

PAULO CESAR DA SILVA MOREIRA

ONTOLOGIA E BUSCA SEMÂNTICA

**CURITIBA
2008**

PAULO CESAR DA SILVA MOREIRA

ONTOLOGIA E BUSCA SEMÂNTICA

Trabalho de conclusão de curso apresentado à disciplina Pesquisa em Informação II do Curso de Gestão da informação da Universidade Federal do Paraná como um dos requisitos para a aprovação.

Orientador: Prof. Dr. Mauro José Belli

**CURITIBA
2008**

TERMO DE APROVAÇÃO

PAULO CESAR DA SILVA MOREIRA

ONTOLOGIA E BUSCA SEMÂNTICA

Trabalho de Conclusão de Curso aprovado como requisito parcial para obtenção do grau de Bacharel no Curso de Bacharelado em Gestão da Informação, Setor de Ciências Sociais Aplicadas da Universidade Federal do Paraná, pela seguinte banca examinadora:

Orientador: Prof. Dr. Mauro José Belli
Departamento de Ciência e Gestão da Informação

Prof. Egon Walter Wildauer
Departamento de Ciência e Gestão da Informação

Prof. Dra. Helena de Fátima Nunes Silva
Departamento de Ciência e Gestão da Informação

Curitiba, de dezembro de 2008.

RESUMO

Realiza estudos exploratórios sobre o conceito de ontologia, verificando a possibilidade de utilização de ontologias como meio para diminuir dificuldades relacionadas com a recuperação da informação. Caracteriza-se por ter uma abordagem científica qualitativa e caráter descritivo exploratório quanto à forma de pesquisa. A metodologia adotada na execução contemplou as seguintes fases: identificação dos temas chaves relacionados com ontologia; levantamento de informações; análise dos conteúdos; busca de argumentos que demonstrem vantagens e meios de se utilizar ontologias; elaboração de um protótipo que apoiado em uma ferramenta de construção de ontologias buscou utilizar ontologias para a indexação e recuperação de documentos eletrônicos com o objetivo de melhorar a localização de informação. Como resultado deste trabalho, relata-se que a utilização de ontologias traz benefícios para processos de recuperação de informação além de permitir maior interoperabilidade e compartilhamento de informações entre sistemas distintos. Permite a construção de filtros mais eficientes e uma navegação intuitiva pelos conceitos descritos na ontologia. Identifica a possibilidade do emprego de ontologias como forma de disponibilizar informações e conhecimento de forma mais estruturada em portais organizacionais e novas possibilidade de estudos com relação ao tema.

Palavras-Chave: Ontologia. OWL. Semântica. Recuperação de Informação. Informação. Conhecimento.

LISTA DE ILUSTRAÇÕES

FIGURA 1 – Representação Gráfica de uma Tripla RDF.....	17
FIGURA 2 – Representação Gráfica de uma Afirmação RDF.....	17
FIGURA 3 – Representação de Afirmação RDF Codificada em RDF/XML.....	17
FIGURA 4 – Exemplo de Código OWL.....	25
FIGURA 5 – Arquitetura do Sistema OntoWiki.....	27
FIGURA 6 – Resultado de Busca Textual Pelo Termo <i>Ontology</i>	30
FIGURA 7 – Resultado de Busca Pelo Termo <i>Article</i>	31
FIGURA 8 – Estrutura da Base de Dados da Ferramenta OntoWiki Modificada.....	33
FIGURA 9 – Interface Para Filtrar Conceitos.....	34
FIGURA 10 – Interface Para Relacionar Conceitos a um Documento.....	34
FIGURA 11 – Demonstração de Utilização de Filtro.....	35
FIGURA 12 – Interface de Resultados da Busca.....	35

LISTA DE ABREVIATURAS E SIGLAS

AIFB – *Institute of Applied Informatics and Formal Description Methods*

API – *Application Programming Interface*

DAML – *Darpa Agent Markup Language*

DL – *Description Logic*

HTML – *Hypertext Markup Language*

IA – *Inteligência Artificial*

KIF – *Knowledge Interchange Format*

MYSQL – *My Structured Query Language*

OIL – *Ontology Inference Layer*

OWL – *Web Ontology Language*

PHP – *Personal Home Page Hypertext Preprocessor*

RDF – *Resource Description Framework*

RDFS – *Resource Description Framework Schema*

SQL – *Structure Query Language*

SWRC – *Semantic Web Research Communities*

URI – *Uniform Resource Identifiers*

W3C – *World Wide Web Consortium*

XML – *eXtensible Markup Language*

SUMÁRIO

1	INTRODUÇÃO	7
1.1	PROBLEMA DE PESQUISA.....	7
1.2	JUSTIFICATIVA.....	8
1.3	OBJETIVOS.....	9
2	LITERATURA PERTINENTE	10
2.1	ONTOLOGIA.....	10
2.2	DIFERENÇAS ENTRE ONTOLOGIA E TESAUSOS	13
2.3	COMPONENTES BÁSICOS DE UMA ONTOLOGIA.....	14
2.4	TIPOS DE ONTOLOGIA.....	16
2.5	MODELO PARA DESCRIÇÃO DE RECURSOS	16
2.6	LINGUAGEM DE DESCRIÇÃO DE VOCABULÁRIO RDF	19
2.7	LINGUAGENS PARA A CRIAÇÃO DE ONTOLOGIAS	19
2.8	MARCADORES DO OWL.....	24
2.9	FERRAMENTA ONTOWIKI	26
3	METODOLOGIA	28
3.1	CARACTERIZAÇÃO DA PESQUISA.....	28
3.2	BASES DE INFORMAÇÃO.....	28
3.3	PROCEDIMENTOS METODOLÓGICOS	28
4	TESTE COM A FERRAMENTA ONTOWIKI	30
5	PROTÓTIPO DESENVOLVIDO	32
6	ANÁLISE E DISCUSSÃO DOS RESULTADOS	37
7	CONSIDERAÇÕES FINAIS	39
	REFERÊNCIAS.....	41
	APÊNDICE A – Ontologia teste construída com o editor Protege	44

1 INTRODUÇÃO

A comunicação envolvendo seres humanos é impregnada de abstração e subjetivismo e só é possível quando as partes envolvidas compartilham, utilizam ou formulam uma base de conceitos com significados explícitos e aceitos por ambos.

Esta característica se mantém presente quando se tenta buscar informação disponibilizada pela tecnologia da informação e ferramentas de recuperação de informação disponíveis. Os resultados da busca nem sempre são satisfatórios e perde-se muito tempo minerando e analisando os resultados gerados. Outro fator que também cria barreiras e aumenta as dificuldades relacionadas com a recuperação de informação é a explosão informacional que faz com que o tempo gasto para identificar a informação desejada seja maior.

O termo ontologia ganha significados diferentes quando analisado de óticas diferentes. No contexto da informática, para Gruber (1993 *apud* FLÓRIDO, 2004), uma ontologia visa formalizar e deixar explícita uma conceituação ligada a um determinado domínio, construída consensualmente, por meio de relações, restrições, axiomas, funções e propriedades, sendo codificada em uma linguagem legível por máquinas.

Este trabalho tem como proposta realizar estudos exploratórios sobre o conceito de ontologia, verificando a possibilidade de utilização de ontologias como meio para diminuir as dificuldades relacionadas com a recuperação da informação.

1.1 PROBLEMA DE PESQUISA

Desde o surgimento da ciência da informação, na segunda guerra mundial até hoje, um dos problemas estudados tem sido a explosão informacional e as dificuldades relacionadas com disponibilização/recuperação de informação e conhecimento.

Bush (1945 *apud* Saracevic, 1996) define em seu artigo este problema como sendo a tarefa massiva de tornar mais acessível, um acervo crescente de conhecimento e propõe uma solução tecnológica para o mesmo.

A partir de então, mais precisamente durante os anos 50 e 60, programas estratégicos financiaram esforços para controlar a explosão informacional focando primeiramente na ciência e tecnologia e posteriormente em outros campos, desenvolvendo a indústria da informação.

Mesmo após muitos anos de estudos e inovações tecnológicas, ainda existe certo grau de dificuldade na tarefa de localização e recuperação de informação. É com base nesta dificuldade que defini-se como problema de pesquisa para este trabalho, verificar a possibilidade de utilizar ontologias com o objetivo de diminuir as dificuldades relacionadas com a recuperação da informação.

1.2 JUSTIFICATIVA

Em tempos extremamente competitivos, dinâmicos e globalizados, variáveis como tempo, agilidade e eficiência ganham maior importância em momentos de obtenção de informação e conhecimento.

Por questão de sobrevivência, o ser humano tende a buscar inovações, aprimorar seu conhecimento e buscar respostas para os problemas que enfrenta. O que por consequência, gera de certa forma um ciclo de competitividade e necessidades informacionais sobre o qual estamos submetidos. A cada novo problema que temos, surge uma nova necessidade de busca de informação e o emprego de esforços para tal tarefa.

Diante deste contexto, estudos que propiciem ou colaborem com esforços para diminuir as dificuldades relacionadas com a recuperação da informação são pertinentes e contribuem com o progresso da ciência da informação, tecnológico e da sociedade.

1.3 OBJETIVOS

Este trabalho tem como objetivo principal verificar se é possível utilizar ontologias com o objetivo de diminuir as dificuldades relacionadas com a recuperação da informação e para aumentar a relevância dos resultados considerando o contexto da Web.

Este trabalho tem como objetivos específicos:

- a) compreender o que é uma ontologia;
- b) identificar linguagens e ferramentas de construção e manipulação de ontologias;
- c) verificar se é possível utilizar ontologias como base conceitual de um domínio para indexar e localizar documentos;
- d) elaborar uma aplicação informatizada apoiada em tecnologias de construção de ontologia de maneira a avaliar as possibilidades da utilização de modelos baseados em ontologias para otimizar a recuperação de informações através da *World Wide Web*.

2 LITERATURA PERTINENTE

A revisão bibliográfica se constitui de uma abordagem sobre Ontologia, linguagem OWL utilizada para construção de Ontologias para Web, conceitos relacionados com OWL e ontologias que se tornem necessários para o desenvolvimento e compreensão do trabalho.

2.1 ONTOLOGIA

O termo ontologia sobre a ótica da Filosofia, dependendo do autor, se relaciona com diferentes definições como o estudo do ser enquanto ser ou ao caráter primário da ciência. Para Florido (2004, p. 37), na inteligência artificial o termo ganha um significado diferenciado, relacionado com a representação do conhecimento.

