedimento intuitivo e de fácil assimilação aos atuais e novos usuários do sistema.

5 – Conclusões e Trabalhos Futuros

Desde a criação da Web, sua estrutura básica sofreu poucas alterações e a Web semântica propõe a mudança necessária para que a Web se torne realmente uma fonte de informação confiável, onde buscas por informações sejam mais fáceis e retornem um conjunto de resultados mais abrangente e preciso. A complexidade envolvida nas diversas tecnologias relacionadas com a Web semântica ainda é um grande empecilho, mas poderá ser contornada pela contínua melhoria das ferramentas disponíveis para facilitar a sua utilização.

A Web Semântica é fundada sobre o princípio de que o conteúdo on-line será anotado semanticamente de acordo com um conjunto de ontologias, o que tornará o conteúdo da Web processável por máquinas. As anotações semânticas consistem na associação de metadados rigorosamente definidos de acordo com uma ontologia.

Este trabalho apresenta uma pesquisa sobre a Web semântica, dando ênfase ao processo de anotação semântica de documentos e sua aplicação. A anotação semântica automática e semi-automática é uma alternativa para contornar a atual ausência de semântica nos documentos, com potencial de contribuir para a melhoria da precisão e da revocação da recuperação de informações em relação às ferramentas de buscas atuais.

A anotação semântica manual não é uma boa alternativa para grandes volumes de dados, pois sempre surgem novos documentos e novos termos, tornando o processo de anotação interminável. Outro ponto negativo é que a anotação manual exige grande conhecimento do domínio e da ontologia utilizada por parte do usuário que efetua as anotações. Ainda assim é uma boa alternativa para desenvolvedores de sistemas Web que queiram ter seus documentos com algum tipo de anotação semântica relevante.

Ferramentas de anotação automática são as mais indicadas para grandes volumes de documentos. Contudo geralmente estas ferramentas ainda estão sujeitas a falhas, ocasionando anotações errôneas e indesejadas.

As ferramentas de anotação semi-automática ou híbridas são uma boa alternativa quando não é possível realizar o processo de anotação de maneira totalmente automática. Isso infelizmente acontece em muitas aplicações práticas, requerendo interações com o usuário no processo de anotação, seja para capturar sua forma de trabalho, seja para validar um conjunto de anotações candidatas geradas automaticamente.

5.1 – Contribuições

As principais contribuições deste trabalho são:

1. Revisão bibliográfica de conceitos, técnicas e ferramentas relativos à anotação semântica de documentos;
2. Estudo comparativo de ferramentas para a anotação semântica automática e semi-automática de documentos;
3. Anotação semântica de documentos para o domínio jurídico utilizando o esboço de uma ontologia para este domínio e ementas de processos jurídicos fornecidos;
4. Avaliação da ferramenta de anotação semântica Smore no estudo de caso para o domínio jurídico.

A análise das ferramentas e das anotações efetuadas permitiu observar que o seu emprego pode ser efetivado em ambientes corporativos, por exemplo. É possível efetuar anotações semânticas necessárias para que a realização de buscas semânticas em documentos se torne realidade, melhorando a precisão na recuperação das informações contidas nos documentos.

Uma observação quanto à documentação das ferramentas pesquisadas é que a maioria das ferramentas é carente de documentação adequada, o que tornou o trabalho de instalação e testes das ferramentas uma tarefa desgastante e desestimulante. Outra observação a ser feita está relacionada à falta de atualização destas documentações.

5.2 – Trabalhos futuros

O trabalho apresentado traçou um panorama geral da Web semântica e seus conceitos e restringiu-se ao problema das anotações semânticas de documentos. Não se preocupou com a criação de ontologias, ou de instâncias. A anotação semântica realizada também ficou limitada ao documento, por se tratar de anotações intrusivas.

Como sugestões ficam a procura e análise de outras ferramentas para a anotação semântica de documentos e a evolução da ontologia Ontojuris, com a criação de novas classes e instâncias.

Outra sugestão é a realização de testes com ferramentas de anotação automáticas e a documentação de todo o processo realizado, já que para algumas ferramentas a documentação existente é deficiente.

Também fica em aberto a implementação de uma ferramenta de busca para os documentos anotados com a ontologia Ontojuris. Uma boa interface para a ferramenta é fundamental para que os usuários consigam realizar buscas mais precisas e avaliar adequadamente os resultados retornados.

6 – Referências

AGUIAR, André Vieira. **Mitos e Verdades sobre a Web Semântica**. Universidade Federal de Santa Catarina - Curso de Sistemas de Informação. 2007.

ALVARENGA, Lídia; SOUZA, Renato Rocha. **A Web Semântica e suas Contribuições para a Ciência da Informação**. Universidade Federal de Minas Gerais – Brasília. 2004.

ANTONIOU, G. & HARMELEN, F. V. **A Semantic Web Primer**. Massachusetts Institute of Technology, 2004.

BAX, M. P. **Introdução às linguagens de marcas**. *Ciência da Informação*, Brasília, v. 30, n.1, p. 32-38. Jan/abr. 2001.

BAYERI, Petra S.; GUT, Ulrike; LÜNGEN, Herald. **Methodology for Reliable Schema Development and Evaluation of Manual Annotations**. In: *The Second International Conference on Knowledge Capture (KCAP 2003)*. Florida. 2003.

BECHHOVER, S., HARMELEN, F., HENDLER, J. HORROCKS, I., MCGUINNESS, D., SCHNEIDER, P., STEIN, L., **OWL Web Ontology Language Reference**. *W3C Recommendation*. 2004.

BENJAMINS V. R., CASANOVAS P., BREUKER J., GANGAMI A. **Law and the Semantic Web, an Introduction**. In *Law and the Semantic Web*. Springer Verlag, Berlin. 2005

BERNERS-LEE, Tim. **Metadata architecture: documents, metadata and link**. 1997.

BERNERS-LEE, T., et al., **Semantic Web Development Proposal**, 2001. Disponível em: <<http://www.w3c.org/2001/sw/>>.

