

**PONTIFÍCIA UNIVERSIDADE CATÓLICA DO RIO GRANDE DO SUL
FACULDADE DE INFORMÁTICA
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO**

**DESCRIÇÃO E RECUPERAÇÃO DE
ONTOLOGIAS EM LÍNGUA PORTUGUESA:
UMA ADAPTAÇÃO DO MODELO *ONTOLOGY*
*METADATA VOCABULARY***

ANDERSON BESTTETI SANTOS

Dissertação apresentada como requisito parcial à obtenção de título de Mestre em Ciência da Computação na Pontifícia Universidade Católica do Rio Grande do Sul.

Orientadora: Profa. Dra. Renata Vieira

**Janeiro
2010**

Dados Internacionais de Catalogação na Publicação (CIP)

S237d Santos, Anderson Bestteti
Descrição e recuperação de ontologias em língua portuguesa : uma adaptação do modelo ontology metadata vocabulary / Anderson Bestteti Santos. – Porto Alegre, 2010. 76 f.

Diss. (Mestrado) – Fac. de Informática, PUCRS.
Orientador: Prof^ª. Dr^ª. Renata Vieira.

1. Informática. 2. Ontologia. 3. Metadados.
4. Representação do Conhecimento (Computação).
I. Vieira, Renato. II. Título.

CDD 006.35

**Ficha Catalográfica elaborada pelo
Setor de Tratamento da Informação da BC-PUCRS**



Pontifícia Universidade Católica do Rio Grande do Sul
FACULDADE DE INFORMÁTICA
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

TERMO DE APRESENTAÇÃO DE DISSERTAÇÃO DE MESTRADO

Dissertação intitulada "Descrição e Recuperação de Ontologias em Língua Portuguesa: Uma Adaptação do Modelo *Ontology Metadata Vocabulary*", apresentada por Anderson Bestteti Santos, como parte dos requisitos para obtenção do grau de Mestre em Ciência da Computação, Inteligência Computacional, aprovada em 15/03/10 pela Comissão Examinadora:

Profa. Dra. Renata Vieira –
Orientadora

PPGCC/PUCRS

Prof. Dr. Marcelo Blois Ribeiro –

PPGCC/PUCRS

Profa. Dra. Maria Luiza Machado Campos –

UFRJ

Homologada em...18.../05.../10..., conforme Ata No. 08/10 pela Comissão Coordenadora.

Prof. Dr. Fernando Gehm Moraes
Coordenador.

PUCRS

Campus Central

Av. Ipiranga, 6681 – P32- sala 507 – CEP: 90619-900

Fone: (51) 3320-3611 – Fax (51) 3320-3621

E-mail: ppgcc@pucrs.br

www.pucrs.br/facin/pos

DEDICATÓRIA

Dedico este trabalho a minha esposa Patrícia e a minha filha Lívia.

“Aqueles que sabem tudo, mas desconhecem a si próprios, são absolutamente carentes.”

Jesus Cristo

AGRADECIMENTOS

A minha esposa Patrícia por todo amor, carinho e atenção que me são oferecidos durante anos de namoro, noivado e casamento. Posso afirmar, sem sombra de dúvida, que você sempre contribuiu para o meu crescimento como pessoa sobre diversos aspectos. Também devo agradecer-te por toda ajuda que me destes durante os meus anos de estudo.

A nossa filha Lívia, eterna alegria em nossas vidas, que eu possa lhe acompanhar e educar durante a sua caminhada pela vida. Rogo a Deus que Ele possa sempre protegê-la.

A minha mãe Marli por todo o esforço e dedicação com a minha educação, mas principalmente pelo exemplo de perseverança, honestidade e humildade. Devo à senhora todas as minhas conquistas. Meu eterno obrigado.

Ao meu irmão Jeferson pelo companheirismo, pelo apoio em momentos de incerteza e pelas horas de descontração.

A Prof^a Dra^a Renata Vieira por me dar a oportunidade entrar no programa de Pós-Graduação em Ciência da Computação da PUCRS, bem como incentivo, compreensão e dedicação fundamentais à realização deste trabalho.

Ao Prof. Dr. Marcelo Blois Ribeiro pelas importantes contribuições e sugestões para o desenvolvimento deste trabalho.

Aos colegas do laboratório de Processamento da Língua Natural pelo companheirismo e horas de estudos em conjunto. Aos amigos André L. Silva e Luciano Blomberg, que tanto me apoiaram nos momentos mais críticos desta jornada chamada mestrado.

Por fim, agradeço a Ação Sistemas de Informática Ltda, em especial, ao senhor Luis Carlos de Moraes Leite pela compreensão e apoio para que pudesse ir adiante com os meus estudos.

DESCRIÇÃO E RECUPERAÇÃO DE ONTOLOGIAS EM LÍNGUA PORTUGUESA: UMA ADAPTAÇÃO DO MODELO *ONTOLOGY METADATA VOCABULARY*

RESUMO

Este trabalho apresenta a adaptação de um modelo de metadados de descrição de ontologias. Essa adaptação tem por objetivo enriquecer o modelo selecionado com novos dados para descrição de ontologias, oferecendo aos pesquisadores da área uma gama maior de informações para auxiliá-los no processo de identificação e seleção de ontologias em língua portuguesa. O modelo resultante da adaptação é utilizado em uma aplicação, cujo objetivo é oferecer novos serviços ao Portal OntoLP para descrição e recuperação de ontologias. Para avaliação do protótipo proposto, foi realizada uma *survey* onde os participantes foram questionados sobre os serviços de busca de ontologias do Portal.

Palavras Chave: Ontologia, Metadados, Portal OntoLP, OMV

DESCRIÇÃO E RECUPERAÇÃO DE ONTOLOGIAS EM LÍNGUA PORTUGUESA: UMA ADAPTAÇÃO DO MODELO *ONTOLOGY METADATA VOCABULARY*

ABSTRACT

This work presents an adaptation of a metadata model to describe ontologies. This adaptation aims to enrich the model with new data to describe ontologies, offering to researchers a wider range of information to help them to find out and select ontologies. The resulting model is used on an application whose goal is to support description and retrieval of ontologies from the ontology portal OntoLP. To assess the proposed prototype, a survey was conducted, where the participants were asked about the new search functions of the OntoLP.

Keywords: Ontology, Metadata, Portal OntoLP, OMV

LISTA DE FIGURAS

Figura 1 - Exemplo de uma taxonomia de veículo.....	18
Figura 2 - Exemplo de uma hierarquia "é-um".....	21
Figura 3 - Diagrama da ontologia OMV Core.....	27
Figura 4 - Metodologia de Desenvolvimento de Sistemas.....	34
Figura 5 - Metadados OMVR. Adaptação do OMV Core.....	42
Figura 6 - Página inicial do Portal OntoLP.....	48
Figura 7 - Fluxo de armazenamento de uma ontologia no Portal OntoLP.....	49
Figura 8 - Estrutura do OMVR API.....	51
Figura 9 - Modelo UML do OMVR API.....	54
Figura 10 - Interface de submissão de ontologias.....	55
Figura 11 - Interface de exploração de ontologias.....	56
Figura 12 – Interface inserção de avaliação de ontologias.....	57
Figura 13 - Interface visualização das avaliações de ontologias.....	57
Figura 14 - Interface de visualização das métricas calculadas para a ontologia.....	58
Figura 15 - Exemplo de visualização da documentação sobre a métrica selecionada.....	59
Figura 16 - Visualização do gráfico das métricas calculadas para a ontologia.....	59
Figura 17 - Interface de pesquisa avançada de ontologias.....	60
Figura 18 - Gráfico de distribuição de respondentes.....	61
Figura 19 - Número de respondentes por universidade.....	62
Figura 20 - Percentual de respondentes por educação formal.....	62
Figura 21 - Avaliação dos novos serviços de submissão e busca do Portal OntoLP.....	63
Figura 22 - Avaliação das novas funções em relação aos serviços vigentes do portal OntoLP.....	64
Figura 23 – Sugestões sobre a inclusão de novas informações no OMVR.....	65

Figura 24 - Ranking das informações mais úteis para localização de ontologias, segundos os respondentes.66

LISTA DE SIGLAS

ACM	-	<i>Association of Computing Machinery</i>
API	-	<i>Application Programming Interface</i>
DAML	-	<i>DARPA Agent Markup Language</i>
DARPA	-	<i>Defense Advanced Research Projects Agency</i>
DL	-	<i>Description Logic</i>
IEEE	-	<i>Institute of Electrical and Electronics Engineers</i>
HTML	-	<i>HyperText Markup Language</i>
NAICS	-	<i>North American Industry Classification System</i>
NETGLOS	-	<i>The Multilingual Glossary of Internet Terminology</i>
P2P	-	<i>Peer-to-Peer</i>
OIL	-	<i>Ontology Interchange Language</i>
OMV	-	<i>Ontology Metadata Vocabulary</i>
OWL	-	<i>Web Ontology Language</i>
RDF	-	<i>Resource Description Framework</i>
ROMEO	-	<i>Requirements-Oriented Methodology for Evaluation Ontologies</i>
SPARQL	-	<i>SPARQL Protocol and RDF Query Language</i>
SUO	-	<i>Standard Upper Ontology</i>
UML	-	<i>Unified Modeling Language</i>
URI	-	<i>Uniform Resource Identifier</i>
W3C	-	<i>World Wide Web Consortium</i>
XML	-	<i>Extensible Markup Language</i>