Para Gruber (1993 *apud* FLÓRIDO, 2004), uma Ontologia é uma especificação formal e explícita de uma conceituação compartilhada. "Formal" significa legível para computadores; "especificação explícita" diz respeito a conceitos, propriedades, relações, funções, restrições, axiomas, explicitamente definidos; "compartilhado" quer dizer conhecimento consensual, e "conceituação" implica um modelo abstrato de algum fenômeno do mundo real.

A conceituação acima permite concluir que poderia ser construída uma ontologia do domínio da medicina e definindo todos os termos ou conceitos utilizados de forma consensual pelas pessoas atuantes nesta área e explicitado relações entre estes conceitos, restrições, axiomas de tal forma que estas restrições, relações e axiomas pudessem ser checados por agentes computacionais (programas, sistemas).

Um recorte simplificado desta possível ontologia poderia conter a definição do conceito de gripe, de febre e o relacionamento entre os conceitos de gripe e febre. Consideremos que quando uma pessoa está com gripe, obrigatoriamente ela terá febre. Esta situação então seria um axioma e estaria explicitada na ontologia de

medicina, permitindo a sua identificação e conseqüentemente sua utilização em processos de buscas.

Uschold (1996 *apud* FLÓRIDO, 2004) menciona que uma Ontologia é uma reunião explícita de conhecimento compartilhado em uma área específica.

Bueno (2005) diz que o termo Ontologia, no campo da tecnologia, faz referência à formulação de esquema conceitual, dentro de determinado domínio, com a finalidade de facilitar a comparação, a classificação, a organização e o armazenamento dos textos analisados. Assim, as Ontologias permitem que o sistema computacional realize raciocínios indutivos, semelhantes ao do especialista. Como a representação do conhecimento é um dos principais procedimentos da Inteligência Artificial, as Ontologias aparecem como um eficiente meio de efetivar tal representação.

A partir do que expôs Bueno (2005), é possível concluir que sua declaração reforça a possibilidade de utilização de ontologias para melhorar os resultados dos mecanismos de busca.

Segundo Guarino (1998 *apud* FLÓRIDO, 2004, p. 38), uma ontologia está para uma dependência de linguagem enquanto uma conceituação seria independente de linguagem e isto reflete o aspecto de que duas ontologias podem compartilhar o mesmo conceito, porém com palavras diferentes. Isto significa também que duas ontologias podem usar as mesmas palavras, porém podem não tratar da mesma conceituação.

Com relação ao aspecto mencionado por Guarino (1998 *apud* FLÓRIDO, 2004, p. 38) de que uma conceituação seria independente de linguagem e que duas ontologias podem compartilhar o mesmo conceito, porém com palavras diferentes, cita-se uma situação em que existem duas ontologias, uma escrita em língua portuguesa e outra escrita em língua inglesa, que definem de forma idêntica o conceito de “Homem”, agregando o significado “Homem é todo ser humano do sexo masculino”. Neste caso o conceito de Homem na ontologia de língua inglesa estaria associado a palavra “*Men*” e na ontologia de língua portuguesa estaria associado a palavra “Homem”.

Para exemplificar o que diz Guarino (1998 *apud* FLÓRIDO, 2004, p. 38) com relação à possibilidade de existir duas ontologias diferentes que definam conceitos diferentes utilizando as mesmas palavras, cita-se uma situação imaginária em que o termo “Quadrado” signifique “Polígono regular de quatro lados de mesmo

comprimento” em uma ontologia X e em outra ontologia Y este mesmo termo signifique “Trajetória que descreve uma circunferência”. Neste caso a palavra “Quadrado” teria um significado totalmente diferenciado de acordo com cada ontologia.

“Uma ontologia é criada por especialistas e define as regras que regulam a combinação entre termos e relações em um domínio do conhecimento.” (ALMEIDA; BAX, 2003).

A participação de especialistas do domínio da ontologia, como citado anteriormente, permite diminuir chances de criar definições incoerentes e aumenta a qualidade e precisão das descrições dos conceitos porque os especialistas possuem conhecimento mais aprofundado sobre determinados domínios.

2.2 DIFERENÇAS ENTRE ONTOLOGIA E TESAUROS

“Tesauros e ontologias são modelos de representação do conhecimento que, servindo como instrumentos de controle terminológico, auxiliam o processo de indexação e recuperação de informações por assunto.” (CAFÉ; SALES, 2008).

Segundo estes autores são inúmeros os objetivos atribuídos às ontologias e estes ultrapassam os objetivos dos tesauros. Os autores citam os seguintes objetivos:

- a) representar o mundo real ou parte dele em um determinado domínio;
- b) proporcionar a interoperabilidade entre sistemas;
- c) potencializar os níveis de serviços ofertados na web sustentando a web semântica;
- d) propiciar ganhos na engenharia de sistemas (reusabilidade, confiabilidade e especificação);
- e) fornecer base conceitual para projetos de pesquisa e execução;
- f) melhorar a comunicação e o aprendizado, fornecendo estruturas conceituais que sustentam o ensino;
- g) possibilitar serviços semânticos;
- h) ligar parte do conhecimento humano aos processamentos computacionais;
- i) descrever os significados de um domínio de modo que seja compreensível por homens e máquinas;
- j) melhorar a consistência e o reuso da informação e o compartilhamento do conhecimento;
- k) recuperação da informação e raciocínio automático de um dado domínio;
- l) capturar significados e relações entre os significados;
- m) servir como um catálogo dos tipos de coisas que existem em um domínio determinado, considerando a perspectiva de uma pessoa que use uma língua determinada, com a finalidade de falar sobre este domínio;
- n) representar os predicados, os sentidos das palavras, ou os tipos de conceitos e relações da língua utilizada para modelar os conceitos quando usados para discutir tópicos no domínio da ontologia;

- o) na web semântica, serve para fornecer uma conceitualização parcial de um dado domínio de forma compartilhada entre usuários comuns e que seja formalmente definida por uma linguagem processável por máquinas;
- p) oferecer meios concisos e sistemáticos para definir a semântica dos recursos web e;
- q) proporcionar regras de inferência e deduções racionais para que sistemas sejam inteligíveis e processados automaticamente por meio das várias linguagens de marcação.

Para Café e Sales (2008), os tesauros são voltados para a relação que existe entre a linguagem do usuário da informação com a linguagem utilizada pelos sistemas de informação que tem a preocupação de conceder um sistema simbólico que esclareçam a relação entre os termos e os conceitos. As ontologias ultrapassam esta meta de padronização da linguagem utilizada na indexação e na recuperação da informação, propondo ser um mapa semântico, uma estrutura formal para um determinado domínio, possuindo como objetivo principal viabilizar uma base de conhecimento.

2.3 COMPONENTES BÁSICOS DE UMA ONTOLOGIA

De acordo com Gruber (1996 *apud* ALMEIDA; BAX, 2003 p. 09) / Noy e Guinness (2001 *apud* ALMEIDA; BAX, 2003, p.09), os componentes básicos de uma ontologia são classes (organizadas em uma taxonomia), relações (representam o tipo de interação entre os conceitos de um domínio), axiomas (usados para modelar sentenças sempre verdadeiras) e instâncias (utilizadas para representar elementos específicos, ou seja, os próprios dados).

Isto indica que em uma ontologia encontramos a existência de vários conceitos (classes) organizados de forma hierárquica, indicações de relações existentes entre os conceitos do domínio da ontologia, indicação de quais situações ou sentenças serão sempre verdadeiras e exemplos dos conceitos da ontologia (instâncias).

Outro elemento importante que se faz presente nas ontologias são as propriedades as quais permitem completar o significado dos conceitos. Se utilizarmos como exemplo o conceito de pessoa, algumas das possíveis propriedades deste conceito seriam: idade, sexo, nome e outras.

Segundo Helfiln (2004 *apud* LORZA, 2007) ontologias devem apresentar as seguintes propriedades comuns:

- a) interoperabilidade - diferentes ontologias podem conter conceitos que possuem mesmo significado, porém descritos (modelados) de formas diferentes e para que softwares sejam capazes de integrar informação utilizando ontologias diferentes é necessário um mecanismo de mapeamento de equivalência entre conceitos dessas ontologias;
- b) compartilhamento - ontologias devem ser publicamente disponíveis e fontes de dados diferentes devem ser capazes de concordar com definições e regras de uma mesma ontologia;
- c) extensibilidade e reuso - ontologias devem ser capazes de usar e estender outras ontologias no sentido de fornecer definições adicionais;
- d) evolução - assim como informações na Web, ontologias podem mudar durante seu ciclo de vida e, portanto, uma fonte de dados deve especificar a versão da ontologia com a qual ela concorda;
- e) balanço de expressividade e escabilidade - ontologias devem não apenas ser capazes de expressar uma grande quantidade de conhecimento, mas também de fornecer meios eficientes para interpretá-la;
- f) facilidade de uso - ontologias devem apresentar conceitos e significados de forma clara, independentemente da sintaxe da linguagem de ontologia utilizada;
- g) detecção de inconsistências - diferentes ontologias ou fontes de dados podem ser contraditórias e ontologias devem, portanto fornecer um mecanismo que auxiliem as ferramentas de detecção dessas inconsistências;
- h) compatibilidade com padrões - para promover sua disseminação, ontologias devem utilizar padrões Web e da indústria.

2.4 TIPOS DE ONTOLOGIA

Segundo Hinz (2006), não existe um consenso sobre os tipos de ontologias e segundo ele as definições podem ser sintetizadas da seguinte forma:

- a) ontologias de Representação: definem as primitivas de representação - como *frames*¹, axiomas, atributos e outros – de forma declarativa;
- b) ontologias Gerais (ou de topo): contêm definições abstratas necessárias para a compreensão de aspectos do mundo, como tempo, processos, papéis, espaço, seres, coisas, etc;
- c) ontologias centrais ou genéricas de domínio (*core ontologies*): descrevem ramos de estudo de uma área e seus conceitos mais genéricos e abstratos;
- d) ontologias de domínio e de aplicação: tratam de um domínio específico de uma área genérica de conhecimento, como direito tributário, microbiologia, etc.

2.5 MODELO PARA DESCRIÇÃO DE RECURSOS

Resource Description Framework (RDF) é uma recomendação do W3C e segundo Klyne e Carroll (2004) é uma estrutura capaz de representar as informações na web que tem uma sintaxe abstrata que reflete um modelo de dados baseado em grafo com uma semântica formal e rigorosa definição de notação provendo a base para deduções em dados RDF.