BERNERS-LEE, Tim; HENDLER, J.; LASSILA, O. **The semantic web: a new form of web content that is meaningful to computers will unleash a revolution of new possibilities**. *Scientific American*, New York, may 2001.

BAEZA-YATES, R., RIBEIRO-NETO, B.; **Modern Information Retrieval**. Addison Wesley. New York 1999.

BIWAS, G., BEZDEK, J., MARQUES, M. & SUBRAMANIAN, V. **“Knowledge-Assisted Document Retrieval: II. The Retrieval Process”**, *Journal of the American Society for Information Science (JASIS)*, Vol. 38, No 2, 1987.

BOOKESTEIN, A. **“Fuzzy Requests: An Approach to Weighted Boolean Searches”**, *Journal of the American Society for Information Science (JASIS)*, Vol. 31, No 7, 1980.

BRAY, Tim, et. Al. **Extensible Markup Language (XML) 1.0 (Third Edition)**. *W3C Recommendation*. 2004

BROEKSTRA, J.; KAMPMAN, A.; HARMELEN, F. van. **Sesame: A Generic Architecture for Storing and Querying RDF and RDF Schema.** *The Semantic Web-ISWC*. Springer, v. 2342, p. 54–68, 2002.

CARDOSO, Olinda N. P.; **Recuperação de Informação**, Departamento de Ciência da Computação - Universidade Federal de Lavras, 2003.

CASTRO, E., **XML para World Wide Web**. Rio de Janeiro. Editora Campus. 2001.

CHANDRASEKARAN, B., JOSEPHSON, J.R., BENJAMINS, V.R. **What Are Ontologies, and Why Do We Need Them?** *IEEE Intelligent Systems*. p. 20-26. 1999.

DACONTA, M.C., OBRST, L.J., SMITH, K.T. **The Semantic Web.** *Wiley Publishing*. Indianópolis, 2003.

DAVIES, John; FENSEL, Dieter; HARMELEN, Frank van. **Towards The Semantic Web: Ontology-driven Knowledge Management.** England: John Wiley & Sons Ltd, 2003.

DAVIES, J., STUDER, R., WARREN, Paul. **Semantic Web technologies : trends and research in ontology-based systems.** *E.d Wiley*. London, England, 2006.

DEITEL, H. M; DEITEL, P. J; NIETO, T. R. et al. **XML: How to program.** Porto Alegre: Bookman, 2003.

DINGLI, Alexiei; CIRAVEGNA, Fabio; WILKS, Yorick. **Automatic Semantic Annotation using Unsupervised Information Extraction and Integration.** *In: Proceedings of The Second International Conference on Knowledge Capture (KCAP 2003)*. Florida. 2003.

FENSEL, D. **Ontologies: A Silver Bullet for Knowledge Management and Electronic Commerce.** Springer-Verlag. Berlin. 2001.

FONSECA, F. & FIDALGO, R. **Gerenciamento de Dados e Informação**, Universidade Federal de Pernambuco, 2006.

FRAKES, W. B. & BAEZA-YATES, R. **Information Retrieval Data Structures & Algorithms**, Prentice Hall, 1992.

GLONVEZYNSKI, R. A. & GAUTHIER, F. O., **Modelo de Anotação de Documentos Para a Codificação do Conteúdo Semântico no Processo de Autoria.** Universidade Federal de Santa Catarina – Programa de Pós-Graduação em Ciência da Computação, 2005.

GOMES-PEREZ, A. **Tutorial on Ontological Engineering**, *International Joint Conference on Artificial Intelligence, IJCAI*, 1999.

GRUBER, T. R. **Toward principles for the design of ontologies used for knowledge sharing.** *International Journal of Human-Computer Studies*, v. 43, 1995.

GUHA, R.; MCCOOL, Rob; MILLER, Eric. **Semantic Search**. Proceedings of the twelfth international conference on World Wide Wibe. Budapest, Hungary. Páginas 20-24. 2003.

HANDSCHUH, S., STAAB, S. and CIRAVOGNA, F., **S-CREAM-Semi-automatic CREATION of Metadata in SAAKM**. *Semantic Authoring, Annotation & Knowledge Markup - Preliminary Workshop Programme 2002*.

HECK, Rachel M., LUEBKE, Sarah M., and OERMARK, Chad H.. **A survey of web annotation systems**. Technical report, Grinnell College, Grinnell, Iowa, 1999.

HORROCKS, I., FENSEL, D., BROEKSTRA, J., DECKER, S., ERDMANN, M., GOBLE, C., HARMELEN, F., KLEIN, M., STAAB, S., STUDER, R., MOTTA, E. 2000. **The Ontology Inference Layer OIL**. University of Amsterdam, 2000. (Technical report).

KAHAN, Jose; KOIVUNEN, Marja-Riitta; PRUD'HOMMEAUX, Eric; and SWICKD, Ralph R. **Annotea: An Open RDF Infrastructure for Shared Web Annotations**. In *Proceedings of the Tenth Inter-national World Wide Web Conference*, Hong Kong, China, May, 2001. pp. 623-632.

KALYANPUR, A., PARSIA, B., HENDLER, J., GOLDBECK, J. **Smore – Semantic Markup, Ontology, and RDF Editor**. *Proceedings of 3rd International Semantic Web Conference (ISWC-2004)*, Japan, 2004.

KARP, P., CHAUDRI, V., THOMERE, J. 1999. **XOL: an XML-based ontology exchange language**. Version 0.4. www.ai.sri.com/~pkarp/xol.

KIETZ, J.-U. and VOLZ, R., **Extracting a Domain-Specific Ontology from a Corporate Intranet**. In *Fourth Conference on Computational Natural Language Learning and of the Second Learning Language in Logic Workshop*, Lisbon, Portugal, 2000. pp. 167-175.

KIRYAKOV, Atanas. POPOV, Borislav. OGNJANOFF, Damyan. **Semantic Annotation, Indexing, and Retrieval**. *2nd International Semantic Web Conference (ISWC2003)*, Florida, USA. Páginas 484 a 499, Outubro, 2003.

KLYNE, Graham; CARROL, Jeremy; MCBRIDE, Brian. **Resource Description Framework (RDF): Concepts and AbstractSyntax**. *W3C Recommendation 10 February 2004*. Disponível em <http://www.w3.org/TR/rdf-concepts/>.