SUMÁRIO

1	INTRODUÇÃO	12
1.1	MOTIVAÇÃO DO TRABALHO	13
1.2	OBJETIVOS	14
1.3	ORGANIZAÇÃO DESTE DOCUMENTO	15
2	ONTOLOGIAS	16
2.1	CLASSIFICAÇÕES DE ONTOLOGIAS	16
2.1.1	<i>Taxonomias versus Ontologias</i>	18
2.1.2	<i>Tesauros versus Ontologias</i>	18
2.1.3	<i>Tipos de Informações Representadas pelas Ontologias</i>	19
2.2	ONTOLOGIAS E A <i>WEB SEMÂNTICA</i>	20
2.3	A LINGUAGEM OWL	22
3	TRABALHOS RELACIONADOS	24
3.1	O MODELO DE METADADOS OMV	24
3.2	DESCRIÇÃO DO MODELO OMV	25
3.3	PADRONIZAÇÃO DA DESCRIÇÃO DE ONTOLOGIAS	28
3.4	A PROPOSTA DE BONTAS E CO-AUTORES	28
3.5	EXEMPLOS DE APLICAÇÕES DO MODELO OMV	29
3.6	O PORTAL <i>SWOOGLE</i>	29
3.7	A FERRAMENTA <i>WEBCORE</i>	30
3.8	ANÁLISE DE NATALYA F. NOY E CO-AUTORES	31
4	MÉTODO	32
4.1	DESIGN E CRIAÇÃO DE FERRAMENTA DE TI	33
4.2	SURVEY	35
4.3	ANÁLISE DOS DADOS	38
5	PROPOSTA DE ADAPTAÇÃO DO MODELO OMV	40
5.1	DESCRIÇÃO DAS CLASSES E PROPRIEDADES DO MODELO OMVR	43
5.1.1	<i>Classe OntologyEvaluation</i>	43
5.1.2	<i>Classe EvaluationCriteria</i>	43
5.1.3	<i>Classe Metric</i>	44

5.1.4 Classe <i>OntologyMetric</i>	45
5.1.5 Classe <i>Project</i>	46
5.2 DESCRIÇÕES DAS RELAÇÕES DO MODELO OMVR	46
6 PROTÓTIPO DA FERRAMENTA DE DESCRIÇÃO E RECUPERAÇÃO DE ONTOLOGIAS.....	48
6.1 O PORTAL ONTOLP.....	48
6.2 ESPECIFICAÇÃO DAS FERRAMENTAS DE MANIPULAÇÃO DO MODELO OMVR	49
6.2.1 <i>Organização dos componentes de infra-estrutura</i>	49
6.2.2 <i>Especificação da Interface de Navegação sobre o Modelo OMVR</i>	54
6.3 AVALIAÇÃO DE USO DO PROTÓTIPO	61
6.3.1 <i>Perfil dos entrevistados</i>	61
6.3.2 <i>Resultados da Survey</i>	63
7 CONSIDERAÇÕES FINAIS E TRABALHOS FUTUROS	68
REFERENCIAL BIBLIOGRÁFICO.....	71
APÊNDICE A – PÁGINA DE CONTATO DO PORTAL ONTOLP	75
APÊNDICE B – QUESTIONÁRIO	76
APÊNDICE C – PÁGINA DE SUBMISSÃO DE ONTOLOGIAS DO PORTAL ONTOLP-OMVR.....	77

1 INTRODUÇÃO

Na área de Computação, ontologias são definidas como especificações formais e explícitas de conceitos, as quais capturam o vocabulário utilizado nas aplicações de um sistema de informação. Além disso, elas proporcionam um bloco de construção para expressar a semântica de um determinado conceito sem apresentar ambigüidade. O W3C define que ontologias devem propiciar descrições dos conceitos de classes, relacionamentos entre classes e as propriedades que essas classes devem possuir.

Nesse contexto, uma das principais características das ontologias é permitir a compreensão das informações através das formalizações semânticas tanto por humanos quanto por máquinas, facilitando o compartilhamento e reuso dessas informações. Gruber [GRU08] define uma ontologia como: “uma especificação formal e explícita de uma conceitualização compartilhada”. A conceitualização representa um modelo abstrato de algum fenômeno, identificando os conceitos relevantes para o mesmo. Formal significa que a ontologia deve ser passível de processamento automático. Explícita quer dizer que as suas restrições estão claramente definidas. Finalmente, compartilhada quer dizer que uma ontologia captura um conhecimento aceito por um grupo de pessoas.

Nos últimos anos, as ontologias passaram a ser amplamente utilizadas em várias áreas da tecnologia da informação, tais como inteligência artificial, banco de dados, engenharia de software e em diversos domínios do conhecimento como, por exemplo, biologia, medicina e direito.

As principais promessas do projeto da *Web Semântica* são permitir o processamento computacional de recursos da *Web* por meio de artefatos de conhecimento como, por exemplo, ontologias, e organizar uma grande quantidade de informações esparsas [GIB07]. Em 1999 Tim Berners-Lee apresenta a sua visão da *Web Semântica* [BER01]:

"I have a dream for the Web in which computers become capable of analyzing all the data on the Web – the content, links, and transactions between people and computers. A 'Semantic Web', which should make this possible, has yet to emerge, but when it does, the day-to-day mechanisms of trade, bureaucracy and our daily lives will be handled by machines talking to machines. The 'intelligent agents' people have touted for ages will finally materialize."

O projeto da *Web Semântica* colaborou para a popularização das ontologias como artefato de representação do conhecimento. Nesse contexto, o reuso é um aspecto importante a ser considerado ao se construir uma nova ontologia. O compartilhamento e

reúso de ontologias existentes ajudam a assegurar a qualidade dos recursos, uma vez que elas tenham sido utilizadas, acessadas e revisadas por muitas pessoas. De forma análoga à disciplina de Engenharia de *Software*, o reúso eficiente de ontologias reduz o custo relacionado com o seu desenvolvimento, pois evita a reconstrução de componentes já disponíveis na *Web*, podendo ser integrados a uma nova ontologia (em alguns casos com pequenas adaptações). Ainda, acessar ontologias disponíveis através da *Web* é um requisito fundamental para a disseminação de serviços *Web* semânticos, os quais pretendem permitir o uso automático de recursos e descrever seus serviços e processos de intercomunicação, através de referências a essas ontologias.

A carência de uma documentação abrangente sobre as ontologias ou a má qualidade da documentação existente, muitas vezes, corrobora para uma difícil compreensão da ontologia por usuários que pretendem utilizá-la. Frequentemente os repositórios existentes oferecem interfaces limitadas no que diz respeito à navegação e à localização de ontologias [DIN04]. Além disso, a função de recuperação, oferecida por alguns portais de ontologias, é limitada pela ausência de um modelo de metadados sobre as ontologias [HAR06].

Uma possível abordagem para fazer frente ao problema apontado é o uso de um vocabulário e anotação padronizada de ontologias, com informações provenientes de um modelo de metadados. Esses são os pré-requisitos básicos para uma boa descrição de ontologias na *Web*. Dessa forma, o modelo de metadados tem grande importância na construção de ferramentas especificamente projetadas para descrição e recuperação de ontologias, proporcionando, entre outros benefícios, o reúso em larga escala desses artefatos.

1.1 Motivação do trabalho

Conforme a discussão apresentada percebe-se que há necessidade de criar artefatos que promovam a descrição de ontologias com base em um modelo de metadados bem definido e, preferencialmente, que o mesmo já tenha passado por um processo de validação e verificação. Dessa forma, espera-se que um modelo de descrição já deva estar maduro o suficiente para ser utilizado como ponto de partida para o desenvolvimento de novas aplicações de recuperação, descrição e compartilhamento de ontologias.

Iván Cantador *et al.* [CAN07] enfatizam que o desenvolvimento de novas ontologias é um processo complexo, requerendo o trabalho conjunto de dois profissionais: (i) um engenheiro de ontologias, que é responsável por criar um modelo correto e consistente durante o desenvolvimento; e (ii) um especialista do domínio do conhecimento que se deseja modelar, o qual é responsável por capturar a semântica de um campo específico dentro da ontologia. Isto posto, fica evidente que o reúso de ontologias é um objetivo a ser alcançado, a fim de minimizar o esforço de desenvolvimento e atender às necessidades e requisitos especiais dos engenheiros e usuários de ontologias.

A utilização de um modelo de metadados consolidado para descrição e recuperação de ontologias tem uma importância significativa no sentido de promover o reúso de ontologias. Um dos benefícios que pode ser ressaltado é que o emprego de um modelo de metadados possibilita aos engenheiros de ontologias centralizarem a descrição das mesmas em um ambiente organizado e controlado, facilitando a localização dessas ontologias. Além disso, um modelo de metadados também pode ser usado como um artefato para construção de sistemas de armazenamento, descrição e recuperação de ontologias. Embora existam iniciativas de construção de modelos de metadados para descrição de ontologias, percebe-se que ainda há oportunidade para enriquecer o conjunto de informações atualmente disponível. Por exemplo, dados relativos ao reúso de ontologias como informações sobre versões, grupos que as utilizaram, avaliações por especialistas, entre outras, não são tratados nos modelos de metadados existentes. Por esse motivo, este trabalho tem por objetivo propor a extensão do modelo de metadados OMV para contemplar novas informações que possam auxiliar usuários e pesquisadores da área a localizarem e compartilharem ontologias. Ainda, no contexto desta pesquisa, o modelo resultante do trabalho poderá contribuir na extensão dos serviços do Portal OntoLP¹.

1.2 Objetivos

O objetivo deste trabalho é propor uma adaptação do modelo de metadados OMV para suportar novos elementos para descrição de ontologias, com o objetivo de auxiliar especialistas da área a recuperarem ontologias em língua portuguesa, permitindo a incorporação das mesmas no Portal OntoLP. Para tanto, este trabalho considera as

¹ <http://www.inf.pucrs.br/~ontolp/> - Edital MCT/CNPq/CT-INFO 07/2007

propostas apresentadas em [BON05], [DIN04], [HAR05], [HAR06], [OAT06], [PAL06], [SIM09], [YU09], [TAR07], [GAN06a], [YAN06], [FRE09] e [PAL08] para embasar o desenvolvimento da extensão do OMV. Cabe salientar que, a partir da extensão do modelo, será desenvolvido um protótipo de aplicação cujo objetivo é permitir capturar informações sobre ontologias e apresentar as informações do novo modelo, através de mecanismos de consultas simples e avançadas. Por fim, são avaliados o novo modelo e o protótipo da aplicação.

1.3 Organização deste documento

Este trabalho está organizado como segue. No Capítulo 2 são apresentados os conceitos básicos sobre ontologias, suas classificações e relação com a *Web Semântica*, finalizando com a linguagem de implementação de ontologias OWL. No capítulo 3 são apresentados os trabalhos relacionados com essa pesquisa, os quais servirão de apoio para o desenvolvimento da extensão do modelo de metadados OMV. No Capítulo 4 apresenta o método de pesquisa utilizado para condução deste trabalho, o qual tem por objetivo produzir um protótipo de software que será avaliado conforme metodologia sugerida em [OAT06]. No Capítulo 5 é apresentada a proposta de adaptação de um modelo de metadados sobre ontologias, objeto de estudo deste trabalho. No Capítulo 6 apresenta o Portal OntoLP e o serviço de compartilhamento de recursos, bem como a especificação e implementação do protótipo do novo portal, onde são detalhados os serviços de descrição e busca de ontologias. Além disso, discute a análise dos resultados obtidos da pesquisa de comparação entre as versões do Portal OntoLP. Por fim, no Capítulo 7, são apresentadas as considerações finais e trabalhos futuros.