Segundo Lima e Schwabe (2004), RDF tem a finalidade de proporcionar um modo de afirmar coisas sobre recursos da web ou do mundo real, restringindo a descrição dos recursos a triplas (sujeito, predicado, objeto). O sujeito é sobre quem será feita a afirmação, predicado é a propriedade do recurso que está sendo descrito e objeto é o valor que será atribuído à propriedade do recurso.

¹ Para Almeida e Bax (2003), *Frames* são estruturas de dados que contêm variáveis pertencentes a um escopo.

De acordo com Klyne e Carroll (2004), um conjunto de triplas pode ser chamado de grafo RDF e pode ser representado da seguinte maneira:

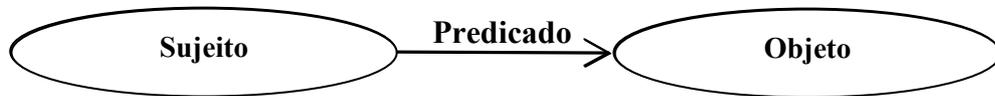


FIGURA 1 – REPRESENTAÇÃO GRÁFICA DE UMA TRIPLA RDF

Fonte: Klyne e Carroll (2004).

Seguindo esta linha de raciocínio, a representação de forma gráfica da afirmação de que o autor de um livro é determinada pessoa pode ser representada como a seguir:

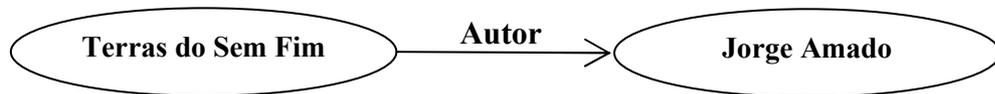


FIGURA 2 – REPRESENTAÇÃO GRÁFICA DE UMA AFIRMAÇÃO RDF

Fonte: O Autor.

Uma forma de representação da afirmação ilustrada anteriormente utilizando o vocabulário RDF seria:

```

01 <?xml version="1.0"?>
02
03 <rdf:RDF
04 xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
05 xmlns:livro="http://www.umdominioqualquer.com.br/livro#">
06
07 <Description about="http://www.terrasdosemfim.com.br/livro.pdf">
08   <livro:autor>Jorge Amado</autor>
09   <livro:titulo>Terras do Sem Fim</titulo>
10 </Description>
11 </rdf>

```

FIGURA 3 – REPRESENTAÇÃO DE AFIRMAÇÃO RDF CODIFICADA EM RDF/XML

Fonte: O Autor.

Segundo Lima e Schwabe (2004), para tornar as afirmações sobre os recursos processáveis por máquina são necessários identificadores para cada elemento da afirmação de modo a impedir qualquer possibilidade de confusão com outros similares. Segundo estes autores, um documento RDF é uma lista de

descrições onde cada descrição corresponde a um recurso e contém uma lista de propriedades.

Uma descrição corresponde a um elemento XML `<rdf:Description>` com um dos seguintes atributos: `rdf:about` ou `rdf:ID`. Se não existir nenhum destes atributos o elemento é um elemento anônimo.

Tomando como base um exemplo descrito por Refsnes Data (2008), o exemplo da Figura 3, contém na primeira linha uma declaração `xml`. Esta declaração passa a informação de que o documento foi codificado utilizando a linguagem de marcação XML. Na terceira linha, temos o elemento `rdf`, chamado de elemento *root* ou elemento de nível mais alto no documento RDF. Ou seja, é o primeiro elemento a ser acessado ou visualizado em uma hierarquia.

Segundo Bray et al (2006), em um documento XML existe a possibilidade de existir o uso de marcadores (vocabulários) iguais, mas que possuem estruturas diferentes e quando ocorre uma situação como esta, dá-se o nome de colisão. Para evitá-la, utiliza-se um recurso chamado *namespace* que é um marcador XML composto de um nome e um endereço (URI) que define a localização e um apelido para este *namespace* dentro do documento onde a declaração *namespace* foi utilizada.

O atributo `xmlns:rdf`, presente na quarta linha da Figura 3, especifica que os elementos com o prefixo `rdf` pertencem ao *namespace* "`http://www.w3.org/1999/02/22-rdf-syntax-ns#`" ou seja, indica em qual documento os elementos com o prefixo `rdf` estão definidos. O mesmo acontece na quinta linha desta figura, na qual é especificado que todos os elementos do documento com o prefixo `livro` pertencem ao *namespace* "`http://www.umdominioqualquer.com.br/livro#`". O elemento `<rdf:Description>` contém a descrição do recurso identificado pelo atributo `rdf:about`. Que no caso descreve que o recurso possui um autor chamado "Jorge Amado" e um título chamado "Terras do Sem Fim".

2.6 LINGUAGEM DE DESCRIÇÃO DE VOCABULÁRIO RDF

Para Brickley e Guha (2004), a linguagem de descrição de vocabulário RDFS, *RDF Schema*, é uma extensão semântica do RDF. Fornece mecanismos para descrever grupos de recursos e as relações entre esses recursos. Descrições de vocabulário *RDF Schema* são escritas em RDF usando os termos descritos no documento "*RDF Vocabulary Description Language 1.0: RDF Schema*". Estes recursos são usados para determinar características de outros recursos, tais como domínio (onde pode ser aplicado) e intervalo (valores possíveis) das propriedades.

Ou seja, de acordo com a definição acima, *RDF Schema* acrescenta vocabulário ao RDF permitindo a descrição de detalhes estruturais dos recursos, como, por exemplo, quais recursos podem utilizar as propriedades descritas e quais valores estas propriedade podem possuir.

Ainda para Brickley e Guha (2004), propriedades RDF podem ser definidas como atributos de recursos e neste sentido correspondem ao tradicional par atributo-valor que também podem representar relacionamentos entre recursos.

A linguagem de descrição de vocabulário RDF é similar às linguagens de programação orientadas a objeto. A diferença é que ao invés de descrever as classes de acordo com suas propriedades, RDF descreve as propriedades em função das classes. Ou seja, ao invés de definir que existe uma classe livro com as propriedades autor cujo valor é do tipo pessoa defini-se uma propriedade cujo domínio é livro e o intervalo (*range*) é pessoa. Esta característica facilita as definições de novas propriedades que possam existir para os recursos sem ter a necessidade de redefinir a classe.

2.7 LINGUAGENS PARA A CRIAÇÃO DE ONTOLOGIAS

Existem várias linguagens que podem ser utilizadas para formalizar as ontologias e a seguir são apresentadas algumas destas linguagens.

Linguagem	Descrição
Cycl	Linguagem formal que expressa conhecimento por meio de um vocabulário combinados em expressões, sentenças e finalmente bases de conhecimento (LENAT; GUHA, 1990).
Flogic (<i>Frame Logic</i>)	Integra <i>frames</i> e lógica de primeira ordem. Trata de uma forma declarativa os aspectos estruturais das linguagens baseadas em <i>frames</i> orientadas a objeto (identificação de objetos, herança, tipos polimórficos, métodos de consulta, encapsulamento etc.). Permite a representação de conceitos, taxonomias, relações binárias, funções, instâncias, axiomas e regras (KIFER; LAUSEN; WU, 1990).
LOOM	Descendente da família KL-ONE (<i>Knowledge Language One</i>), é baseada em lógica descritiva e regras de produção. Permite a representação de conceitos, taxonomias, relações n-árias, funções, axiomas e regras de produção (BRILL, 1993).
CARIN	Trata-se de uma combinação da Datalog (linguagem baseada em regras) e lógica descritiva ALN. Uma ontologia CARIN é construída por dois componentes terminológicos: um conjunto de conceitos com declarações de inclusão e um conjunto de regras que usam os conceitos (LEVY; ROUSSET, 1996).
GRAIL	É uma linguagem que especifica uma ontologia do domínio médico (Galen). É uma linguagem baseada em lógica descritiva, terminologicamente limitada, que permite a construção de hierarquias de primitivas e axiomas de inclusão de conceitos (RECTOR et al, 1997).
Ontolíngua	Combina paradigmas das linguagens baseadas em <i>frames</i> e lógica de primeira ordem. Permite a representação de conceitos, taxonomias de conceitos, relações n-árias, funções, axiomas, instâncias e procedimentos. Sua alta expressividade causa problemas na construção de mecanismos de inferência (CHAUDHRI et al, 1998).
OCML	Permite a especificação de funções, relações e classes,

	instâncias e regras. Utilizada em aplicações de gerenciamento do conhecimento, desenvolvimento de ontologias, comércio eletrônico e sistemas baseados em conhecimento. Aplicada em medicina, ciências sociais, memória corporativa, engenharia, portais da Web etc. (DOMINGUE et al, 1999).
OML (<i>Ontology Markup Language</i>)	Linguagem baseada em lógica descritiva e grafos conceituais que permite a representação de conceitos organizados em taxonomias, relações e axiomas (KENT, 1999).
RDF (<i>Resource Description Framework</i>) / RDFS (<i>RDF Schema</i>)	Desenvolvidos pelo W3C, têm por objetivo a representação de conhecimento por meio da idéia de redes semânticas. São linguagens que permitem a representação de conceitos, taxonomias de conceitos e relações binárias (LASSILA; SWICK, 1999).
NKRL (<i>Narrative Knowledge Representation Language</i>)	Linguagem de representação baseada em <i>frames</i> especialmente desenvolvida para descrever modelos semânticos de documentos multimídia (BERTINO; BARBARA; ZARRI, 1999).
SHOE (<i>Simple HTML Ontology Extensions</i>)	Utiliza extensões ao HTML, adicionando marcações para inserir metadados em páginas Web. As marcações podem ser utilizadas para a construção de ontologias e para anotações em documentos da Web. (HEFLIN; HENDLER, 2000).
XOL	É uma linguagem que pode especificar conceitos, taxonomias e relações binárias. Não possui mecanismos de inferência e foi projetada para o intercâmbio de ontologias no domínio da biomédica (KARP, 1997).
OIL (<i>Ontology Interchange Language</i>)	Precursor do DAML+OIL e base para uma linguagem para a Web Semântica. Combina primitivas de modelagem das linguagens baseadas em <i>frames</i> com a semântica formal e serviços de inferência da lógica descritiva. Pode verificar classificação e taxonomias de conceitos (FENSEL et al, 2001).
DAML (<i>Darpa Agent Markup Language</i>) + OIL	DAML+OIL é uma linguagem de marcação semântica para a Web que apresenta extensões a linguagens como o DAML, RDF e RDFS, por meio de primitivas de modelagem baseadas em

	linguagens lógicas. (HORROCKS et al, 2001).
FOML (<i>Formal Ontology Markup Language</i>)	Trata-se de uma linguagem de marcação, baseada em XML, que conecta documentos da Web com ontologias formais. O objetivo é a aquisição automática de conhecimento de domínios específicos (OGATA, 2001).