KOGUT, Paul; HOLMES, William. **AeroDAML: Applying Information Extraction to Generate DAML Annotagions from Web Pages**. Philadelphia. 2001. <http://semannot2001.aifb.uni-karlsruhe.de/positionpapers/AeroDAML3.pdf>

KOIVUNEN, M., MILLER, E. 2001. *W3C Semantic Web Activity*. <http://www.w3.org/2001/12/semweb-fin/w3csw>

KWOK, K. L. “**A Network Approach to Probabilistic Information Retrieval**”, *ACM Transactions on Information Systems*, Vol. 13, No 3, 1995.

REEVE, L., HAN, H. **Survey of semantic annotation platforms**. ACM Symposium on Applied Computing (SAC), 2005.

LEE, Tim-Berners. **Uniform Resource Identifiers (URI): Generic Syntax**. *Internet Engineering Task Force Request for Comments RFC1630*, disponível em <http://www.ietf.org/rfc/rfc1680.txt?number=1680>.

LUKE, Sean; SPECTOR, Lee; RAGER, David. 1996. **Ontology-Based Knowledge Discovery on the World-Wide Web**. *Proceedings of the Workshop on Internet-based Information Systems*, AAAI-96 (Portland, Oregon). USA.

MARTIN, D., et al. **Professional XML**. Editora Ciência Moderna Ltda. Rio de Janeiro. 2001.

MENDONÇA, Eduardo. **Extração Resiliente de Dados RDF a partir de Fontes Dinâmicas em Linguagem de Marcação**. Dissertação de Mestrado – Universidade Federal do Ceará – Departamento de Computação. 2004

MCGUINNESS, Debora L.; FIKES, Richard; HENDLER, James et al. **DAML+OIL: an ontology language for the Semantic Web**. IEEE – Intelligent Systems. Volume 17. Páginas 72-80. 2002.

MCGUINNESS, Debora L.; VAN HARMELEN, Frank. **OWL Web Ontology Language Overview**. W3C Recommendation 10 February 2004. Disponível em <http://www.w3.org/TR/owl-features/>.

MOTTA, E., VARGAS-VERA, M., DOMINGUE, J., LANZONI, M., STUTT, A., CIRAVEGNA, F. **MnM: Ontology driven semi-automatic and automatic support for semantic markup**. In *13th International Conference on Knowledge Engineering and Knowledge Management (EKAW02)*, Siguenza, Spain, pp 379–391. 2002.

MOURA, Ana Maria de C. **A Web Semântica: Fundamentos e Tecnologias**. Congresso Internacional de Ciências de la Computación – CICC 2001. Universidad de Aquino. Aquino, Bolívia. 2001.

NOVELLO, Taisa Carla. **Ontologias, Sistemas baseados em Conhecimento e Modelos de Banco de Dados**. 2003.

NUNES, Ancelmo M., FILETO, Renato. **Uma Arquitetura para Recuperação de Informação Baseada em Semântica e sua Aplicação no Apoio a Jurisprudência**. INE – Universidade Federal de Santa Catarina - 2006

OLIVEIRA, D. H. de. **Introdução ao XML e suas aplicações**. 2002. Disponível em: http://www.xml.com.br/docs/intro_xml_apli.pdf.

POPOV, B. et al. **KIM: Semantic annotation platform**. In: *The Second International Semantic Web Conference (ISWC2003)*. Flórida: 2003. http://www.ontotext.com/publications/KIM_SAP_ISWC168.pdf

POPOV, B., KIRYAKOV, A., KIRILOV, A., MANOV, D., OGNJANOFF, D., GORANOV, M. **KIM – Semantic Annotation Platform**. *Lecture Notes in Computer Science* - Springer Berlin / Heidelberg, Volume 2870, pp 834-849. 2003.

REEVE, L., HAN, H. **Survey of Semantic Annotation Plataforms**. In: *The 2005 ACM Symposium on Applied Computing*. 2005

SMITH, M., WELTY, C., MCGUINNESS, D. **OWL Web Ontology Language Guide**. *W3C Recommendation*. February 10, 2004

SOWA, J. F. **Building, sharing and merging ontologies**. Tutorial. [S. 1. : s. n.], 1999.

UREN, V., CIMIANO P., IRIA, J., HANDSCHUH, S., MOTTA, E., CIRAVEGNA, F., VARGAS-VERA, . **Semantic annotation for knowledge management: Requirements and a survey of the state of the art**. *Journal of Web Semantics*, 4(1), pp. 14-28, Springer, 2005.

YESILADA, Yeliz. **Annotation and Transformation of Web Pages to Improve Mobility for Visually Impaired Users**. *University of Manchester*. Computer Science School. Doctor Degree Thesis. 2005. Disponível em: <http://www.cs.man.ac.uk/~yesilady/publication/>

Anexo 1: Exemplo de acórdão

Apelação criminal n. 99.001792-3, de São José

Relator: Des. Genésio Nollí

FURTOS. EMBORA TENHAM SIDO COMPROVADAS AUTORIA E MATERIALIDADE PELOS DEPOIMENTOS DAS VÍTIMAS, EM CONSONÂNCIA COM O RESTANTE DA PROVA, AO RÉU FOI APLICADA A MEDIDA DE SEGURANÇA DE INTERNAÇÃO, PORQUANTO O LAUDO DO EXAME DE SANIDADE MENTAL RECONHECEU SER ELE INTEIRAMENTE IRRESPONSÁVEL PELOS ATOS PRATICADOS. DECISÃO CORRETA RECURSO DESPROVIDO.

"Absolvido o agente inimputável, autor de fato punível com pena de reclusão, é obrigatória a imposição de medida de segurança de internação em casa de custódia e tratamento psiquiátrico" (Recurso criminal n. 98.000088-2, de Lages - Relator: Des. Nilton Macedo Machado).

Vistos, relatados e discutidos estes autos de Apelação Criminal n. 99.001792-3, da Comarca de São José, em que é apelante JOÃO MARCOS COIMBRA, e apelado(a) A JUSTICA POR SEU PROMOTOR:

ACORDAM, em Primeira Câmara Criminal, por votação unânime, negar provimento ao recurso.