2 ONTOLOGIAS

A fim de detalhar o conceito sobre ontologias, dentro da área da Ciência da Computação, este capítulo descreverá diferentes classificações ligadas às ontologias e também apresentará, brevemente, a linguagem OWL para especificação de ontologias.

Ontologias têm sido empregadas para descrever artefatos com vários graus de estruturação e para propósitos variados, independentes da forma que são descritas pelos arquitetos. Elas podem ser generalizações simples ou formar metadados complexos, como, por exemplo, o Dublin *Core*², ou formarem modelos descritos em linguagem lógica.

O consórcio W3C define que ontologias devem prover descrições dos seguintes tipos de conceitos:

- a) Classes nos vários domínios de interesse;
- b) Relacionamento entre classes, e;
- c) Propriedades que essas classes devem possuir.

Por sua vez, Maedche e Staab [MAE01] propõem uma descrição de ontologias composta de cinco elementos primitivos:

- a) Conceitos;
- b) Relacionamentos;
- c) Hierarquia de conceitos;
- d) Função que relaciona conceitos, e;
- e) Conjunto de axiomas.

2.1 Classificações de Ontologias

As ontologias são classificadas de acordo com os aspectos semânticos, hierárquicos, de gênero, tipo de informações e relacionamentos semânticos [BRE05], a saber:

² <http://dublincore.org/>

- a) **Vocabulários controlados:** lista finita de termos, exemplo: catálogo norte-americano (NAICS) de produtos e serviços oferecidos em diversas áreas, como agricultura, finanças, varejo, etc.
- b) **Glossários:** lista de termos em linguagem natural, exemplo: o NetGlos que reúne terminologia relacionada a recursos da Internet.
- c) **Tesauros:** lista de termos com suas definições, que padroniza a utilização de palavras para indexação. Além disso, um tesauro fornece relacionamentos entre termos, podendo ser do tipo hierárquico, associativo ou de equivalência.
- d) **Hierarquia “tipo-de” informais:** hierarquias que utilizam relacionamento de generalização (tipo-de) de maneira informal. Nesse tipo de hierarquia, conceitos relacionados podem ser agregados a uma categoria, mesmo não respeitando integralmente o relacionamento de generalização, como, por exemplo, a classe “viagem” poderia ser usada para classificar os termos “aluguel de carro” e “hotel”.
- e) **Hierarquias “tipo-de” formais:** hierarquias que incluem instâncias de um domínio. Nessas hierarquias os relacionamentos de generalização são completamente respeitados.
- f) **Frames:** modelos que incluem classes e propriedades, segundo a representação proposta por Minsky [MIN74]. As primitivas desse modelo são as classes e respectivos atributos. Atributos não têm escopo global e aplicam-se apenas nas classes para as quais foram definidas. Cada *frame* fornece o contexto para modelar um aspecto do domínio.
- g) **Ontologias que exprimem restrições de valores:** fornecem subsídios para restringir os valores assumidos pelas propriedades de suas classes. Um exemplo típico é a restrição que se impõe a uma informação do tipo data, onde o valor de atribuição deve ser numérico.
- h) **Ontologias que exprimem restrições lógicas:** ontologias que permitem que sejam expressas restrições em lógica de primeira ordem.

2.1.1 Taxonomias *versus* Ontologias

Segundo Breitman [BRE05], uma taxonomia é a “classificação de entidades de informação no formato de uma hierarquia, de acordo com relacionamentos que estabelecem com entidades do mundo real que representam”. Os termos de uma taxonomia são relacionados através de generalizações, por exemplo: um mamífero é um vertebrado, que é um cordado, que é um tipo de animal. Outro exemplo de taxonomia é uma árvore de diretórios em um determinado sistema de arquivos. A Figura 1 representa graficamente uma taxonomia de veículo.

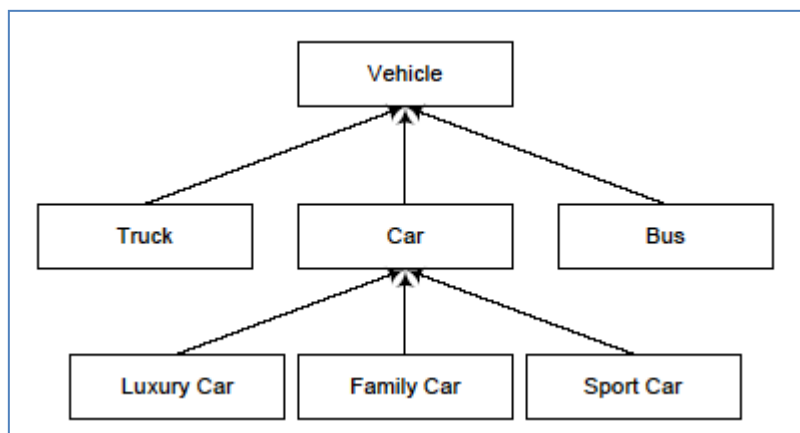


Figura 1 - Exemplo de uma taxonomia de veículo.

Fonte [GAN06b]

2.1.2 Tesouros *versus* Ontologias

Um tesouro reúne um conjunto de relacionamentos entre termos que estão organizados em uma taxonomia. É definido como uma taxonomia adicionada de um conjunto de relacionamentos semânticos entre seus termos. Tesouros são utilizados por indexadores para atribuir uma terminologia a vários repositórios de dados.

O objetivo de um tesouro é garantir que conceitos sejam descritos de forma consistente, de modo a permitir que usuários possam refinar buscas e localizar a informação desejada. Para facilitar esse processo, tesouros contam com relacionamentos extras que auxiliam na organização da informação. É importante observar que em um tesouro os tipos de relacionamentos entre seus termos são finitos e bem definidos. Isso é útil para a criação de vocabulários, mas insuficiente para modelar todos os aspectos do mundo real. Por esse motivo, um tesouro não permite ao usuário a criação de novos tipos

de relacionamentos, por exemplo: parte-de, fase-processo, lugar-região, material-objeto, causa-efeito, entre outros. Logo, faz-se necessário usar uma ontologia.

A maior expressividade de uma ontologia em relação às taxonomias e aos tesauros advém de três propriedades essenciais:

- a) **Estrita hierarquia de sub-conceitos:** toda a instância de uma classe tem de ser uma instância do nodo pai (hierarquia tipo-de formal). A estrutura mestre da ontologia é organização dos termos segundo o relacionamento tipo-de (generalização).
- b) **Interpretação livre de ambigüidade para os significados e relacionamentos:** as propriedades de cada nodo podem ser definidas pelos usuários, tendo valores restritos por um leque de valores também definidos pelos usuários. Ainda, ontologias podem possuir relacionamentos mais sofisticados como disjunção e parte-de.
- c) **Extensível:** ontologias são definidas com um vocabulário controlado, mas permitindo estendê-lo.

2.1.3 Tipos de Informações Representadas pelas Ontologias

Outra classificação de ontologias refere-se ao tipo de informação a ser modelado [IEE08]:

- a) **Ontologias para representação do conhecimento:** capturam primitivas de representação do conhecimento, fornecendo o apoio para modelagem de linguagens baseadas em frames, tais como: classes, subclasses, atributos, valores e axiomas.
- b) **Ontologias gerais de uso comum:** representam conhecimento de senso comum, podendo ser utilizadas em vários domínios. Essas ontologias incluem um vocabulário relacionado a classes, eventos, espaço, causalidade, comportamento, entre outros.
- c) **Ontologias de topo (*upper ontologies*):** descrevem conceitos gerais, podendo ser utilizadas em uma ampla variedade de domínios [IEE08]. A fim de padronizar ontologias desse tipo, o IEEE propôs a criação de um grupo de trabalho focado em ontologia de topo padrão, o SUO.

- d) **Ontologias de domínio:** podem reutilizar conceitos dentro de um domínio específico, exemplo: médico, administrativo, financeiro, engenharia, entre outros. Os termos e as propriedades de uma ontologia de domínio são obtidos através da especialização de conceitos de uma ontologia de topo.
- e) **Ontologias de tarefas:** descrevem o vocabulário ligado a uma tarefa específica.
- f) **Ontologias de domínio-tarefa:** são ontologias que podem ser utilizadas em um dado domínio, porém não em domínio similares.
- g) **Ontologias de métodos:** fornecem definições para os conceitos e relacionamentos relevantes para um processo de modo a atingir um objetivo.
- h) **Ontologias de aplicação:** contêm a informação suficiente para modelar o conhecimento necessário a uma aplicação qualquer. Esse tipo de ontologia é utilizado para especializar e estender ontologias de domínio ou tarefa para uma dada aplicação.

2.2 Ontologias e a Web Semântica

A principal razão do recente interesse por ontologias é o desenvolvimento da *Web* semântica, a qual pode ser vista como o gerenciamento do conhecimento em escala global [FEN06]. Para Tim Berners-Lee [BER01], a *Web* Semântica será a próxima geração da *Web* atual, pois permitirá que as máquinas façam melhor uso das informações, além de permitir a automação de serviços. Através da representação semântica explícita dos dados, programas, páginas e outros recursos da *Web*, será possível criar uma *Web* baseada em conhecimento, gerando uma nova gama de serviços. Ontologias provêm essa representação semântica explícita, as quais combinadas com a *Web* potencializam a superação de muitos dos problemas de compartilhamento e de reúso do conhecimento, assim como a integração de informações.

Um dos principais aspectos das ontologias é que através da formação da semântica do mundo real e terminologias consensuais, elas permitem que as informações sejam tratáveis tanto por humanos, quanto por máquinas [FEN06]. Essa propriedade facilita o compartilhamento e o reúso de ontologias entre essas duas entidades.