QUADRO 1 - LINGUAGENS PARA CONSTRUÇÃO DE ONTOLOGIAS

Fonte: Almeida e Bax (2003).

A linguagem no qual será mantido foco neste trabalho é a OWL devido ao fato de ser uma recomendação do W3C e porque segundo McGuinness e Harmelen (2004), OWL está relacionada com o contexto da Web e Web Semântica.

De acordo com McGuinness e Harmelen (2004), a OWL foi projetada para ser usada com aplicações que necessitam de processamento de conteúdos informacionais ao invés de apenas apresentar informações para o homem e para satisfazer a necessidade de uma Linguagem Ontológica para Web. OWL proporciona maior interoperabilidade de máquina do que conteúdos web construídos com XML, RDF, e RDF *Schema* (RDF-S) por fornecer vocabulário adicional junto com uma semântica formal. OWL possui três expressivas sub-linguagens: *OWL Lite*, *OWL DL* e *OWL Full* e faz parte de um conjunto de recomendações W3C relacionadas com a Web Semântica.

Segundo McGuinness e Harmelen (2004), as sublinguagens do OWL possuem algumas características distintas. *OWL Lite* suporta principalmente usuários que necessitem de uma classificação hierárquica e restrições simples. Por exemplo, ao mesmo tempo em que suporta restrições de cardinalidade, ela só permite valores de cardinalidade 0 ou 1. É mais simples fornecer ferramentas que suportem *OWL Lite* que seus parentes mais expressivos, e ela também permite um caminho de migração mais rápido de tesouros e outras taxonomias. *OWL Lite* também tem uma menor complexidade formal que *OWL DL*.

OWL DL, de acordo com McGuinness e Harmelen (2004), suporta usuários que querem o máximo de expressividade, embora mantendo a completude computacional (garante-se que todas as conclusões sejam computáveis) e decidibilidade (todos os cálculos serão feitos em tempo finito). *OWL DL* inclui todas as construções da linguagem OWL, mas elas podem ser usadas apenas em certas restrições (por exemplo, enquanto uma classe pode ser uma subclasse de várias

classes, uma classe não pode ser uma instância de uma outra classe). *OWL DL* é assim chamado devido à sua correspondência com a lógica descritiva, um campo de pesquisa que tem estudado as lógicas que formam a base formal da OWL.

OWL Full, de acordo com McGuinness e Harmelen (2004), foi projetada para os usuários que querem máxima expressividade e liberdade sintática do RDF sem garantia computacional. Por exemplo, em *OWL Full* uma classe pode ser tratada simultaneamente como uma coleção de indivíduos e como um indivíduo por si mesma. *OWL Full* permite uma ontologia expandir o vocabulário pré-definido (RDF ou OWL). No entanto, é pouco provável que qualquer software consiga efetuar ações de raciocínio para cada característica da *OWL Full*.

Segundo McGuinness e Harmelen (2004), o seguinte conjunto de relações é verdadeiro, já seu inverso não é:

- a) toda ontologia *OWL Lite* válida é uma ontologia *OWL DL* válida;
- b) toda ontologia *OWL DL* válida é uma ontologia *OWL Full* válida;
- c) toda conclusão *OWL Lite* válida é uma conclusão *OWL DL* válida;
- d) toda conclusão *OWL DL* válida é uma conclusão *OWL Full* válida.

McGuinness e Harmelen (2004) afirmam também que *OWL Full* pode ser vista como uma extensão de RDF, enquanto *OWL Lite* e *OWL DL* podem ser vistas como extensões de uma visão restrita do RDF. Todo documento OWL (Lite, DL, Full) é um documento RDF, e todo documento RDF é um documento *OWL Full*, mas apenas alguns documentos RDF serão um documento válido *OWL Lite* ou *OWL DL*. Devido a isso, alguns cuidados precisam ser tomados quando um usuário deseja migrar de um documento RDF para OWL.

Com relação à questão de migrar um documento RDF para *OWL Lite* ou *OWL DL*, McGuinness e Harmelen (2004) mencionam que algumas precauções devem ser tomadas para garantir que o documento RDF original esteja em conformidade com as restrições impostas pelas linguagens *OWL DL* e *OWL Lite*. Entre outros, cada URI que é usado como o nome de uma classe deve ser expressamente associado ao tipo *owl:Class* (e o mesmo acontece com propriedades), cada indivíduo tem que ser tipado (estar associado ao marcador correspondente no padrão OWL), ou seja, deve pertencer a pelo menos uma classe (ainda que apenas

owl:Thing), As URI's utilizadas para classes, propriedades e os indivíduos devem ser mutuamente *disjoint*, ou seja, não podem possuir nenhuma instância em comum.

2.8 MARCADORES DO OWL

A seguir são apresentados de acordo com McGuinness e Harmelen (2004), alguns dos marcadores do OWL:

- a) *Class* - uma classe define um grupo de indivíduos que possuem algumas características em comum;
- b) *rdfs:subClassOf* - hierarquias de classes podem ser criadas fazendo uma ou mais declarações indicando que uma classe é uma subclasse de outra classe;
- c) *rdf:Property* - propriedades podem ser utilizadas para indicar as relações entre indivíduos ou de indivíduos para os valores dos dados;
- d) *rdfs:subPropertyOf* - hierarquias de propriedades podem ser criadas fazendo uma ou mais declarações de que uma propriedade é uma subpropriedade de uma ou várias outras propriedades;
- e) *rdfs:domain* - o domínio de uma propriedade limita os indivíduos sobre os quais a propriedade pode ser aplicada;
- f) *rdfs:range* - o intervalo de uma propriedade limita os indivíduos que uma propriedade pode ter como valor;
- g) *owl:ObjectProperty* - subclasse de *rdf:Property* utilizada para relacionar uma instância de uma classe com uma instância de outra classe.
- h) *owl:DatatypeProperty* - subclasse de *rdf:Property* utilizada para relacionar uma instância de uma classe com um valor de dado (texto – conjunto de caracteres).
- i) *indivíduo* - os indivíduos são instâncias de classes, e propriedades podem ser usadas para relacionar um indivíduo a outro.

É apresentado a seguir, com o objetivo de demonstrar um exemplo de codificação OWL, códigos OWL que descrevem os conceitos de Livro e Pessoa:

```

01 <owl:Class rdf:ID="Livro">
02   <rdfs:label>Livro</rdfs:label>
03 </owl:Class>
04
05 <owl:DatatypeProperty rdf:ID="Titulo">
06   <rdfs:label>Título</rdfs:label>
07   <rdfs:domain rdf:resource="#Livro"/>
08   <rdfs:range rdf:resource="&xsd:string"/>
09 </owl:DatatypeProperty>
10
11 <owl:ObjectProperty rdf:ID="Autor">
12   <rdfs:label>Autor</rdfs:label>
13   <rdfs:domain rdf:resource="#Livro"/>
14   <rdfs:range rdf:resource="#Pessoa"/>
15 </owl:ObjectProperty>
16
17 <owl:Class rdf:ID="Pessoa" />
18
19 <owl:DatatypeProperty rdf:ID="Nome">
20   <rdfs:label>Nome</rdfs:label>
21   <rdfs:domain rdf:resource="#Pessoa"/>
22   <rdfs:range rdf:resource="&xsd:string"/>
23 </owl:DatatypeProperty>
24
25 <Pessoa rdf:ID="JorgeAmado">
26   <Nome rdf:datatype="&xsd:string">Jorge Amado</Nome >
27 </Pessoa>
28
29 <Livro rdf:ID="LivroTerrasDoSemFim">
30   <Titulo rdf:datatype="&xsd:string">Terras do Sem Fim</Titulo >
31   <Autor rdf:resource="#JorgeAmado"/>
32 </Livro>

```

FIGURA 4 – EXEMPLO DE CÓDIGO OWL

Fonte: O autor.

No exemplo da Figura 4, criou-se o conceito de Livro por meio do marcador *owl:Class*, definiu-se uma propriedade chamada Título que deve conter como valor um texto (uma string) e que pode ser associada ao conceito de Livro. Definiu-se uma propriedade chamada Autor que pode ser associada ao conceito de Livro e que deve conter como valor uma Pessoa. Criou-se também o conceito de Pessoa e uma propriedade chamada Nome que deve conter como valor um texto e pode ser

associada a uma Pessoa. Em seguida criou-se duas instâncias, uma do conceito de Pessoa e outra do conceito de livro. A instância de Pessoa representa uma pessoa cujo nome é Jorge Amado e a instância de Livro representa um livro cujo título é Terras do Sem Fim e seu autor é Jorge Amado.

A indicação de possibilidade de utilizar a propriedade Título com o conceito de Livro foi feita na linha 07 utilizando o marcador *rdfs:domain*. A indicação de que o valor da propriedade Título deve ser um texto foi feita na linha 08 por meio do marcador *rdfs:range*.

Na linha 13, na definição da propriedade Autor, o marcador *rdfs:domain* indica que a propriedade Autor pode ser associada ao conceito de Livro e na linha 14 por meio do marcador *rdfs:range* defini-se que o valor da propriedade Autor deve ser uma Pessoa.

Um exemplo de uma ontologia contendo mais exemplos de construções pode ser encontrado no endereço “<http://www.mindswap.org/2003/owl/swint/terrorism>”.

2.9 FERRAMENTA ONTOWIKI

OntoWiki é uma ferramenta de apoio a engenharia do conhecimento para ambiente web. Chama-se OntoWiki porque possui características dos sistemas Wiki existentes. No entanto, o projeto é independente e completa as tecnologias convencionais Wiki. OntoWiki tem como abordagem simplificar a apresentação e construção de bases de conhecimento (AUER; DIETZOLD; RIECHERT, 2006).