Custas legais.

1. Perante o Juízo de Direito da comarca de São José (Vara Criminal), João Marcos Coimbra foi denunciado como incurso nas sanções do artigo 155, *caput*, c/c o § 1º, em relação à vítima Guaraci Edson Fagundes, e artigo 155, *caput*, e art. 155, *caput*, c/c o artigo 14, II, em relação à vítima Adilson Rocha, tudo na forma do art. 71, todos do Código Penal.

Narrou a denúncia:

"...o acusado, com o objetivo de se apossar do patrimônio alheio, em data de 13 de abril do corrente ano, por volta das 03:00 horas da madrugada, o acusado invadiu, clandestinamente, o interior da residência da vítima Guaraci Edson Fagundes, sito a Rua Manoel M. Silva, n. 498, Praia Comprida, nesta cidade, aproveitando-se do repouso noturno da vítima e sua esposa. Do interior da residência de Guaraci, o acusado subtraiu uma calça de cor branca, pertencente à vítima, ocultando-a no pátio da casa. Surpreendida pela vítima, que acordado por sua esposa, levantou-se, dirigiu-se a sala de casa, onde viu o acusado fugir pela janela. Notou, ainda, que o acusado portava uma faca. Imediatamente a vítima acionou a polícia.

"O acusado, insatisfeito, após deixar a casa de Guaraci e trocar de roupa, dirigiu-se até a residência de Adilson Rocha, sito à Rua Oscar Teodoro da Silva, n. 578, próximo ao Bebidas Elimos, Praia Comprida, nesta cidade, ingressou no interior dela, de onde apossou-se das chaves do veículo Voyage de Adilson, de um rádio toca-fita tipo gaveta que usava no carro e de um vídeo-cassete; ato contínuo, o acusado ingressou no carro de Adilson, estacionado no pátio da casa, e quando tentava acionar o motor do veículo, objetivando subtraí-lo, foi surpreendido pela vítima, empreendendo fuga, deixando no carro as chaves e o aparelho de som. Na residência vizinha da vítima Guaraci foi encontrado o aparelho vídeo cassete, já adremente separado pelo réu.

"A Autoridade Policial que fora acionada por Guaraci (a primeira vítima), quando chegou no local, encontrou o acusado imobilizado por Guaraci e seu cunhado, dando-lhe voz de prisão, após ouvir o relato das vítimas.

"O acusado subtraiu para si uma calça jeans pertencente à vítima Guaraci Edson Fagundes e um aparelho de vídeo-cassete pertencente a Adilson Rocha, ocultando tais objetos no pátio da residência da primeira vítima, em duas condutas autônomas e distintas. As vítimas somente encontraram e recuperaram tais pertences após a prisão do réu. Os referidos objetos, portanto, integraram a posse, mesmo que efêmera, do acusado, razão pela qual, estes dois delitos restaram devidamente consumados.

"Por outro lado, o acusado não logrou apossar-se do veículo de propriedade da vítima Adilson Rocha, posto que surpreendido por este, deixando, portanto, de consumir tal delito, por circunstância alheia à sua vontade".

Processado, ao final restou absolvido com fulcro no artigo 386, V, do CPP, por ausência de imputabilidade e, na forma do art. 26, caput, do CP, foi-lhe aplicada a medida de segurança de internação em Hospital de Custódia e Tratamento Psiquiátrico, pelo prazo mínimo de 03 (três) anos.

Inconformado recorreu do decreto absolutório impróprio, pleiteando a absolvição fundada no art. 386, VI, do CPP, pela absoluta ausência de provas do delito, revogando-se a medida de segurança aplicada.

Rebatido o inconformismo, rumaram os autos a este segundo grau de jurisdição, onde a Procuradoria de Justiça opinou pelo seu improvimento.

É o relatório.

2. A decisão vergastada não merece reparos.

Ao contrário do que sustenta a insígne defesa, os delitos restaram efetivamente comprovados.

Embora tenha o apelante negado a autoria em ambas as fases, os depoimentos das vítimas, apoiados nos depoimentos dos policiais, autorizam a formação de um juízo de convicção, senão vejamos:

A materialidade restou evidenciada pelos termos de apreensão e de reconhecimento e entrega de fls. 20/23 e 26, bem como pelo laudo pericial de fls. 48.

Quanto à autoria, asseverou a vítima Guaraci:

"...na data dos fatos, por volta das duas e quarenta horas, a esposa do depoente observou uma pessoa dentro da residência, acordando-o para que averiguasse. Quando o depoente chegou no local em que a esposa havia visto dita pessoa, deparou-se com um indivíduo vestindo somente uma bermuda, já do lado de fora, próximo a uma janela, pela qual o depoente perguntou o que estava se passando. Este indivíduo então lhe disse que encontrava-se armado. Com receio o depoente foi ao telefone comunicar o fato ao COPOM e a um cunhado seu ...; Neste meio tempo este indivíduo saiu do local; o depoente saiu fazendo uma busca no pátio de sua residência...; Observou então que uma pessoa passava em frente a sua casa vestindo calça, camisa e tênis, tentando esconder seu rosto; como o depoente é bom fisionomista reconheceu que se tratava daquele mesmo indivíduo que tinha visto momentos antes; (...); ao retornar para sua residência novamente passou a procurar objetos nas imediações, encontrando um videocassete..." (fls. 68/69).

A vítima Adilson Rocha confirmou que teve furtado seu video-cassete, encontrando posteriormente no pátio da residência da outra vítima, uma rua após. Confirmou ainda, que presenciou um indivíduo, que pelo conjunto de circunstâncias faz crer ser o réu, tentando furtar seu carro.

Os policiais quando chegaram, ao passo de já ter sido preso o réu, apreenderam com ele uma faca igual à descrita pela vítima Guaraci, além de corroborarem sua versão.

Outrossim, a alegação do apelante de que estava passando pela localidade, pois almejava dirigir-se ao Hospital Colônia Santana para adquirir remédio, e foi abordado pelas vítimas que o teriam prendido, parece pouco crível, além de não restar comprovada em qualquer momento destes autos.