O significado para um humano é representado pelo próprio termo, o qual normalmente é uma palavra em linguagem natural, e através do relacionamento semântico entre termos. Um exemplo é o entendimento humano sobre a relação superconceito-subconceito (normalmente referenciada pelo termo “é-um”). Esse relacionamento denota o fato que um conceito (o superconceito) é mais genérico que outro (o subconceito) [FEN06]. Por exemplo, o conceito Pessoa é mais genérico que Estudante. Logo, conceito descreve um conjunto de objetos do mundo real. A Figura 2 ilustra os relacionamentos do termo “é-um”.

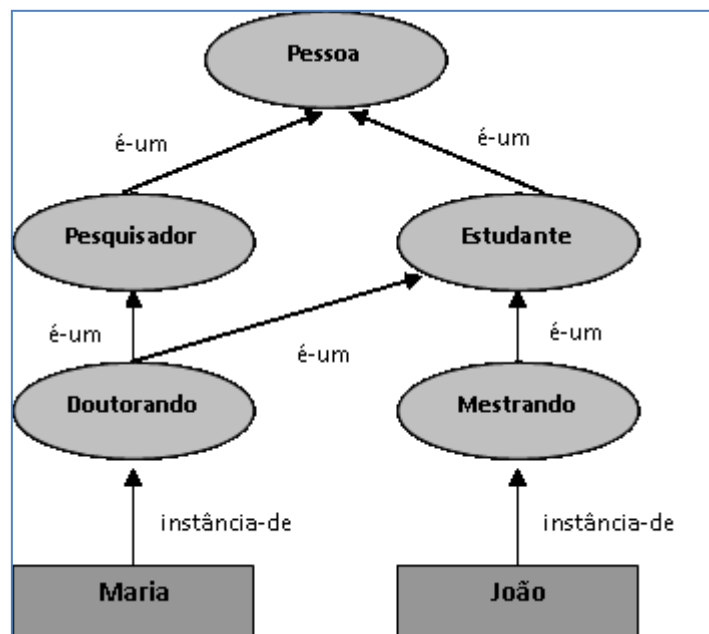


Figura 2 - Exemplo de uma hierarquia "é-um".

Fonte: Adaptado de [FEN06]

Essa relação é facilmente compreendida por um leitor humano, porque o significado da relação está formalmente definido. Assim como uma máquina pode, em algumas situações específicas, fazer um raciocínio sobre essa relação e chegar à mesma conclusão de um humano. Essa relação, implicitamente conhecida por um humano (por exemplo, um humano sabe que todo estudante é uma pessoa), é formalmente codificada em uma estrutura que também possa ser entendida por uma máquina. De certa forma a máquina não tem o real entendimento, mas a compreensão dos seres humanos é codificada de tal forma que uma máquina pode processá-las e chegar a conclusões através do raciocínio lógico.

2.3 A Linguagem OWL

A OWL é uma linguagem de marcação semântica para compartilhar e publicar ontologias na *Web*, desenvolvida como uma extensão do vocabulário de RDF e é derivada das linguagens DAML + OIL [BEC08], incorporando lições aprendidas a partir da concepção e aplicação de DAML + OIL. É designada para ser usada em aplicações que precisam processar o conteúdo de informações em páginas da *Web*, indo além da simples apresentação de informações para humanos [MCG08]. Também facilita sobremaneira a interpretação do conteúdo da *Web* por máquinas, superando o suporte dado por XML, RDF e RDF Schema (RDF-S) e provendo um vocabulário adicional junto com uma semântica formal [MCG08].

A OWL possui três linguagens, cada uma delas com um nível de expressividade maior, a saber [MCG08]:

- a) **OWL *Lite***: propicia aos usuários a criação de hierarquias e restrições simplificadas. OWL *Lite* tem uma complexidade formal inferior a OWL DL;
- b) **OWL DL**: gera suporte aos usuários que desejam o máximo em termo de expressividade, mantendo um baixo custo computacional devido à garantia de conclusões computáveis e decidíveis, finalizadas em um tempo finito. OWL DL inclui todas as estruturas de OWL, as quais só podem ser usadas sob certas restrições. Por exemplo, enquanto uma classe pode ser subclasse de múltiplas classes, uma classe não pode ser uma instância de outra classe; e,
- c) **OWL *Full***: busca garantir aos usuários máxima expressividade e liberdade sintática através de RDF, mas sem garantias computacionais. Por exemplo, em OWL *Full* uma classe pode ser tratada simultaneamente como um conjunto de indivíduos, assim como um único indivíduo. Outra de suas funcionalidades é permitir que uma ontologia aumente o significado do vocabulário pré-definido em RDF ou OWL.

Os desenvolvedores de ontologias que adotam a OWL devem escolher, entre as opções de linguagens, qual é a mais adequada às suas necessidades. A escolha entre OWL *Lite* e OWL DL depende do grau de expressividade que o usuário necessita, ou seja, quando a maior expressividade requerida pela OWL DL passa a ser a opção natural. A escolha entre OWL DL e OWL *Full* depende principalmente do grau de facilidade

exigido pelo usuário para modelar um metamodelo através de RDF *Schema*, por exemplo, definindo classes de classes, ou propriedades inerente às classes.

Conforme o consórcio W3C [MCG08], o OWL *Full* pode ser visto como uma extensão de RDF. Enquanto que OWL *Lite* e OWL DL podem ser vistos como extensões de uma visão restrita de RDF. Todo documento OWL (*Lite*, DL e *Full*) é um documento RDF. Porém, somente alguns documentos RDF serão documentos OWL *Lite* ou OWL DL válidos. Em função disso, alguns cuidados devem ser tomados quando um usuário deseja migrar um documento RDF para OWL. Quando a expressividade de OWL DL ou OWL *Lite* for considerada apropriada, algumas precauções devem ser tomadas para garantir que o documento RDF original esteja em conformidade com as restrições impostas por OWL DL e OWL *Lite*. Entre outras restrições, cada URI, usado como um nome de classe, deve ser afirmado expressamente como sendo do tipo *owl:Class* (o mesmo deve ocorrer com as propriedades) e cada indivíduo deve pertencer a pelo menos uma classe (ainda que apenas a *owl:Thing*).

Uma maior descrição de OWL pode ser obtida no site da W3C, ver referências [MCG08] e [BEC08], pois não é o escopo deste trabalho abordar em profundidade as características e estruturas de OWL.

Este capítulo procurou apresentar algumas definições sobre ontologias, sob a ótica da Ciência da Computação. Também foi apresentado como as ontologias são classificadas, a proposta da *Web Semântica* e a sua relação com ontologias. Por fim, foi apresentada a linguagem OWL como recurso para estruturação do conhecimento de forma declarativa. As próximas seções desse trabalho apresentarão um modelo de metadados de descrição de ontologias e uma proposta de adaptação desse modelo, além de um protótipo de aplicação, baseado no modelo adaptado, e, por fim, a avaliação da aplicação implementada.

3 TRABALHOS RELACIONADOS

Neste capítulo, serão apresentados alguns trabalhos pesquisados sobre o tema recuperação, descrição e reuso de ontologias.

3.1 O modelo de metadados OMV

Antes de construir qualquer aplicação para gerenciamento de ontologias, é necessário definir uma estrutura que possa atender aos requisitos de padronização da descrição das mesmas. Neste sentido foi criado o projeto OMV [PAL08], que tem por objetivo especificar um modelo de metadados que possa abranger um largo espectro de propriedades sobre ontologias, colaborando para sua descrição.

O OMV foi proposto pelo consórcio formado por *University of Bremen* (TZI), *University of Karlsruhe* (UKARL), *Universidad Politécnica de Madrid* (UPM) e *Stanford Center for Biomedical Informatics Research* (BMIR). Esse consórcio também é responsável pela manutenção do OMV. Além disso, o grupo também definiu regras para construção e validação da ontologia, com o objetivo de oferecer uma ontologia com um bom nível de estabilidade já em suas primeiras versões. A definição na íntegra do OMV está disponível no *site* do projeto³.

O modelo de metadados OMV foi identificado como um forte candidato a atender às necessidades de descrição, recuperação e reuso de ontologias. O principal benefício do OMV refere-se às técnicas de validação e verificação utilizadas durante o seu desenvolvimento, conferindo-lhe consistência e abrangência na descrição de ontologias [PAL06] [PAL08], além de permitir que o modelo seja estendido por outros grupos de pesquisa. Esse nível de maturidade e flexibilidade de extensão é desejável para a construção de aplicações que se destinam a gerenciar ontologias.

³ <http://ontoware.org/projects/omv/>

3.1.1 Descrição do modelo OMV

O modelo OMV *Core* pretende transformar-se em um esquema de ampla aceitação para descrição de ontologias para a *Web Semântica*. Nesse sentido, o OMV concentra-se na conceitualização de ontologias para representação das suas principais informações [HAR06]. A seguir serão descritas, brevemente, as classes que compõem o modelo de metadados OMV *Core* e, logo a seguir, o diagrama do OMV é apresentado na Figura 3.