OntoWiki é uma aplicação web de código aberto escrita em PHP com a finalidade de criar e manter bases de conhecimento. A seguir é demonstrada a arquitetura do software:

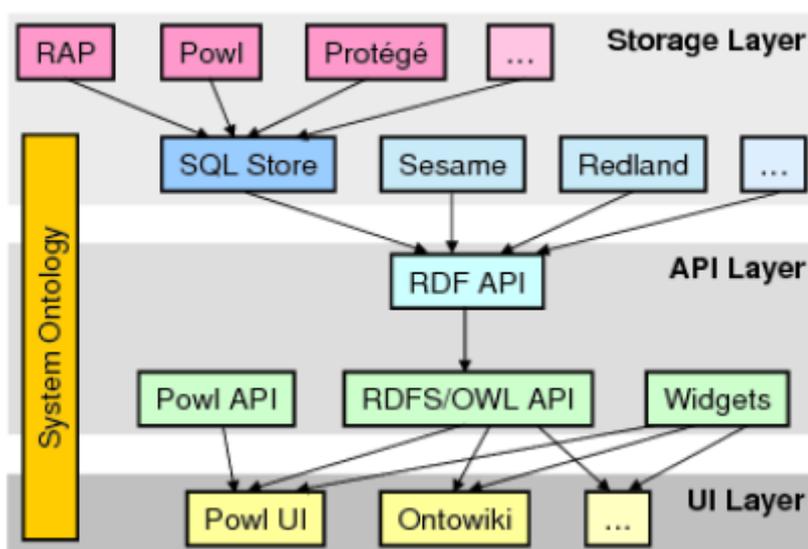


FIGURA 5 – ARQUITETURA DO SISTEMA ONTOWIKI

Fonte: (AUER; DIETZOLD; RIECHERT, 2006).

A arquitetura da ferramenta demonstra que é possível construir ontologias utilizando diferentes ferramentas como Protege ou Powl (*Semantic Web Application Development Platform*) e depois importar para o OntoWiki.

Um dos principais objetivos do OntoWiki é de promover a interação social para desenvolver bases de conhecimento. Isto facilita o intercâmbio de metadados sobre as bases de conhecimento e promove a colaboração em cenários onde a comunicação pessoal é difícil (AUER; DIETZOLD; RIECHERT, 2006)

OntoWiki foi implementado como uma alternativa de interface do usuário para o editor Powl, uma plataforma de desenvolvimento de aplicações da Web Semântica.

3 METODOLOGIA

Nesta seção, apresentam-se a caracterização da pesquisa, bases de informação e procedimentos metodológicos utilizados na execução deste trabalho.

3.1 CARACTERIZAÇÃO DA PESQUISA

A pesquisa se caracterizou por ter uma abordagem científica qualitativa porque visou analisar uma situação, a possibilidade de otimização de busca com a utilização de ontologias, sem quantificar ou enumerar. Caracteriza-se também como sendo descritiva exploratória quanto à forma de pesquisa pelo fato de buscar conhecer os conceitos relacionados com ontologia e obter com isto novas idéias e percepções.

3.2 BASES DE INFORMAÇÃO

Foram utilizados para a elaboração deste trabalho, como referencial teórico, padrões e recomendações estabelecidas pelo W3C, artigos da área de Ciência da Informação, teses e dissertações que abordam o tema ontologia e temas relacionados.

3.3 PROCEDIMENTOS METODOLÓGICOS

Na primeira etapa de execução deste trabalho, definiram-se quais seriam os temas relacionados com o problema de pesquisa e quais seriam pesquisados e descritos.

Elaborado a lista de temas que necessitavam de dados a fim de permitir relacionar conceitos e embasar teóricamente o trabalho, passou-se então para a coleta e análise dos dados buscando argumentos que demonstrassem a possibilidade de atenuação dos problemas relacionados com a recuperação de informação. Em paralelo com esta etapa, buscou-se redigir o texto final com base nos textos coletados e analisados.

De posse de uma base conceitual, procurou-se então identificar uma possível forma de resolução do problema de pesquisa e ao mesmo tempo foram feitas pesquisas exploratórias por tecnologias que poderiam auxiliar na construção de uma solução. Foram identificadas duas ferramentas que poderiam ser utilizadas durante o desenvolvimento do trabalho, o editor de ontologias Protege e a ferramenta OntoWiki.

Buscou-se compreender a estrutura de uma ontologia e posteriormente criar uma ontologia pequena, com poucos conceitos, utilizando o editor de ontologias Protege. Em seguida foram feitos estudos iniciais na ferramenta OntoWiki, importando a ontologia teste criada para esta ferramenta e observada a possibilidade de utilizar a base de dados e o recurso de importação de ontologias do OntoWiki para a construção da solução procurada.

Após as etapas anteriores, foi construído um protótipo que utiliza os conceitos modelados em ontologias para a indexação e localização de documentos. Com o objetivo de verificar se o formato de indexação e localização dos documentos proposto estava funcional, foi realizada a indexação de um documento disponível na internet relacionando este documento a um conceito existente em uma ontologia presente na base de dados do OntoWiki. Para comprovar a indexação do documento, foi realizada, no protótipo, uma busca pelo documento indexado utilizando como filtro o conceito relacionado com o documento.

Posteriormente a construção do protótipo, novas buscas de informações foram feitas, sobre o tema ontologia, e detectado um ontologia chamada SWRC (*Semantic Web for Research Communities*) que propiciou um novo teste com a ferramenta OntoWiki.

4 TESTE COM A FERRAMENTA ONTOWIKI

Com o objetivo de explorar a ferramenta e buscar maiores compreensões sobre o seu funcionamento, uma ontologia chamada SWRC escrita com o padrão OWL, que segundo Oberle et al. (2005), modela entidades de comunidades de pesquisas, como pessoas, organizações, publicações (metadados bibliográficos) e seus relacionamentos, foi importada para a base de dados da ferramenta OntoWiki.

Após a importação foi possível fazer buscas na ontologia de forma textual (utilizando linguagem natural) e navegando entre os conceitos.

A figura a seguir demonstra, no quadro chamado Classes, parte dos conceitos existentes na ontologia importada e, no quadro *Search*, uma lista de resultados de uma busca textual pelo termo *ontology*:

The screenshot displays the OntoWiki search interface. On the left, there is a search form with the text 'ontology' entered and a 'Submit' button. Below the search form are two panels: 'Knowledge Bases' and 'Classes'. The 'Classes' panel lists various ontology classes such as 'swrc:Article (168)', 'swrc:AssistantProfessor (2)', 'swrc:Book (27)', 'swrc:FullProfessor (5)', 'swrc:Graduate (1)', 'swrc:InCollection (94)', and 'swrc:InProceedings (740)'. The main search results area on the right shows five results, each with its title, class name, relevance, and property value. The results are:

- ontology-based knowledge management systems**
swrc:ResearchTopic
Relevance: 100%
Property: swrc:name
Value: ontology-based knowledge management systems
- ontology engineering**
swrc:ResearchTopic
Relevance: 100%
Property: swrc:name
Value: ontology engineering
- ontology learning**
swrc:ResearchTopic
Relevance: 100%
Property: swrc:name
Value: ontology learning
- Alain Giboin**
swrc:Person
Relevance: 100%
Property: swrc:homepage
Value: http://knowledgeweb.semanticweb.org/semanticportal/home.jsp?_origin=%2Fhome.jsp&instance=Alain+Giboin&ontology=Person+Ontology&content=instance.jsp&instance_set=kweb&var_name=instance
- The SWRC Ontology - Semantic Web for Research Communities**
swrc:Publication
Relevance: 100%
Property: swrc:abstract

FIGURA 6 – RESULTADO DE BUSCA TEXTUAL PELO TERMO *ONTOLOGY*

Fonte: OntoWiki 0.8.

A figura a seguir demonstra os resultados de uma busca por instâncias do conceito *Article* resultante da navegação pelos conceitos (Classes):

The screenshot displays the OntoWiki interface. On the left, there is a search bar and a list of knowledge bases (AIFB_OWL, Generica, Ontologia, medicina, Pizza) and classes (swrc:Article (168), swrc:AssistantProfessor (2), swrc:Book (27), swrc:FullProfessor (5), swrc:Graduate (1), swrc:InCollection (94), swrc:InProceedings (740)). The main area shows the search results for the class 'swrc:Article', listing 8 instances with titles and associated class information. On the right, there is a 'Show Properties' panel and a 'Filter' panel.

FIGURA 7 – RESULTADO DE BUSCA PELO TERMO *ARTICLE*

Fonte: OntoWiki 0.8.

A figura anteriormente apresentada demonstra no centro, todas as instâncias do conceito *Article* encontradas. No canto direito, uma caixa denominada *Show Properties* com todas as propriedades que podem ser associadas ao conceito *Article*. Na caixa *Filter*, demonstra os diferentes valores encontrados nas instâncias do conceito *Article* como opção de filtro das instâncias listadas.

Com base no teste realizado, é possível confirmar algumas características e vantagens no uso de ontologias como a interoperabilidade e intercâmbio de informações e tornar mais acessível e processável as informações.

A questão da interoperabilidade e intercâmbio de informações evidenciou-se após acessar uma ontologia disponível na internet, a ontologia no formato OWL do instituto AIFB, importar para o *software* OntoWiki e este *software* ser capaz de processar as informações da ontologia. É interessante observar que o arquivo onde se encontra a ontologia original não está preso a nenhum sistema ou banco de dados, foi possível acessar o arquivo da ontologia por meio de um endereço da internet.

5 PROTÓTIPO DESENVOLVIDO

Com o objetivo de identificar uma forma de utilizar ontologias para melhorar processos de recuperação da informação, foi construído um protótipo (conjunto de interfaces que se comunica com a base de dados da ferramenta OntoWiki) que permite recuperar as Classes (conceitos) de cada ontologia existente na base de dados e utilizar estes conceitos para indexar e localizar documentos eletrônicos.

Analisando a base de dados da ferramenta OntoWiki, identificou-se que as ontologias que eram importadas para o sistema, estavam sendo armazenadas preservando as características do formato RDF e que por meio de consultas SQL era possível reconstruir a hierarquia de conceitos das ontologias e recuperar suas propriedades e relações.

A figura a seguir apresenta a estrutura da base de dados da ferramenta OntoWiki, acrescida das tabelas (*conceitos_documentos* e *documentos*) geradas para o caso particular em estudo, demonstrando apenas os relacionamentos referentes a estas tabelas porque não foi possível reconstruir ou identificar de forma precisa os relacionamentos existentes na base de dados do OntoWiki devido ao tipo de banco de dados usado pela ferramenta ser MySQL e pelo fato dos relacionamentos não existirem na base de dados.