Comprovadas autoria e materialidade, correta foi a decisão quando aplicou a medida de segurança ao apelante, determinando sua internação, eis que conforme a conclusão do exame de sanidade mental realizado (fls. 80), "é portador de um Transtorno Esquizotípico. Do ponto de vista psiquiátrico, deve ser considerado como sendo irresponsável pelos atos praticados e que constam nos Autos de Processo-Crime".

Bem se houve também a Magistrada sentenciante ao determinar a aplicação da medida de segurança de internação.

Em resposta a quesito do MM Juiz, os peritos responderam que a espécie adequada de tratamento ao apelante é, inicialmente, a internação (fls. 80).

Se já não bastasse este argumento, cumpre ressaltar que o apelante praticou delitos punidos com pena de reclusão, e "**Ao inimputável autor de fato punível com pena de reclusão, a medida de segurança cabível é o internamento e não o tratamento ambulatorial**" (TJSP, RT 612/317).

Assim, diante do exposto, nega-se provimento ao recurso.

Presidiu o julgamento o Exmo. Sr. Des. Genésio Nolli, e participaram do mesmo, com votos vencedores, os Exmos. Srs. Des. Francisco Borges e Souza Varella, e lavrou o parecer pela douta Procuradoria-Geral de Justiça o Exmo. Sr. Dr. Demétrio Constantino Serratine.

Florianópolis, 31 de agosto de 1999.

Amaral e Silva

PRESIDENTE P/ O **ACÓRDÃO**

Genésio Nolli

RELATOR

Anexo 2: Padrão do acórdão

[Classe do processo]
n. [Número do processo], de [Foro de origem]
Relator: Des. [Relator atual do processo sem tratamento]

EMENTA

*

Vistos, relatados e discutidos estes autos de [Classe do processo] n. [Número do processo], da comarca de [Foro de origem] ([Vara de origem]), em que é/são [participação da principal parte ativa] [Principal parte ativa], e [participação da principal parte passiva] [Principal parte passiva]:

ACORDAM, em [Órgão julgador atual do processo],
[Decisão do processo]

RELATÓRIO

*

VOTO

*

DECISÃO

Ante o exposto, *.

Participou/Participaram do julgamento, [Participantes].

Florianópolis, [Data da sessão com mês por extenso].

[Presidente do órgão julgador atual do processo]

PRESIDENTE

[Relator atual do processo sem tratamento]

RELATOR

Apêndice 1 – ARTIGO

Anotações Semânticas de Fontes de Dados Heterogêneas Um Estudo de Caso com a Ferramenta Smore

Markus Pereira Eller

Departamento de Informática e Estatística - Universidade Federal de Santa Catarina
Campus Trindade, Florianópolis – SC, Brasil
markus@inf.ufsc.br

Resumo. *O artigo tem por objetivo apresentar as etapas realizadas para fazer a anotação semântica semi-automática de documentos com a ferramenta Smore. A justificativa é que a tarefa de recuperação de informação vem se tornando cada vez mais trabalhosa devido à quantidade de documentos existentes e à falta de padronização para descrever e recuperar tais documentos. As anotações realizadas serão utilizadas posteriormente para efetuar a recuperação de informação de grandes volumes de documentos. Para tanto, foram utilizados os conceitos: de Web-semântica, de ontologias, de recuperação de informação e de anotação semântica.*

1. Introdução

Com o objetivo de estudar os conceitos sobre Web-semântica e conhecer mais profundamente os seus componentes, foram realizadas pesquisas e estudos nessa área, servindo de motivação para a elaboração deste artigo. Mais especificamente, será realizado um estudo sobre a Web-semântica e de ferramentas para fazer a anotação semântica de documentos. Uma das ferramentas pesquisadas será escolhida e será utilizada como um estudo de caso, onde documentos da área jurídica serão efetivamente anotados de acordo com uma ontologia pré-definida.

2. Desenvolvimento

A seguir apresenta-se um descritivo das atividades realizadas durante o desenvolvimento do trabalho, onde se realizou o Estudo de Caso com a anotação semântica de documentos. Os resultados do trabalho fundamentam-se após resultados dos estudos da revisão da literatura e as necessidades do Tribunal de Justiça de Santa Catarina, buscando prover buscas mais precisas aos usuários. Os trabalhos foram desenvolvidos nas seguintes etapas:

2.1. Fundamentação Teórica

Foram realizadas pesquisas bibliográficas em livros, artigos, monografias, apostilas, tutoriais e sites da internet. Estes estudos estão presentes em todas as etapas do trabalho, onde o tempo e o esforço disponibilizados para as pesquisas, em busca de informações sobre os conceitos e ferramentas, resultaram na consolidação dos conceitos referentes à

Web-semântica, como por exemplo: Recuperação de Informação, Web Semântica e Ontologias.

2.1.1. Recuperação de Informação

Recuperação da informação é uma subárea da Ciência da Computação que estuda o processo de armazenamento e recuperação automática de documentos, que são objetos de dados, geralmente textos [1].

O principal objetivo é encontrar, de forma eficiente, os documentos que satisfaçam a necessidade do usuário. A crescente complexidade dos objetos armazenados e o grande volume de dados exigem processos de recuperação cada vez mais sofisticados

2.1.2. Web Semântica

A Web semântica será uma extensão da Web atual; porém apresentará estrutura que possibilitará a compreensão e o gerenciamento dos conteúdos armazenados na Web independente da forma em que estes se apresentem (como texto, som, imagem ou gráfico) a partir da valoração e formalização das descrições semânticas desses conteúdos [2].

Na Web Semântica, os documentos são anotados com meta-informação, que define quais informações ele contém [3]. Esta meta-informação, acompanhada de alguma teoria de domínio, na forma de ontologias, por exemplo, possibilitará uma Web que fornecerá um novo nível de serviços.

2.1.3. Ontologias

Uma ontologia é uma especificação explícita dos objetos, conceitos e outras entidades que assumimos existirem em uma área de interesse, além das relações entre esses conceitos e restrições expressados através de axiomas [4].