- a) **Ontology**: classe principal que identifica o recurso sendo descrito, que é a implementação de um modelo conceitual, ou seja, é o nome dado ao conjunto de classes, relações e instâncias que formam uma ontologia;
- b) **OntologyType**: esta classe agrupa tipos de ontologias de acordo com a sua classificação. O OMV oferece algumas instâncias de classificação de ontologias encontradas na literatura, tais como ontologias de topo, domínio, aplicação e tarefa;
- c) **LicenseModel**: identifica o modelo de licença e as condições de usos da ontologia. O modelo OMV já oferece algumas instâncias de modelos de licenças, como por exemplo, a AFL, CPL, LGPL, GPL, entre outras;
- d) **OntologyEngineeringMethodology**: esta classe representa informações sobre metodologias de engenharia das ontologias. O modelo OMV já oferece algumas instâncias de metodologias de construção de ontologia, como por exemplo, a KACTUS, METHONLOGY, entre outras;
- e) **OntologyEngineeringTool**: representa as ferramentas utilizadas para construção das ontologias. Esta classe oferece algumas instâncias de ferramentas, tais como Protégé⁴, SWOOP⁵, OntoStudio⁶, OilEd⁷, entre outras;
- f) **OntologySyntax**: representa informações da sintaxe usada pelas ontologias. Esta classe já possui as instâncias dos padrões de sintaxe OWL-XML e RDF/XML;

⁴ <http://protege.stanford.edu/>

⁵ <http://code.google.com/p/swoop/>

⁶ <http://semanticWeb.org/wiki/OntoStudio>

⁷ <http://xml.coverpages.org/oilEdANn20001204.html>

- g) **OntologyLanguage**: esta classe já possui as instâncias das linguagens OWL⁸, OWL-DL, OWL-Lite, OWL-Full, DAML-OIL⁹ e RDF-S. Ainda, ela permite criar novas instâncias de linguagens de implementação de ontologias;
- h) **KnowledgeRepresentationParadigm**: são informações sobre paradigmas de representação do conhecimento aderida por uma linguagem em particular. Esta classe já possui as instâncias dos paradigmas de representação DL e *Frames*;
- i) **FormalityLevel**: representa o nível de formalidade de uma ontologia. Esta classe já possui as instâncias dos níveis de formalidade *catalog*, *glossary*, *thesauri*, *taxonomy*, *frames* e *properties*, *value restrictions*, *disjointness* e *general logic constraints*;
- j) **OntologyTask**: são informações sobre para que tarefa uma ontologia se destina. Uma descrição detalhada das tarefas é encontrada em [PAL08];
- k) **OntologyDomain**: enquanto o domínio pode se referir a qualquer tema em ontologias, é aconselhável que sempre seja especificado uma hierarquia de domínio de propósito genérico como o DMOZ¹⁰, ou um assunto em uma hierarquia de domínio específico como ACM para Ciência da Computação. Somente dessa forma pode-se assegurar o correto significado que se pode deduzir ao se comparar diferentes ontologias;
- l) **Party**: esta classe tanto pode representar informações de uma pessoa, quanto de uma organização relacionada ao desenvolvimento da ontologia;
- m) **Person**: representa uma pessoa;
- n) **Organization**: representa algum tipo de organização, e;
- o) **Location**: representa a localização geográfica de uma pessoa ou organização.

⁸ <http://www.w3.org/TR/owl-ref/>

⁹ <http://www.w3.org/TR/dam1+oil-reference>

¹⁰ <http://www.dmoz.org/about.html>

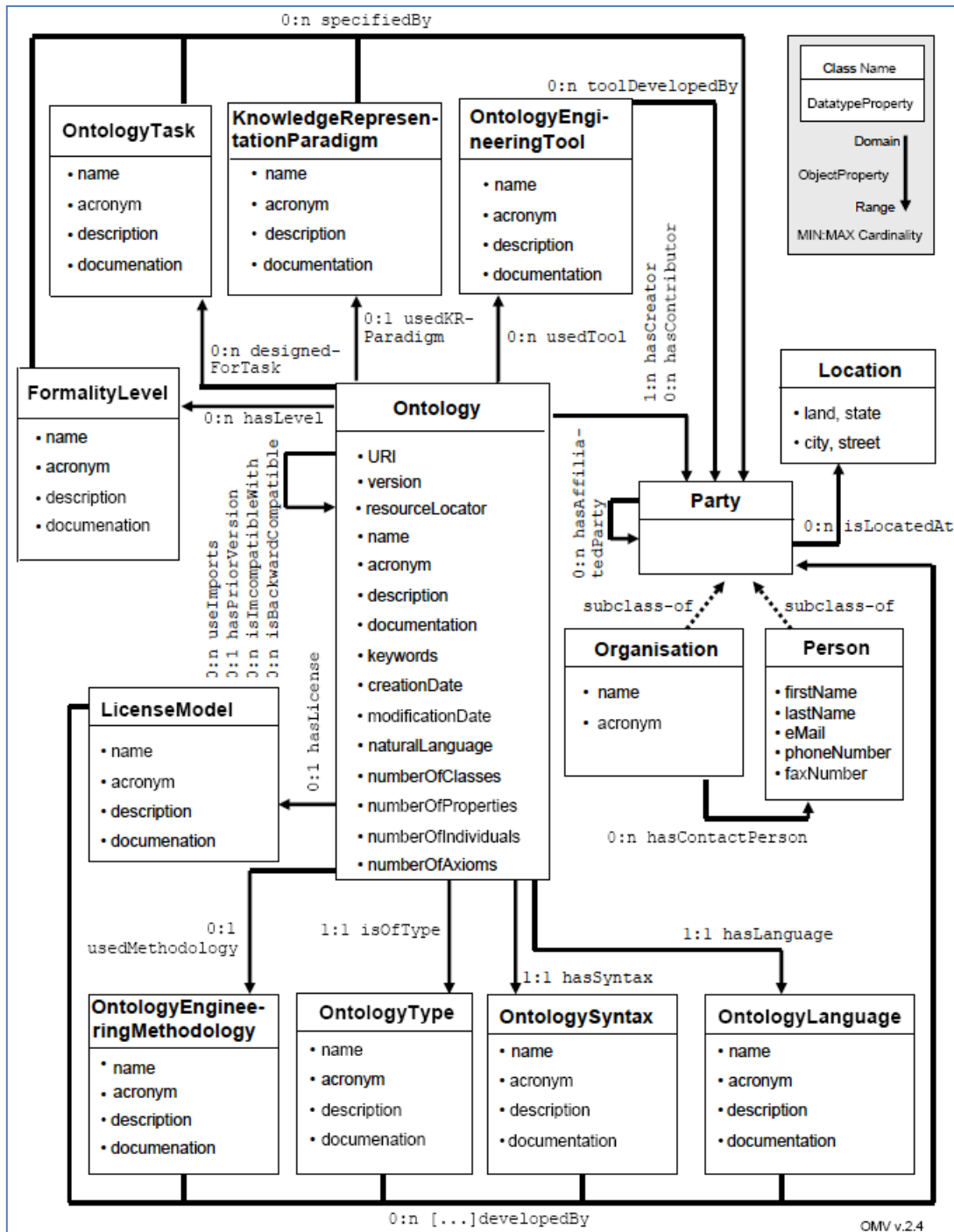


Figura 3 - Diagrama da ontologia OMV Core

Fonte: [PAL08].

Informações detalhadas sobre o modelo OMV Core podem ser obtidas no relatório técnico disponível em [PAL08].

O próximo capítulo apresentará uma breve discussão sobre a vantagem de utilizar um modelo de metadados para descrição de ontologias.

3.2 Padronização da descrição de ontologias

Segundo Hartmann *et al.* [HAR06], a capacidade de compartilhar e reaproveitar o conhecimento de forma eficiente é um pré-requisito para tornar a *Web Semântica* uma realidade. Entretanto, esse pré-requisito não está sendo atendido em função da falta de padronização para documentação e anotação de ontologias a partir de dados provenientes de metadados. O modelo OMV, descrito no subcapítulo anterior, oferece um vocabulário específico para capturar informações sobre as ontologias, representado por uma linguagem bem definida e interpretável tanto por humanos quanto por máquinas. Quando se fala sobre ontologias, há uma distinção entre a representação conceitual de um domínio de aplicação e as suas várias implementações. Como essas duas partes são caracterizadas por propriedades diferentes, metadados sobre ontologias devem diferenciar a conceitualização semântica da sua realização particular como uma ontologia concreta. Além disso, aspectos relacionados com o cenário da aplicação, escopo, propósito, ou avaliação de resultados são essenciais para o reuso bem sucedido de ontologias. Logo, esses aspectos devem ser capturados por um modelo de metadados.

3.3 A proposta de Bontas e co-autores

Na proposta apresentada por Bontas [BON05], são examinados métodos para promover o reuso de ontologias. Inicialmente, os autores apontam que as técnicas atuais de avaliação de ontologias para reuso são pouco exploradas. Essas técnicas, apesar de mencionarem a possibilidade de reuso do conhecimento existente como artefatos de entrada na fase de conceitualização falham ao definir precisamente como devem ser realizadas a descoberta de ontologias e a subsequente avaliação das ontologias candidatas. Os autores salientam que essas técnicas não deixam claro qual será o impacto do reuso de ontologias em todo o processo de desenvolvimento. Os autores ainda abordam outros trabalhos, os quais descrevem novas ontologias, mas que apresentam poucas recomendações para reuso.

Entretanto, o reuso de ontologias é fortemente recomendado nas atuais metodologias e *guidelines* como um fator chave para desenvolver ontologias de alta qualidade com baixo custo. Os autores enfatizam que esse processo de reutilização

diminui os custos de engenharia de ontologias ao evitar a reconstrução de ontologias existentes.

Esse trabalho analisou dois estudos que envolveram o reuso de ontologias e a construção de um modelo formal de representação e contextualização de ontologias (OMV). Além disso, investigou o significado das características sintáticas e a potencial utilização em operações de *matching* para demonstrar a usabilidade da contextualização impulsionada pelo modelo OMV.

3.4 Exemplos de aplicações do modelo OMV

No trabalho de Palma [PAL06], são descritas duas aplicações que utilizam o modelo OMV como padrão de descrição de ontologias. A primeira aplicação desenvolvida é chamada de Oyster¹¹ P2P. Ela utiliza o modelo OMV como um repositório isolado, mas que pode comunicar-se com outros repositórios Oyster através de uma rede P2P. O objetivo do Oyster é prover um ambiente de compartilhamento de conhecimento de forma descentralizada, utilizando tecnologias da *Web Semântica* que permitem aos engenheiros de ontologias compartilhá-las com facilidade [HAR05]. A segunda aplicação, denominada de *A Central Ontology Metadata Portal*, também utiliza o modelo OMV, porém de forma centralizada. Assim, esse sistema permite que as ontologias sejam armazenadas em uma base única e mantidas, por um longo prazo, por comunidades que as utilizam freqüentemente. As ontologias com alto nível de aceitação e reutilização precisam ser armazenadas em um metamodelo central para que possam ser acessadas facilmente por uma grande quantidade de usuários [HAR05].

3.5 O portal SWOOGLE

No artigo de Ding [DIN04] é apresentado o *Swoogle*¹², que é um sistema de indexação e recuperação de ontologias em documentos nos formatos RDF e OWL. O sistema extrai os metadados dos documentos encontrados na *Web* e realiza uma

¹¹ <http://oyster.ontoware.org/>

¹² <http://swoogle.umbc.edu/>

computação da relação entre esses documentos. Além disso, os documentos selecionados pela pesquisa são indexados por um sistema de recuperação o qual pode localizar documentos relevantes tanto por seqüências de caracteres, quanto por palavras chaves descritas em referências URIs, bem como para calcular a similaridade entre um conjunto de documentos. Os metadados utilizados pelo *Swoogle* são mais simples que os propostos pelo OMV. Por exemplo, os metadados do *Swoogle* não contemplam informações sobre o tipo de licenciamento da ontologia, ferramenta de construção, paradigma de representação do conhecimento, metodologia de construção, nível de formalidade, entre outras. Os metadados do *Swoogle* incluem informações sobre a linguagem de construção da ontologia, o tipo de expressividade do OWL empregado (*Lite*, *DL*, e *Full*), e, por fim, as possíveis relações entre ontologias.