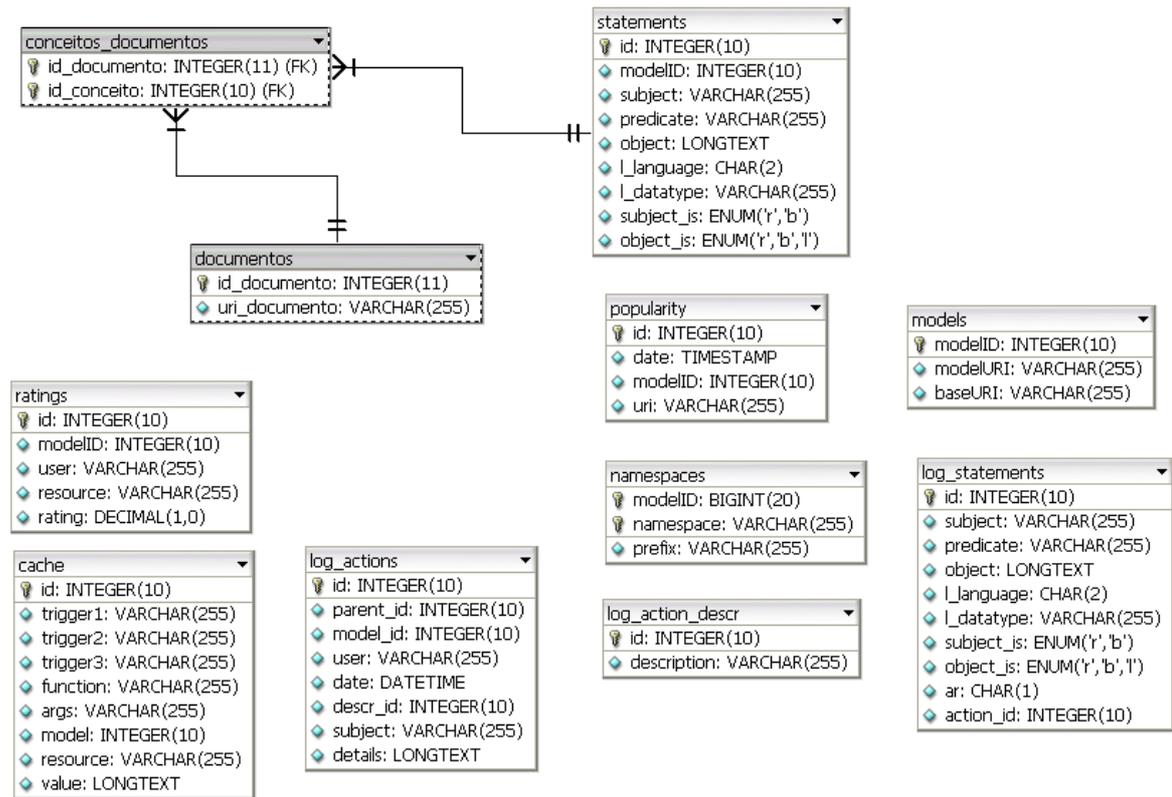


FIGURA 8 – ESTRUTURA DA BASE DE DADOS DA FERRAMENTA ONTOWIKI MODIFICADA

Fonte: O Autor.

Analisando a estrutura das tabelas e os dados existentes em cada uma, identificou-se que na tabela *models* ficam armazenados os endereços URI e uma abreviação do URI (*alias*) de todas as ontologias importadas para a base de dados. Na tabela *namespaces* ficam armazenados as declarações *namespaces* presentes em cada ontologia. A tabela *statements* armazena as declarações da ontologia no formato RDF.

Foram criadas duas tabelas, uma chamada *documentos* e outra chamada de *conceitos_documentos*. A tabela *documentos* foi criada com o objetivo de cadastrar os documentos eletrônicos e a tabela chamada *conceitos_documentos* foi criada com o objetivo de relacionar os conceitos aos documentos e indicar com isto quais conceitos existem em cada documento.

O protótipo foi construído utilizando as linguagens de programação PHP, HTML, Javascript e biblioteca JQuery. A seguir são apresentadas as telas criadas para a indexação dos documentos:

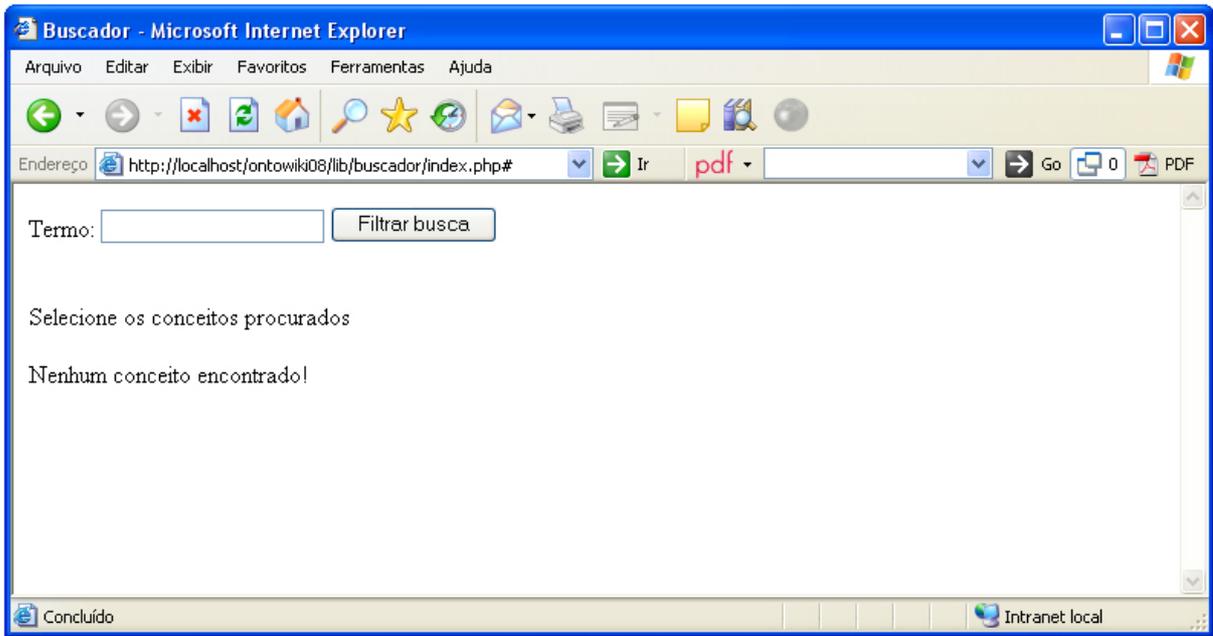


FIGURA 9 – INTERFACE PARA FILTRAR CONCEITOS

Fonte: Protótipo.

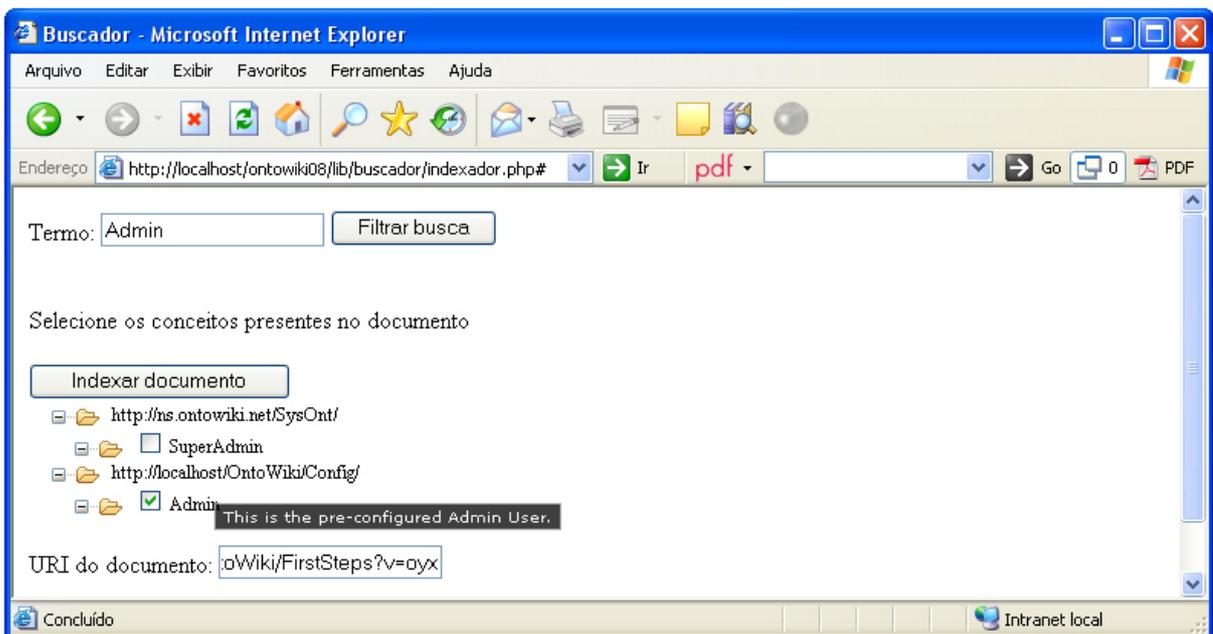


FIGURA 10 – INTERFACE PARA RELACIONAR CONCEITOS A UM DOCUMENTO

Fonte: Protótipo.