Esta é a camada mais importante e pesquisada da Web semântica. Nos últimos anos as ontologias vêm ganhando grande ênfase no campo da Ciência da Computação e Inteligência Artificial como meio de representar, compartilhar e reusar o conhecimento de forma legível para um computador.

2.2. Anotação Semântica

A anotação semântica de documentos possui como objetivo facilitar a busca dos documentos no repositório digital. Com ela, é possível correlacionar termos (conceitos, instâncias ou propriedades) da ontologia a palavras, simples ou compostos, do texto que passou pelo processo de anotação semântica. Ela atribui às palavras que aparecem no documento ligações com suas descrições semânticas na ontologia.

As anotações semânticas possuem papel fundamental no desenvolvimento da Web Semântica, seja no sentido de criar novos documentos já com conteúdo semântico definido ou para prover semântica a documentos já existentes. A estrutura das anotações semânticas deve seguir as recomendações do W3C, na elaboração de sua estrutura e na utilização do modelo RDF como a linguagem para descrever os recursos.

A Web semântica precisa que o máximo de conteúdo relevante esteja anotado semanticamente para que se possam desenvolver sistemas que utilizem ao máximo este recurso e proporcionem novos serviços, ou melhorem os já existentes, sobre os recursos disponíveis atualmente

2.2.1. Ferramentas para anotação semântica

Nesta etapa foram estudados conceitos referentes as ferramentas de anotação semântica e pesquisadas e analisadas algumas das ferramentas de anotação semântica existentes e que se encaixavam com o proposta do trabalho.

Existem pelo menos três tipos de ferramentas que podem ser utilizadas para anotação semântica de documentos [5]:

- Semi-automáticas: Associam palavras do texto a classes, instâncias e propriedades da ontologia, utilizando-se do julgamento humano. Esta associação geralmente é efetuada através de interfaces “arraste-e-solte”.
- Automática: Aplica técnicas de processamento de linguagem natural (PLN), aprendizado de máquina e extração de informação, entre outras, para associar palavras à ontologia.
- Híbrida: Utiliza as definições de anotação semântica semi-automática e automática para combiná-las em uma só ferramenta, ou seja, pode utilizar tanto o julgamento humano quanto técnicas de PLN para determinar as associações de palavras do texto com classes e propriedades.

As Plataformas de anotação semântica podem incluir um subconjunto de vários recursos, tais como APIs de acesso a ontologias, interfaces gráficas menos ou mais intuitivas aos usuários, editores para a base de conhecimento, armazenamento de recursos (por exemplo, repositórios RDF), etc. Nem sempre todas as características estarão presentes em todas as Plataformas.

As pessoas que estudam formas de anotação semântica, seja manual ou automática, buscam diferentes pontos de vista sobre o problema. No processo de anotação Manual a investigação incide mais sobre a representação da anotação, compartilhamento e armazenamento, bem como mecanismos amigáveis de interface com o usuário para ajudar e facilitar as pessoas a escreverem as suas notas. Por outro lado, a investigação sobre ferramentas automatizadas de anotação incide mais sobre as formas de criar anotações, de acordo com determinados domínios de ontologias.

Abaixo, são apresentadas as ferramentas pesquisadas:

Ont-O-Mat [6]: Ont-O-Mat adota uma técnica automática de extração de dados baseado no Amilcare, que é uma adaptação de IE (*Information Extraction*), um sistema designado para suportar anotações em documentos. Com o Ont - O - Mat. É possível acessar ontologias especificadas em mais de um tipo de marcação, como RDF e DAML + OIL. Mas cada ontologia somente pode ser acessada individualmente.

MnM [7] é muito semelhante ao Ont - O - Mat. Fornece suporte para anotações automatizadas ou semi-automáticas. De acordo com os autores, a diferença entre os dois sistemas, Ont - O - Mat e MnM, é a sua filosofia. Embora Ont - O - Mat adota a filosofia que as marcações devem ser incluídas como parte dos documentos, MnM armazena suas anotações tanto como marcações em documento como itens de uma base de conhecimento.

KIM [8]: é uma parte do projeto SWAN (*Semantic Web ANnotator*). KIM consiste em uma ontologia KIM, uma base do conhecimento KIM, um KIM Server (com uma API para acesso remoto ou incorporação), e *front-ends* que fornecem acesso total às funcionalidades do KIM Server ao usuário. Os autores optaram RDF (S) como a sua língua de representação para ontologias. O KIM básico contém uma entidade de descrição de informação para fins de anotação. Durante o processo anotação, KIM emprega uma técnica NLP para fazer a extração da informação, que se baseia no GATE (General Architecture of Text Engineering) para extrair, indexar e anotar dados instanciados. É possível utilizar várias ontologias no KIM, podendo ser instalado e configurado para atender um domínio específico. Possui um *plug-in* para o *Internet Explorer* que permite carregar a ontologia do servidor e efetuar anotação da página web carregado no navegador.

Smore [9]: é uma ferramenta que permite aos usuários anotarem os seus documentos em RDF usando ontologias, fazendo a associação com termos específicos do documento com os elementos da ontologia. O objetivo deste programa é proporcionar ao usuário um ambiente flexível, em que ele possa criar sua página da web sem esforços demasiados que envolvam a anotação dos documentos. O usuário também pode criar sua própria ontologia a partir do zero, aproveitando fragmentos de ontologias existentes. O formato de linguagem para representação de ontologias suportado pela ferramenta é o OWL, conforme indicado pelo W3C.

Annotea [10]: faz parte dos esforços da W3C para firmar a Web semântica. A ferramenta permite a criação de anotações com metadados, gerando assim um ambiente compartilhado de anotações. A ferramenta permite a criação de anotações com metadados, gerando assim um ambiente compartilhado de anotações. O servidor de anotação armazena as anotações realizadas em uma base de dados RDF e toda a comunicação entre a cliente e servidor utiliza métodos HTTP.

A Tabela 1 abaixo faz uma síntese das características relevantes sobre as ferramentas pesquisadas.