3.6 A ferramenta *WebCORE*

No trabalho realizado por Iván Cantador [CAN07], é discutido o desenvolvimento de uma extensão da ferramenta *CORE*, que é uma ferramenta para reúso e avaliação colaborativa de ontologias. A *WebCORE* implementa uma nova abordagem para localização de ontologias a partir de um repositório. Além disso, ela procura recomendar ao usuário qual é a ontologia mais apropriada para representação de um dado domínio.

A aplicação *WebCORE* foi dividida em três módulos que, combinados em um *pipeline*, constituem uma arquitetura de localização de ontologias. O primeiro módulo recebe uma descrição de um problema como um conjunto de termos e permite ao usuário refiná-la e incrementá-la usando a *WordNet*¹³. O segundo módulo aplica múltiplos critérios automáticos para avaliação das ontologias de um repositório, e determina qual delas melhor se enquadra no critério de busca informado. Este módulo retorna uma lista de ontologias ranqueadas a partir do critério de busca. Finalmente, o terceiro módulo usa as validações manuais do usuário a fim de incorporar uma avaliação humana das ontologias. Portanto, este último módulo leva em conta a opinião do usuário sobre as características das ontologias, permitindo ao sistema recomendar uma lista de ontologias gerada a partir de uma estratégia de busca múltipla.

¹³ <http://wordnet.princeton.edu/>

3.7 Análise de Natalya F. Noy e co-autores

No trabalho de Natalya e co-autores [NOY04], são discutidos os benefícios que os repositórios de ontologias podem oferecer para a comunidade de pesquisadores das ciências biomédicas, quando estes são combinados e/ou integrados. Tal combinação serviria como um diretório capaz de prover uma visão unificada das ontologias distribuídas através da *Web Semântica*. Os autores ainda destacam que um conjunto distribuído de repositórios de ontologias e outras fontes de conhecimento ajudariam os biólogos a terem um entendimento melhor das informações disponíveis.

Por fim, os autores relacionam as funcionalidades desejáveis que um repositório de ontologias deve oferecer, tais como: sumários de ontologias, qualidade de ontologias, navegação on-line de ontologias, mecanismo de visualização gráfica das ontologias armazenadas, mecanismos de busca de ontologias, mapeamento e alinhamento de ontologias.

4 MÉTODO

Este capítulo descreve o método empregado para execução desta pesquisa, ou seja, as estratégias, as técnicas e os instrumentos para coleta de dados, bem como os procedimentos adotados para a análise dos dados. Para tanto, apresenta alguns princípios metodológicos aplicáveis à área de Ciência da Computação e Sistemas de Informação.

Oates [OAT06] propõe um modelo de processo de pesquisa com os seguintes componentes: experiência pessoal e motivação, questão de pesquisa, revisão da literatura, *framework* conceitual, estratégias, métodos de coleta de dados e análise dos dados.

O primeiro corresponde às diferentes razões para a realização da pesquisa, a qual está descrita na introdução deste trabalho, no subcapítulo motivação. A revisão da literatura consiste em identificar áreas tópicos a pesquisar e também dar suporte à construção de um novo conhecimento, manter ou refugar conceitos e argumentos. Assim, os capítulos sobre ontologias e trabalhos relacionados apresentam alguns conceitos, trabalhos e modelos que embasaram esta pesquisa. A estrutura conceitual representa a forma de estruturação do pensamento sobre a pesquisa e o processo de construção do conhecimento, que neste trabalho consiste na apresentação do método da pesquisa. Por sua vez, a estratégia de pesquisa corresponde à abordagem adotada para responder a questão de pesquisa, podendo ser a realização de *survey*, experimento, estudo de caso, pesquisa-ação, etnografia, ou mesmo, o *design* e a criação de produtos de *software* da área de Sistemas de Informação (SI). Assim, esta pesquisa envolveu a extensão do modelo OMV, a prototipação e o desenvolvimento de novos serviços de descrição e recuperação de ontologias do Portal OntoLP, com base no modelo estendido para descrição de ontologias, e a realização de uma *survey* para avaliação do protótipo implementado.

As principais técnicas e instrumentos para coletar dados e informações envolvem a aplicação de questionários, a realização de entrevistas e de observações, além da análise de documentos. A análise dos dados consiste na forma de interpretação e na apresentação dos dados obtidos na pesquisa. A primeira pode ser qualitativa ou quantitativa. A segunda envolve a realização de relatórios, artigos, apresentações, etc. Por fim, a análise descritiva dos dados, a ser relatada nos capítulos a seguir, refere-se à

apresentação da adaptação do modelo OMV, à documentação do protótipo da ferramenta de descrição e recuperação de ontologias do Portal OntoLP, e aos resultados obtidos na aplicação da *survey*. [OAT06]. Cabe salientar que os principais instrumentos utilizados neste trabalho foram a análise documental e a aplicação de questionários.

4.1 Design e criação de ferramenta de TI

Para March e Smith [MAR95], a estratégia de design e criação pode envolver a proposição de um construto, um modelo, método ou instanciações de um produto ou artefato de TI ou SI. O construto refere-se aos conceitos ou vocabulário utilizados em determinada área de domínio da TI, enquanto os modelos resultam da combinação de construtos para entendimento ou solução de um problema. Por sua vez, o método guia o modelo a ser produzido e os estágios dos processos a serem seguidos para a solução de um problema usando a TI. As instanciações demonstram teorias, construtos, modelos ou métodos que podem ser implementadas por um sistema baseado em computador.

A abordagem de desenvolvimento de um artefato normalmente envolve a identificação e a solução de um problema, podendo adotar princípios de desenvolvimento de sistemas [OAT06]. Nunamaker *et al.* [NUN91] propõem uma metodologia de desenvolvimento de sistemas, iniciando no entendimento de problemas e questões de pesquisa, passando por processo de prototipação de sistemas, podendo resultar em contribuições conceituais ou práticas como é apresentado na Figura 4.

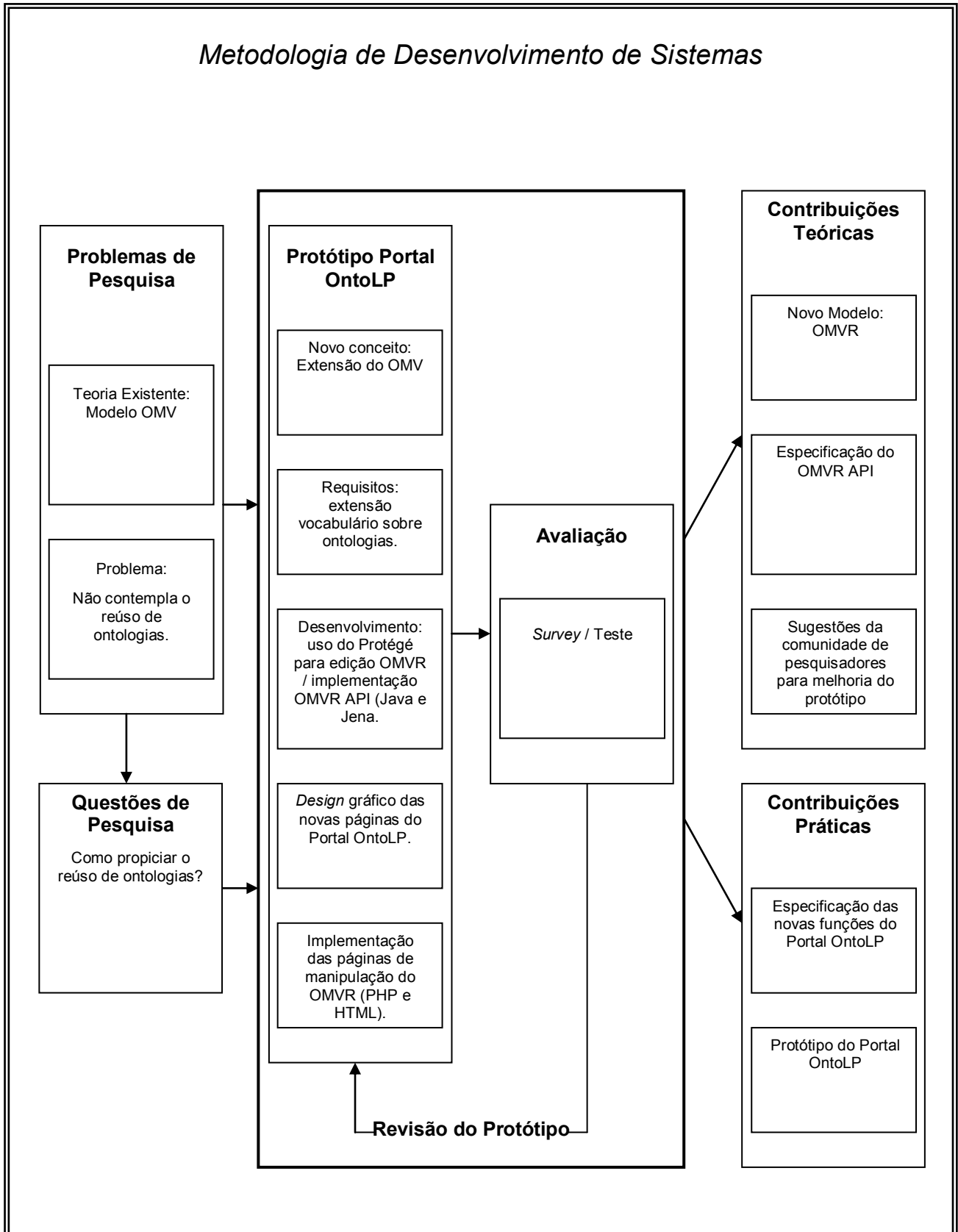


Figura 4 - Metodologia de Desenvolvimento de Sistemas.