A interface de indexação permite relacionar conceitos distintos e de ontologias distintas a um documento. Após a etapa de indexação, se torna possível realizar buscas por estes documentos no protótipo. A seguir são apresentadas as interfaces gráficas para localização de documentos:

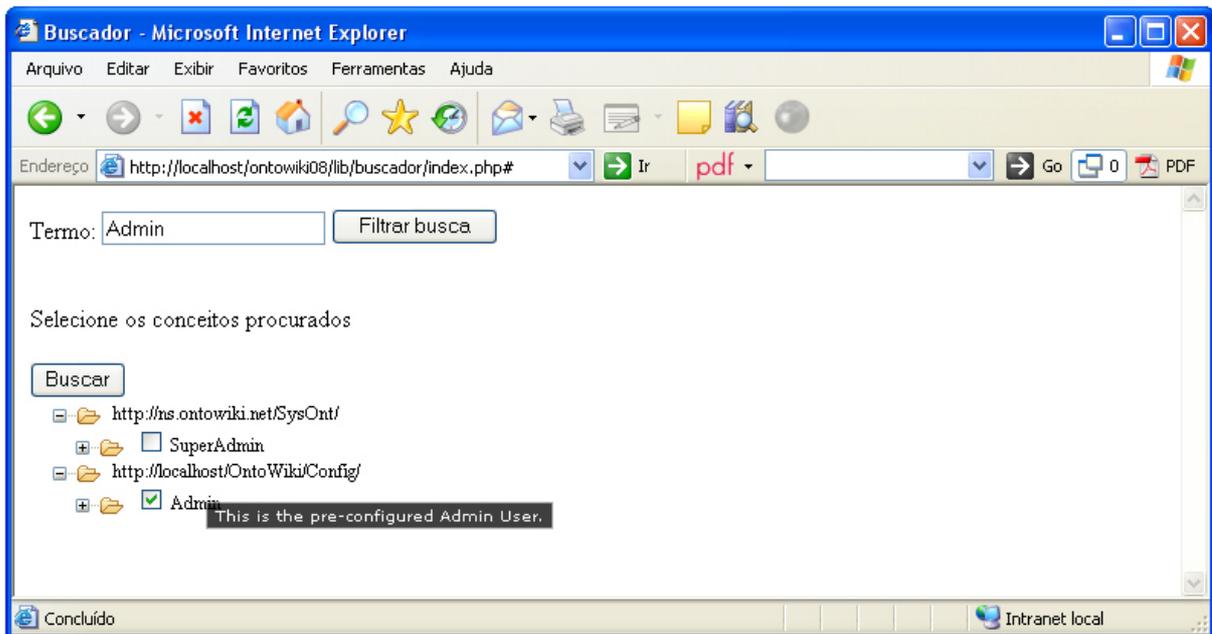


FIGURA 11 – DEMONSTRAÇÃO DE UTILIZAÇÃO DE FILTRO

Fonte: Protótipo.

Efetuada a filtragem desejada e clicando no botão *Buscar*, são listados na tela, os resultados encontrados como na figura a seguir:

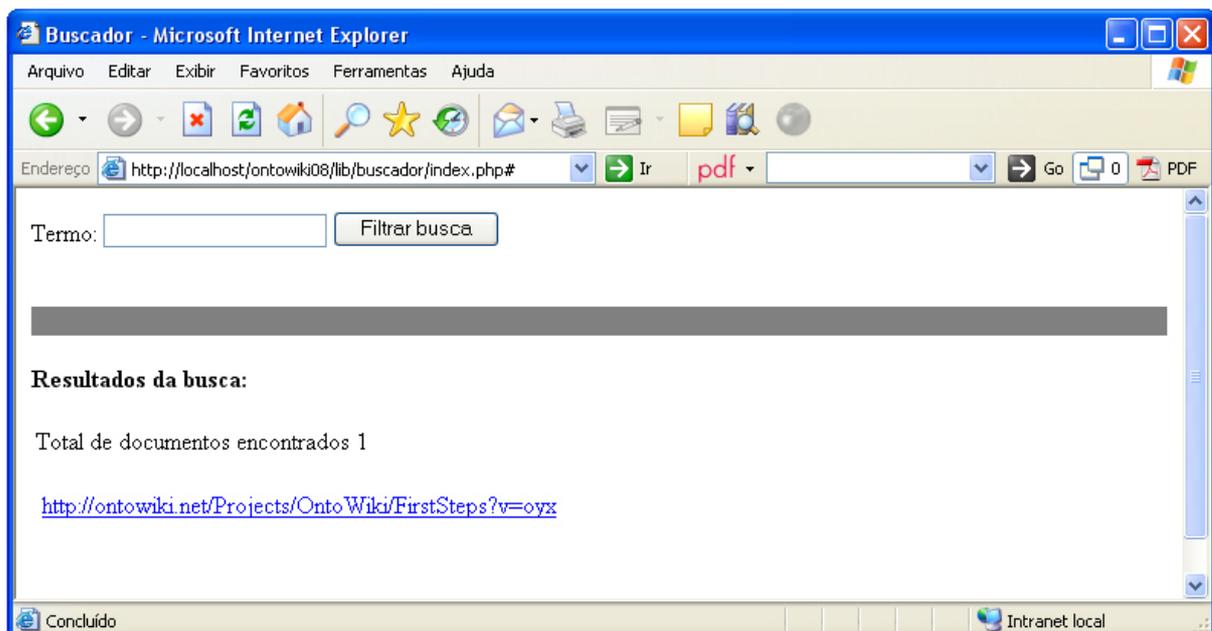


FIGURA 12 – INTERFACE DE RESULTADOS DA BUSCA

Fonte: Protótipo.

A Figura 12 apresentada anteriormente contém em forma de listagem todos os documentos que contém a utilização do conceito *Admin*, que contenham conteúdo relacionado com “usuário administrador pré-configurado”.

6 ANÁLISE E DISCUSSÃO DOS RESULTADOS

Após estudos sobre o assunto e exploração da ferramenta OntoWiki, identificou-se algumas vantagens e comprovou-se outras já mencionadas pelos autores citados com relação a utilização de ontologias:

- a) ontologias permitem a construção de filtros mais eficientes, filtrar de acordo com as propriedades de cada conceito;
- b) uma ontologia permite criar um mapa semântico entre os conceitos pelo fato de permitir a descrição detalhada e relacionamento entre os conceitos;
- c) as ontologias formalizam os conceitos utilizados em determinado contexto permitindo a busca por estes conceitos;
- d) ontologias permitem realizar buscas de forma intuitiva por meio de navegação entre os conceitos;
- e) permite a navegação entre todas as informações existentes na ontologia sem a necessidade de criar relatórios por meio de programação;
- f) as ontologias permitem fazer ligações entre recursos como uma página de internet, um arquivo PDF, foto etc;
- g) ontologias são construídas em um formato que permite o processamento por outras aplicações – interoperabilidade entre sistemas;
- h) as ontologias permitem inferência;
- i) as ontologias facilitam o reuso da informação e o compartilhamento do conhecimento;
- j) as ontologias visam ser compreensíveis por máquinas e por humanos;
- k) as ontologias viabilizam a construção de uma base de conhecimento sobre um determinado domínio.

Ao analisar a ferramenta OntoWiki verificou-se que esta serviu como exemplo prático que permitiu gerar maiores conclusões com relação a possibilidade de diminuir problemas ligados a recuperação da informação por meio da utilização de ontologias.

Observou-se a possibilidade de utilizar as camadas API e de Armazenamento da arquitetura do OntoWiki como modelo para a construção de interfaces de

navegação dos conceitos presentes em ontologias e associar tal mecanismo, por exemplo, a um portal de internet fornecendo informações de forma mais estruturada para os usuários e contribuindo para otimizar a recuperabilidade de informações destes contextos e domínios modelados.

A respeito da camada de persistência de dados, observou-se que a forma padrão utilizada pela ferramenta OntoWiki, armazenamento em base de dados, demonstra-se ser eficiente devido ao fato de permitir a recuperação dos dados das ontologias de acordo com a demanda, ou seja, aos poucos. Este mecanismo permite uma navegação mais rápida, concedendo a ferramenta maior velocidade no processamento destas informações.

Constatou-se também a necessidade de realização de testes referentes às questões de integridade dos dados das ontologias que são importadas para a ferramenta OntoWiki e ou exportadas desta ferramenta, verificando se todas as construções possíveis permitidas pelo padrão OWL são mantidas durante estes processos.

Com relação ao protótipo desenvolvido, comprovou-se que é mais eficiente criar o relacionamento entre os recursos como páginas de internet, documentos ou qualquer outro recurso eletrônico na própria ontologia ao invés de fazer relações na base de dados tendo em vista que a relação existente na ontologia pode ser processada por outra aplicação com acesso à ontologia.

Chegou-se a conclusão de que existem possibilidades para a construção de um mecanismo que utilize ontologias para a indexação e localização de documentos eletrônicos com o objetivo de facilitar a recuperação destes documentos. Observou-se também que o protótipo construído poderia ter utilizado a camada API presente na arquitetura do OntoWiki para poder recuperar os conceitos das ontologias existentes na base de dados e trabalhar com estes conceitos.

7 CONSIDERAÇÕES FINAIS

Este trabalho teve como proposta buscar identificar a possibilidade de utilização de ontologias como meio para diminuir as dificuldades relacionadas com a recuperação de informação no ambiente web.

Ao finalizar este trabalho, verificou-se que a utilização de ontologias como meio para melhorar processos de recuperação da informação é possível e se justifica porque as ontologias permitem fazer um mapa de significados ao invés de fazer um mapa de links, permitem que o usuário “aponte” o significado e contexto desejado e fornece meios para a construção de filtros mais eficientes.

Concluí-se que a associação de significado aos conceitos é conseguida por meio das descrições, relacionamentos e detalhamentos dos conceitos feitos nas ontologias e que a captação destes significados por parte do homem é feita a partir da observação destas descrições e estruturas conceituais criadas, ou seja, de forma análoga a que o ser humano está acostumado a aprender. Por exemplo, observando a instância do conceito de Livro apresenta na Figura 4, e a forma como os conceitos foram modelados nesta figura, podemos identificar que um livro pode ter uma pessoa associada como autor e um livro pode ter um título (conjunto de caracteres que representam um título). Ou seja, quanto maior a descrição dos conceitos maiores são os significados transmitidos sobre os conceitos descritos.

Observou-se a possibilidade para a utilização de ontologias como forma de disponibilizar informações e conhecimento de forma mais estruturada em portais organizacionais e de melhorar a recuperação, processamento e compartilhamento destas informações. Um portal de uma instituição de pesquisa, por exemplo, poderia utilizar a ontologia SWRC para disponibilizar as informações sobre suas pesquisas, publicações e pesquisadores, facilitando a recuperação e o compartilhamento destas informações com outras aplicações.