Nome da Ferramenta	Tipo de Ferramenta	Linguagem da Ontologia	Entrada de Ontologias	Representação das Anotações	Evolução de Ontologia	Formato do Documento	Tipo de Anotação
Ont-O-Mat	Semi-automática	DAML + OIL/RDF	Apenas uma por tarefa	Dentro de páginas da web	Sim	HTML	Intrusivo
MnM	Híbrida	DAML + OIL/RDF	Uma ou mais de uma por tarefa	Dentro de páginas da Web e em uma base de conhecimento	Sim	HTML, texto	Intrusivo
KIM	Automática	RDF/OWL	Ontologia do KIM (Proton)	Dentro de páginas da web	Sim	HTML	Não-intrusivo
Smore	Semi-automática	OWL	Uma ou mais de um por tarefa	Dentro de páginas da web	Sim	HTML, e-mail, imagens, texto	Intrusivo
Annotea	Semi-automática	RDF	Uma ou mais de um por tarefa	Servidores de anotações	Não	HTML	Não-intrusivo

Tabela 1: Comparação de ferramentas de anotação semântica

Todas as ferramentas trabalham com documentos HTML sendo que muitas delas funcionam com navegadores Web associados em sua própria interface. Para representação das Ontologias, as linguagens DAML + OIL são usadas pelas ferramentas Ont-O-Mat e MnM, enquanto a linguagem RDF é utilizada nas ferramentas Annotea e KIM. Apenas as ferramentas Smore e KIM utilizam a OWL para representar as ontologias, conforme sugerido pelo W3C.

Na maioria das ferramentas pesquisadas, as anotações são intrusivas, ou seja, são guardadas junto ao documento anotado. As anotações são guardadas separadas do documento apenas no KIM e Annotea. Somente KIM e MnM implementam anotação automática. Todas as ferramentas pesquisadas suportam a evolução da ontologia (ao identificar entidades nomeadas ainda não descritas), exceto a ferramenta Annotea.

3. Estudo de Caso

Depois de ter estudado conceitos sobre Web-semântica, sobre anotações semânticas e ter pesquisado algumas ferramentas que realizam a anotação semântica de documentos, a próxima etapa foi definir um domínio para efetuar anotações semânticas em documentos.

Buscando dar continuidade ao trabalho de [11] e analisando os trabalhos futuros do mesmo, o domínio selecionado foi a área jurídica. O principal objetivo é buscar atender algumas demandas do judiciário por ferramentas computacionais para gerenciar e facilitar a recuperação da informação produzida. A proposta é associar alguns metadados baseados em uma ontologia aos documentos jurídicos para que seja possível estipular e processar consultas mais precisamente.

Os documentos escolhidos para se realizar o processo de anotação semântica são acórdãos de processos do Tribunal de Justiça de Santa Catarina, conseguidos pelo site <http://tjsc6.tj.sc.gov.br/jurisprudencia/PesquisaAvancada.do>. Em Outubro de 2008, o sistema contava com aproximadamente 250.000 Acórdãos em sua base de dados.

A ontologia OntoJuris [11] é o princípio de uma ontologia jurídica que modela o conhecimento jurídico brasileiro para fins de recuperação de informação. Ela representa as principais características dos acórdãos, sentenças e outros tipos de decisões jurídicas.

A ferramenta escolhida para fazer a anotação semântica foi o Smore, que é uma ferramenta para anotação semântica semi-automática projetada para permitir aos usuários a marcação de documentos HTML em RDF utilizando Ontologias em OWL, conforme as indicações da W3C. Além disso, a ferramenta apresenta documentação e tutoriais que facilitaram o procedimento de instalação e operação da ferramenta.

A definição de novas instâncias sempre é realizada desta forma: os termos do documento carregado no *web-browser* do Smore que farão referência às classes da ontologia devem ser identificados e selecionados. Depois disso, segue a definição da tripla RDF que irá efetivamente realizar esse relacionamento e gerar o código da anotação semântica resultante.

Depois que a tripla RDF é definida, o código RDF da anotação é gerado. A anotação faz com que o significado dos termos selecionados no texto deixem de ser analisados apenas sintaticamente, expressamente como são escritos, e passem a ter também um significado semântico, de acordo com as definições das classes da ontologia as quais foram relacionados através do Smore. A figura abaixo representa o procedimento realizado para gerar a anotação.