Fonte: Modelo adaptado de [NUN91].

Conforme apresentado na Figura 4, o método utilizado para o desenvolvimento do protótipo dos novos serviços do Portal OntoLP é constituído de cinco atividades. Na atividade Novo Conceito, foi definido que informações seriam adicionadas ao modelo OMV, a partir das conclusões obtidas do estudo dos conceitos apresentados nos trabalhos relacionados (ver capítulo 3) como, por exemplo, avaliações de ontologias [CAN07], métricas [FRE09] e projetos que utilizam uma determinada ontologia, como requisito desejado do projeto Portal OntoLP (o modelo estendido é apresentado no capítulo 5). O passo seguinte foi a especificação dos requisitos, ou seja, as funcionalidades desejáveis a serem implementadas no Portal OntoLP, a exemplo da pesquisa, das avaliações, das métricas, além de outras funções para descrição e recuperação de ontologias. A seguir, na atividade desenvolvimento, foram implementados os novos conceitos no modelo OMV, através do editor de ontologias Protégé, seguido da implementação do software para manipulação do novo modelo, o qual constitui o núcleo para implementação das interfaces e serviços do Portal OntoLP. Em seguida, foram definidos os *layouts* de cada uma das novas páginas, a fim de atender os requisitos previamente definidos. Por fim, na última etapa, foram implementadas as páginas, com base nos *layouts* definidos, utilizando a linguagem PHP, formando os novos serviços do Portal OntoLP.

No que diz respeito à avaliação do trabalho realizado, foi desenvolvida uma *survey*, a fim de validar os novos serviços do Portal OntoLP, apresentado no subcapítulo 4.2. Essa avaliação previu uma etapa de teste, em que os usuários utilizaram o protótipo proposto para então responder as questões do questionário desenvolvido. Por fim, são apresentadas as contribuições teóricas, que se constitui do presente estudo, bem como das contribuições práticas, formadas pelo protótipo implementado e componentes de *software*.

4.2 Survey

O objetivo de uma *survey* é obter dados a partir de um grupo de participantes, de forma sistematizada e padronizada [OAT06]. A partir da avaliação dos dados, com o objetivo de encontrar padrões, é possível fazer uma generalização dos resultados para um grupo maior do que aquele definido como alvo para execução da pesquisa. Os questionários são uns dos instrumentos mais utilizados para realizar uma *survey*, embora

outros instrumentos de coleta de dados também possam ser empregados como, por exemplo, entrevistas, observações e documentos.

Segundo Briony [OAT06], *surveys* são amplamente empregadas e aceitas na área de Sistemas de Informação. Conseqüentemente, em Computação, é um dos métodos mais utilizados para avaliação de sistemas. O planejamento de uma *survey* pode ser dividido em seis atividades, a saber:

- a) **Requisitos de dados:** corresponde aos tipos de dados que se deseja coletar na pesquisa, por exemplo, se os dados serão relacionados diretamente com a questão de pesquisa, ou indiretamente a exemplo dos dados demográficos sobre os respondentes (sexo, idade, etc). Cabe salientar que essa etapa deve ser cuidadosamente definida, porque não haverá a chance de alterar os dados, ou de acrescentar novos dados ao *survey* após a sua aplicação;
- b) **Métodos para geração de dados:** consiste na decisão da forma de coletar os dados, ou seja, a aplicação de um questionário, a realização de entrevistas, a análise de documentos, ou ainda, a observação. Normalmente, em uma *survey*, se adota apenas um tipo de coleta de dados;
- c) **Base de amostragem:** aborda a definição do público-alvo da pesquisa e as fontes que serão utilizadas para obtenção dos participantes. Por exemplo, para realizar uma *survey* sobre um sistema de ajuda a clientes, o pesquisador pode aproveitar os registros de chamados do aplicativo para base de amostragem;
- d) **Técnicas de amostragem:** refere-se ao tipo de seleção da base de amostragem, ou seja, a utilização de técnicas de amostragem probabilísticas e não-probabilísticas. A primeira técnica sugere que a amostra de respondentes, definida pelo pesquisador, tem representatividade em relação à população total a ser estudada. Na segunda técnica, o pesquisador não sabe determinar se a amostra de pessoas é representativa em relação à população total. Além disso, cada uma das técnicas de amostragem possui técnicas específicas para seleção do tamanho da amostra;
- e) **Taxa de respostas e não-respostas:** nessa atividade são definidas as estratégias para melhorar a taxa de resposta da amostra de respondentes selecionada. Nessa atividade são definidas as formas que serão adotadas para persuadir as pessoas a responderem a pesquisa, de tal forma que o pesquisador possa obter o número de respostas que ele deseja. São sugeridas

algumas técnicas para melhorar a taxa de resposta que vão desde uma boa explicação sobre os propósitos da pesquisa; entrega de questionários pessoalmente; envio de correspondência já com selos, até o acompanhamento do andamento da pesquisa. Nesse último caso, em intervalos de tempos definidos, o pesquisador reenvia mensagens agradecendo às pessoas que já responderam a pesquisa, mas reforçando da importância da participação e do preenchimento do questionário daqueles que ainda não o fizeram;

- f) **Tamanho da amostra:** a definição do tamanho da amostra deve ter um grau de significância aceitável. Para projetos de pesquisas pequenos, uma boa prática é ter uma amostra final de, pelo menos, trinta. Contudo, é comum um percentual inicial de retorno de 10% da base de amostragem.

A fim de avaliar os novos serviços de descrição e recuperação de ontologias, do Portal OntoLP, foi desenvolvido um questionário (ver APÊNDICE B – QUESTIONÁRIO) para coleta das impressões dos participantes sobre os novos serviços do Portal OntoLP em relação ao serviço de descrição de ontologias vigente no portal.

Seguindo as etapas sugeridas por [OAT06], na atividade de **requisitos de dados** foi determinado que os dados coletados deveriam ter relação direta com a questão de pesquisa. Sendo assim, o questionário é constituído de quatro questões: (i) duas objetivas de escala graduada de 1 a 5 (**1 Discordo totalmente; 2 Discordo; 3 Neutro; 4 Concordo; 5 Concordo totalmente**) as quais objetivaram avaliar a satisfação dos respondentes ao utilizarem os novos serviços; e (ii) duas questões dissertativas com o objetivo de obter do participante as suas impressões sobre a abrangência dos metadados do modelo OMVR e quais informação, do modelo, são mais úteis para localização de ontologias.

Na atividade **método para geração de dados**, optou-se pelo questionário como forma de captação das informações de interesse para pesquisa. A coleta das respostas dos participantes foi realizada por intermédio de um formulário eletrônico, publicado em um endereço da *Web* especificamente criado para abrigar o novo Portal OntoLP¹⁴.

Em relação à **base de amostragem**, o público-alvo da pesquisa corresponde a pesquisadores brasileiros que participem de atividades de pesquisa relacionadas a ontologias. A população envolveu um número aproximado de cem (100) pesquisadores, os quais participaram de congressos brasileiros sobre ontologias no período de 2008 a 2009.

¹⁴ <http://www.ontolp.com.br/questionarioOMVR.php>

Na etapa de definição da **técnica de amostragem** optou-se pelo uso de uma técnica não-probabilística por conveniência, pela adesão voluntária dos participantes em participar da pesquisa, visto que não buscou a generalização dos resultados para a população definida.

Para estimular a participação na pesquisa, para aumentar a **taxa de respostas**, adotou-se o método de reenvio do e-mail com o convite da pesquisa à lista de participantes. Além disso, houve o cuidado de agradecer às pessoas que por ventura já tivessem respondido ao questionário, reforçando a importância do preenchimento do formulário para não comprometer a *survey*.

Apesar de uma expectativa inicial de **tamanho da amostra** de trinta por cento (30%) da população, ou seja, 30 pesquisadores; obteve-se um total de 19 respondentes no período de dezembro de 2009. Contudo, o questionário continuará disponível, com o objetivo de obter a opinião dos usuários e continuar a aprimorar os serviços disponibilizados pelo Portal OntoLP.

4.3 Análise dos dados

Esta pesquisa envolveu uma análise descritiva dos dados qualitativos e quantitativos tanto da especificação do modelo OMVR e do protótipo da ferramenta de descrição e recuperação de ontologias quanto dos dados obtidos na aplicação do questionário. A análise dos dados também compreendeu na definição das categorias e variáveis para apresentação dos resultados da pesquisa. Essa preparação dos dados envolveu quatro etapas:

- a) a definição das classes, propriedades e relações dos metamodelo OMVR a serem descritas;
- b) os elementos e serviços do Porto OntoLP a serem apresentados;
- c) os componentes de infra-estrutura e de interface de navegação para especificação do protótipo; e,
- d) a preparação e a tabulação das informações obtidas através da aplicação do questionário.

Nesta última etapa, foi utilizado o aplicativo Microsoft Excel para a geração das estatísticas descritivas tanto das questões fechadas quanto das questões abertas referentes ao questionário, visto que foi realizada a análise de conteúdo para o levantamento e contagem das variáveis nominais das respostas obtidas. Por fim, na apresentação das informações desta pesquisa, foram utilizados recursos como diagramas, gráficos e imagens do Portal OntoLP.

5 PROPOSTA DE ADAPTAÇÃO DO MODELO OMV

Este capítulo apresenta a proposta de adaptação do OMV para descrever, recuperar e auxiliar no processo de seleção de ontologias para reuso. Deste ponto em diante o modelo OMV adaptado é referenciado no texto como OMVR.

Em estudo inicial, identificou-se que o modelo proposto em [PAL08] não contempla informações sobre avaliação de ontologias, métricas mais elaboradas e informação sobre a evolução de uma ontologia.

Elena Simperl [SIM09] realiza um estudo sobre a maturidade dos atuais processos, metodologias e ferramentas voltadas para o reuso de ontologias. A autora apresenta os pontos fortes e as necessidades de melhorias no atual estado da arte da engenharia de ontologias, a partir das evidências encontradas em estudos de casos realizados nas áreas de *e-Health* e *e-Recruitment* e em revisões de publicações importantes da área. A autora relata que é necessário desenvolver novas aplicações que auxiliem os pesquisadores a melhorar o reuso das atuais fontes de ontologias e, além disso, desenvolver ainda mais as metodologias e ferramentas existentes com o objetivo de melhorar o processo de seleção de ontologias.