Como possibilidade para desenvolvimento de trabalhos futuros identificou-se:

- a) fazer um estudo de caso aprofundado sobre a ontologia SWRC e exemplos de utilização;

- b) realizar um estudo mais aprofundado sobre as construções da linguagem *OWL DL* e demonstrar por meio de exemplos a forma de modelar os conceitos de uma ontologia com este padrão;
- c) entender mecanismos de anotações OWL utilizados em portais como o do instituto AIFB, visando identificar a possibilidade de utilizar estas anotações como forma de indexar as páginas de internet em mecanismos de busca e proporcionar com isto o desenvolvimento de mecanismos de busca semânticos focando a web como um todo;
- d) entender conceitos e explorar aspectos de modularização de ontologia visando identificar possibilidades de otimização e escalabilidade;
- e) realizar estudos de identificação de barreiras existentes para utilização de ontologias visando conhecer problemas que precisam ser superados ou considerados;
- f) identificar maneiras de explorar e por em prática mecanismos de inferência sobre ontologias modeladas com OWL;
- g) realizar estudos aprofundados sobre os elementos da camada API utilizada na ferramenta OntoWiki visando compreender a forma de manipulação das ontologias utilizadas na ferramenta e conseqüentemente proporcionar a construção de uma interface de exploração das ontologias que possa ser incorporada em portais corporativos;
- h) realizar estudo exploratorio a respeito de aplicações que utilizam ontologias visando identificar as diferentes formas de explorar e usar ontologias;
- i) fazer um estudo mais aprofundado sobre a ferramenta OntoWiki visando comprovar a manutenção de aspectos de integridade das ontologias que são manipuladas pela camada API da ferramenta.

REFERÊNCIAS

ALMEIDA, Maurício Barcellos. **Uma introdução ao XML, sua utilização na Internet e alguns conceitos complementares**. Ciência da Informação, Brasília, v. 31, n. 2, maio/agos. 2002. Disponível em: <<http://www.scielo.br/pdf/ci/v31n2/12903.pdf>>. Acesso em: 26 setembro 2007.

ALMEIDA, Maurício Barcellos; Bax, P. Marcello. **Uma visão geral sobre ontologias: pesquisa sobre definições, tipos, aplicações, métodos de avaliação e de construção**. Ciência da Informação, Brasília, v. 32, n. 3, set./dez. 2003. Disponível em: <<http://www.scielo.br/pdf/ci/v32n3/19019.pdf>>. Acesso em: 17 maio 2008.

AUER, Sören; DIETZOLD, Sebastian; RIECHERT, Thomas. **OntoWiki – A Tool for Social, Semantic Collaboration**. Disponível em: <<http://www.informatik.uni-leipzig.de/~auer/publication/ontowiki.pdf>>. Acesso em: 01 novembro 2008.

BRAY, Tim; HOLLANDER, Dave; LAYMAN, Andrew; TOBIN, Richard. **Namespaces in XML 1.0 (Second Edition)**, 2006. Disponível em: <<http://www.w3.org/TR/REC-xml-names/#sec-namespaces>>. Acesso em: 08 dezembro 2008.

BRICKLEY, Dan; GUHA, R.V. **RDF Vocabulary Description Language 1.0: RDF Schema**, 2004. Disponível em : <http://www.w3.org/TR/rdf-schema/#ch_introduction>. Acesso em: 08 novembro 2008.

BUENO, Tânia Cristina D'agostini. **Engenharia da Mente: Uma Metodologia de Representação do Conhecimento para Construção de Ontologias em Sistemas Baseados em Conhecimento**. Tese (Doutorado), Universidade Federal de Santa Catarina, Florianópolis, 2005. Disponível em: <http://www.tede.ufsc.br/tedesimplificado//tde_busca/arquivo.php?codArquivo=26>. Acesso em: 17 maio 2008.

CASARE, Sara Jane. **Uma Ontologia Funcional de Reputação para Agentes**. Dissertação (Mestrado), Escola Politécnica da Universidade de São Paulo, São Paulo, 2005. Disponível em: <http://www.teses.usp.br/teses/disponiveis/3/3141/tde-22052006-221632/publico/Dissertacao_Ontologia_Funcional_Reputacao.pdf>. Acesso em: 17 maio 2008.

CAFÉ, Ligia; SALES, Rodrigo. **Semelhanças e Diferenças entre Tesouros e Ontologias**, DataGramaZero - Revista de Ciência da Informação, Rio de Janeiro, v. 9, n. 4, ago. 2008. Disponível em: <http://www.dgz.org.br/ago08/F_I_art.htm>. Acesso em: 08 novembro 2008.

FLÓRIDO, Irapuru Haruo. **Gestão do conhecimento em sistemas de informação legados – ONTOLEGACY**. Dissertação (Mestrado), Pontifícia Universidade Católica do Paraná, Curitiba, 2004.

HINZ, Verlani Timm. **Proposta de Criação de uma Ontologia de Ontologias**. Trabalho Individual I (Pós-Graduação em Informática), Universidade Católica de Pelotas, Pelotas, 2006. Disponível em: <<http://ppginf.ucpel.tche.br/TI-arquivos/2006/VerlaniHinz/PPGINF-UCPel-TI-2006-2-10.pdf>>. Acesso em: 17 maio 2008.

INSTITUTE OF APPLIED INFORMATICS AND FORMAL DESCRIPTION METHODS. **Ontologia OWL**. Disponível em: <http://www.aifb.uni-karlsruhe.de/WBS/dvr/viewAIFB_OWL.owl>. Acesso em: 01 novembro 2008.

JAIME ROBREDO, **Da Ciência da Informação Revisitada aos Sistemas Humanos de Informação**. 1. ed. Brasília: Thesaurus Editora, 2003, p. 39-91.

KLYNE, Graham; CARROLL, Jeremy J. **Resource Description Framework (RDF): Concepts and Abstract Syntax**. Disponível em: <<http://www.w3.org/TR/2004/REC-rdf-concepts-20040210/>>. Acesso em: 01 novembro 2008.

LIFEBOAT FOUNDATION SCIENTIFIC ADVISORY BOARD MEMBER NOVA SPIVACK. **The Third Generation Web is Coming**. Disponível em: <<http://lifeboat.com/ex/web.3.0>>. Acesso em: 20 outubro 2007.

LIMA, Fernanda; SCHWABE, Daniel. **Introdução à Web Semântica - MiniTutorial do evento WWS2004 I Workshop de Web Semântica - SBD/SBES2004 - Brasília**. Disponível em: <<http://ftp.informatik.rwth-aachen.de/Publications/CEUR-WS/Vol-123/WWS2004MiniTutorial.pdf>>. Acesso em: 01 novembro 2008.

LORZA, Augusto Carbol. **Ativação de Componentes de Software com a Utilização de uma Ontologia de Componentes**. Dissertação (Mestrado), Instituto de Ciências Matemáticas e de Computação - ICMC-USP, São Carlos, 2007. Disponível em: <<http://www.teses.usp.br/teses/disponiveis/55/55134/tde-13092007-104023/publico/discarbol.pdf>>. Acesso em: 31 março 2008.

REFSNES DATA. **RDF Tutorial**. Disponível em: <<http://www.w3schools.com/rdf/>>. Acesso em: 02 novembro 2008.

SARACEVIC, Tefko. **Ciência da informação origem, evolução e relações**, Perspectivas em Ciência da Informação, Belo Horizonte, v. 1, n. 1, jan./jun. 1996. Disponível em: <<http://www.eci.ufmg.br/pcionline/index.php/pci/article/viewFile/235/22>>. Acesso em: 25 setembro 2007.

McGuinness, Deborah L.; Harmelen, Frank van. **OWL Web Ontology Language Overview**, 2004. Disponível em: <<http://www.w3.org/TR/owl-features/>>. Acesso em: 26 agosto 2008.

McGuinness, Deborah L.; Welty, Chris; Smith, Michael K. **OWL Web Ontology Language Guide**, 2004. Disponível em: <<http://www.w3.org/TR/owl-guide/>>. Acesso em: 26 setembro 2007.

APÊNDICE A – Ontologia teste construída com o editor Protege

```

<?xml version="1.0"?>

<!DOCTYPE rdf:RDF [
  <!ENTITY owl "http://www.w3.org/2002/07/owl#" >
  <!ENTITY xsd "http://www.w3.org/2001/XMLSchema#" >
  <!ENTITY rdfs "http://www.w3.org/2000/01/rdf-schema#" >
  <!ENTITY rdf "http://www.w3.org/1999/02/22-rdf-syntax-ns#" >
  <!ENTITY j.0 "http://www.owl-ontologies.com/ontologia.owl#" >
  <!ENTITY protege "http://protege.stanford.edu/plugins/owl/protege#" >
]>

<rdf:RDF xmlns="http://www.owl-ontologies.com/generica.owl#"
  xml:base="http://www.owl-ontologies.com/generica.owl"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
  xmlns:protege="http://protege.stanford.edu/plugins/owl/protege#"
  xmlns:xsd="http://www.w3.org/2001/XMLSchema#"
  xmlns:owl="http://www.w3.org/2002/07/owl#"
  xmlns:j.0="http://www.owl-ontologies.com/ontologia.owl#"
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#">

  <owl:Ontology rdf:about="">
    <owl:imports rdf:resource="http://www.owl-ontologies.com/ontologia.owl"/>
  </owl:Ontology>

  <owl:Class rdf:ID="Aplicacao">
    <rdfs:label rdf:datatype="&xsd:string">Aplicação</rdfs:label>
    <rdfs:comment rdf:datatype="&xsd:string">
      >Utilização de algo.</rdfs:comment>
  </owl:Class>

  <owl:Class rdf:ID="Exemplo">
    <rdfs:comment rdf:datatype="&xsd:string">
      >Demonstração de alguma coisa.</rdfs:comment>
  </owl:Class>

  <owl:ObjectProperty rdf:ID="exemploDe">
    <rdfs:comment rdf:datatype="&xsd:string">
      >Especifica qual é o exemplo</rdfs:comment>
  </owl:ObjectProperty>

  <Aplicacao rdf:ID="ind_aplicacao">
    <rdfs:comment rdf:datatype="&xsd:string">
      >Aplicação de ontologias escritas em OWL.</rdfs:comment>
    <utiliza rdf:resource="#ind_ontologia_owl"/>
  </Aplicacao>

```

```
<Exemplo rdf:ID="ind_exemplo">
  <rdfs:comment rdf:datatype="&xsd:string"
    >Exemplo de aplicação que utiliza ontologias escritas em
  OWL.</rdfs:comment>
  <exemploDe rdf:resource="#ind_aplicacao"/>
</Exemplo>

<j.0:ontologia_owl rdf:ID="ind_ontologia_owl">
  <rdfs:comment rdf:datatype="&xsd:string"
    >Instância de uma ontologia escrita em OWL.</rdfs:comment>
</j.0:ontologia_owl>

<owl:ObjectProperty rdf:ID="utiliza">
  <rdfs:comment rdf:datatype="&xsd:string"
    >Especifica uma coisa que é utilizada por outra coisa.</rdfs:comment>
</owl:ObjectProperty>

</rdf:RDF>
```