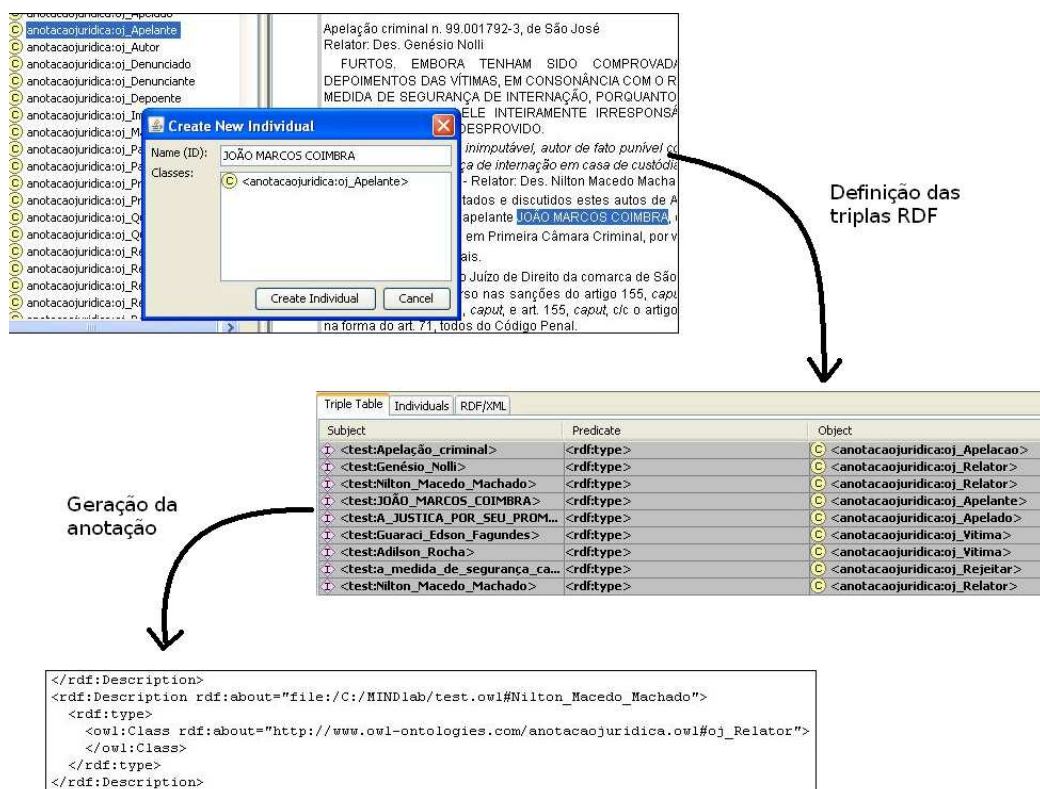


Figura 1: Geração das anotações em RDF na ferramenta Smore

4. Resultados

A realização do Estudo de Caso gerou esforços para: Estudar e entender o funcionamento de ferramentas para anotação semântica disponíveis; Revisão da literatura para embasamento da escolha de domínio e ferramenta; Revisão da literatura sobre os conceitos voltados a Web Semântica e anotação semântica de documentos.

Os benefícios incluem: definição de uma ontologia para a área jurídica, documentação de um modelo para a realização de buscas mais precisas e direcionadas, fazendo com que o usuário não necessite mais realizar demasiado esforço para fazer a filtragem dos resultados obtidos pela ferramenta de busca; Estudo comparativo de ferramentas para a anotação semântica automática e semi-automática de documentos; Avaliação da ferramenta de anotação semântica Smore no estudo de caso para o domínio jurídico.

5. Conclusões

Desde a criação da Web, sua estrutura básica sofreu poucas alterações e a Web semântica propõe a mudança necessária para que a Web se torne realmente uma fonte de informação confiável, onde buscas por informações sejam mais fáceis e retornem um conjunto de resultados mais abrangente e preciso.

Este trabalho apresenta uma pesquisa sobre a Web semântica, dando ênfase ao processo de anotação semântica de documentos e sua aplicação. A anotação semântica automática e semi-automática é uma alternativa para contornar a atual ausência de semântica nos documentos, com potencial de contribuir para a melhoria da precisão e da revocação da recuperação de informações em relação às ferramentas de buscas atuais.

Foi traçado um panorama geral da Web semântica e seus conceitos, porém o trabalho restringiu-se ao problema das anotações semânticas de documentos. Não se preocupou com a criação de ontologias, ou de instâncias. A anotação semântica realizada também ficou limitada ao documento, por se tratar de anotações intrusivas.

Como sugestões de trabalhos futuros ficam a procura e análise de outras ferramentas para a anotação semântica de documentos e a evolução da ontologia Ontojuris, com a criação de novas classes e instâncias.

Outra sugestão é a realização de testes com ferramentas de anotação automáticas e a documentação de todo o processo realizado, já que para algumas ferramentas a documentação existente é deficiente.

Também fica em aberto a implementação de uma ferramenta de busca para os documentos anotados com a ontologia Ontojuris. Uma boa interface para a ferramenta é fundamental para que os usuários consigam realizar buscas mais precisas e avaliar adequadamente os resultados retornados.

6. Referencias Bibliográficas

- [1] CARDOSO, Olinda N. P.; **Recuperação de Informação**, Departamento de Ciência da Computação - Universidade Federal de Lavras, 2003.
- [2] BERNERS-LEE, T., et al., **Semantic Web Development Proposal**, 2001. Disponível em: <<http://www.w3c.org/2001/sw/>>.
- [3] DAVIES, John; FENSEL, Dieter; HARMELEN, Frank van. **Towards The Semantic Web: Ontology-driven Knowledge Management**. England: John Wiley & Sons Ltd, 2003.
- [4] GRUBER, T. R. **Toward principles for the design of ontologies used for knowledge sharing**. *International Journal of Human-Computer Studies*, v. 43, 1995.
- [5] KOGUT, Paul; HOLMES, William. **AeroDAML: Applying Information Extraction to Generate DAML Annotations from Web Pages**. Philadelphia. 2001. <http://semannot2001.aifb.uni-karlsruhe.de/positionpapers/AeroDAML3.pdf>
- [6] HANDSCHUH, S., STAAB, S. and CIRAVOGNA, F., **S-CREAM-Semi-automatic CREATION of Metadata in SAAKM**. *Semantic Authoring, Annotation & Knowledge Markup - Preliminary Workshop Programme 2002*.
- [7] MOTTA, E., VARGAS-VERA, M., DOMINGUE, J., LANZONI, M., STUTT, A., CIRAVEGNA, F. **MnM: Ontology driven semi-automatic and automatic support for semantic markup**. In *13th International Conference on Knowledge Engineering and Knowledge Management (EKAW02)*, Siguenza, Spain, pp 379–391. 2002.
- [8] POPOV, B. et al. **KIM: Semantic annotation platform**. In: *The Second International Semantic Web Conference (ISWC2003)*. Flórida: 2003. http://www.ontotext.com/publications/KIM_SAP_ISWC168.pdf
- [9] KALYANPUR, A., PARSIA, B., HENDLER, J., GOLDBECK, J. **Smore – Semantic Markup, Ontology, and RDF Editor**. *Proceedings of 3rd International Semantic Web Conference (ISWC-2004)*, Japan, 2004.
- [10] KAHAN, Jose; KOIVUNEN, Marja-Riitta; PRUD'HOMMEAUX, Eric; and SWICKD, Ralph R. **Annotea: An Open RDF Infrastructure for Shared Web Annotations**. In *Proceedings of the Tenth International World Wide Web Conference*, Hong Kong, China, May, 2001. pp. 623-632.
- [11] NUNES, Ancelmo M., FILETO, Renato. **Uma Arquitetura para Recuperação de Informação Baseada em Semântica e sua Aplicação no Apoio a Jurisprudência**. INE – Universidade Federal de Santa Catarina - 2006