Especificamente sobre a avaliação de ontologias, [YU09] apresenta uma metodologia para auxiliar no processo de busca e seleção de ontologias apropriadas para reuso, batizada de ROMEO. Essa metodologia propõe um conjunto de critérios de avaliação que uma ontologia deve satisfazer. Esses critérios estão divididos entre resultados que podem ser expressos de forma numérica ou textual. Com o objetivo de exemplificar a aplicação da metodologia, o autor utiliza um conjunto de ontologias obtidas a partir da estrutura do *Wikipedia* onde são aplicados os critérios de avaliação e analisados os resultados obtidos. Para [TAR07], a construção de uma nova ontologia pode ser realizada através de duas abordagens: (i) pelo desenvolvimento a partir do zero; ou (ii) pelo desenvolvimento a partir de uma ontologia existente. Nesses dois casos são necessários o emprego de um conjunto de métricas, como o OntoQA [TAR07], e metodologias, como o ROMEO [YU09], que auxiliem na construção e seleção de ontologias candidatas ao reuso.

Levando em consideração que o *OMV Core* não contempla informações importantes para possibilitar o reúso, e nos valendo da possibilidade de estendê-lo, criaram-se novas classes no *OMV Core*, as quais estão representadas na Figura 5.

Com base nos trabalhos relacionados, passou-se a considerar o conceito de avaliação de ontologias no *OMVR*, pois se trata de uma metodologia desenvolvida especificamente para auxiliar no processo de seleção de ontologias.

A inclusão do conceito de avaliação na adaptação do modelo é estruturado de forma que possam ser instanciados os critérios de avaliação na classe ***EvaluationCriteria*** (ver Figura 5) que, em um segundo momento, servirão para compor um conjunto de avaliações na classe ***OntologyEvaluation*** (ver Figura 5) associadas a uma ontologia. Tal associação permitirá que seja atribuído ao critério de avaliação uma pontuação ou um texto, dependendo da definição do tipo de critério, ou seja, um critério pode exigir um valor numérico ou textual como forma de avaliação.

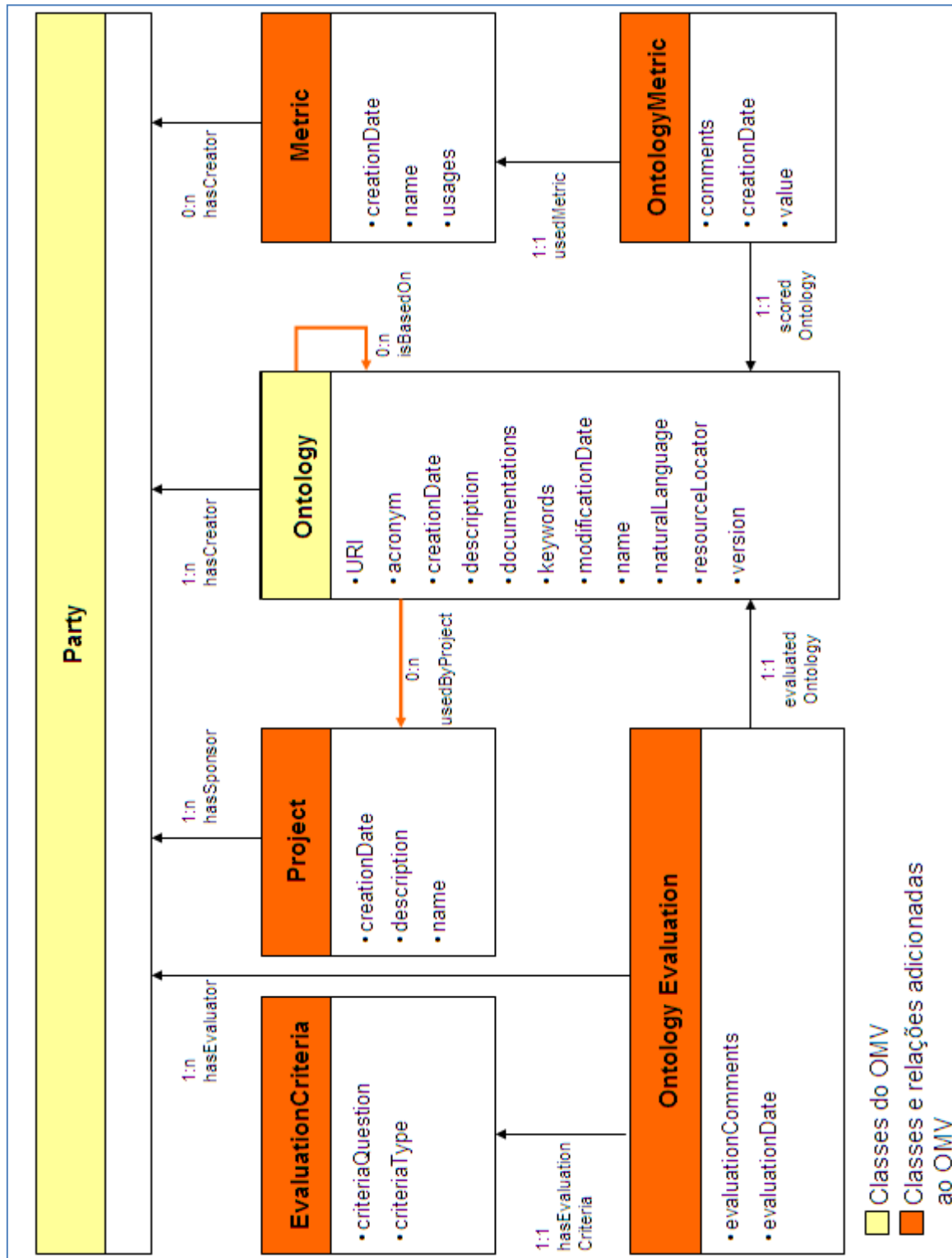


Figura 5 - Metadados OMVR. Adaptação do OMV Core.

A seguir são apresentadas as descrições das novas classes, as respectivas propriedades e relações que foram adicionadas ao OMV Core, as quais compõem o OMVR.

5.1 Descrição das classes e propriedades do OMVR

Neste subcapítulo serão apresentadas as classes que integram o modelo de metadados OMVR.

5.1.1 Classe *OntologyEvaluation*

As instâncias da classe *OntologyEvaluation* representam os critérios de avaliação aplicados à ontologia. O objetivo é permitir que pesquisadores, interessados em criar uma ontologia a partir de uma existente, possam selecionar uma determinada ontologia mediante análise dos valores atribuídos ao conjunto de critérios associados a ela. Além disso, é possível que sejam criadas novas instâncias de avaliações sobre ontologias

5.1.1.1 Propriedades

- a) ***evaluationComments***: texto livre para que possa ser realizado comentário de critério de avaliação, quando o mesmo exige uma avaliação textual;
- b) ***evaluationDate***: data em foi realizada a avaliação da ontologia, e;
- c) ***evaluationValue***: valor que deve ser atribuído ao critério de avaliação, quando o mesmo exige uma avaliação numérica.

5.1.2 Classe *EvaluationCriteria*

A classe *EvaluationCriteria* representa os critérios de avaliação, os quais poderão ser utilizados para avaliar ontologias (conforme descrito no subcapítulo 5.1.1). A atribuição de um critério de avaliação tem por objetivo prover dados qualitativos sobre as

ontologias, os quais podem ser usados como forma de avaliação da ontologia por usuários que pretendem reusá-la.

5.1.2.1 Propriedades

- a) **criteriaQuestion**: pergunta relativa ao critério que se pretende avaliar.
- b) **criteriaType**: define se a pergunta espera uma resposta numérica ou textual.

Essa classe possui um conjunto de instâncias de critérios de avaliação apresentado em [YU09], a saber:

1. Avaliações numéricas:

- a) Quantos erros de referência circular foram encontrados na ontologia?
- b) Quantos conceitos idênticos foram modelados usando nomes diferentes?
- c) Quantas relações redundantes do tipo *subclass-of* foram encontradas na ontologia?

2. Avaliações textuais:

- a) A ontologia tem conceitos claramente definidos?
- b) A ontologia tem instâncias claramente definidas?
- c) As instâncias da ontologia cobrem adequadamente as instâncias do domínio?
- d) A ontologia captura corretamente os conceitos do domínio?
- e) A ontologia captura corretamente os relacionamentos entre as instâncias de um domínio?

Cabe salientar que novos critérios poderão ser instanciados no OMVR, com o objetivo de ampliar o conjunto existente. Dessa forma, com o passar do tempo, os usuários terão à disposição mais opções para avaliar ontologias, bem como para seleção.

5.1.3 Classe *Metric*

Esta classe representa um conjunto de métricas instanciadas. Tais métricas podem ser utilizadas para pontuar uma ontologia, atribuindo um valor quantitativo às ontologias, a fim de auxiliar os usuários durante o processo de seleção de ontologias para reuso. O

modelo adaptado possui como instâncias vinte métricas coletadas de [FRE09]. Alguns exemplos das métricas instanciadas são:

- a) Número de classes raiz [GAN06a];
- b) Número de classes folha [GAN06a];
- c) Profundidade máxima [GAN06a];
- d) Número total de caminhos [GAN06a];
- e) Média de propriedades por classe [YAN06], e;
- f) Profundidade média das classes folha [CRO05].

5.1.3.1 Propriedades

- a) **creationDate**: data aproximada de concepção da métrica;
- b) **name**: nome da métrica, e;
- c) **usages**: texto livre para descrição e orientação ao usuário de como a métrica é usada, como ela deve ser aplicada e a interpretação do resultado.

5.1.4 Classe *OntologyMetric*

As instâncias da classe *OntologyMetric* referem-se às métricas calculadas para a ontologia. Uma ou mais métricas poderão estar associadas à ontologia, proporcionando aos usuários um conjunto de métricas para seleção de ontologias sobre diferentes aspectos.

5.1.4.1 Propriedades

- a) **comments**: texto livre para que possa ser registrado alguma informação textual que auxilie na compreensão da métrica associada à ontologia;
- b) **creationDate**: data em que a métrica foi calculada, e;
- c) **value**: valor atribuído à métrica, conforme definição da métrica (ver propriedade usages da classe *Metric*).

5.1.5 Classe *Project*

As instâncias da classe *Project* representam os projetos de pesquisa, ou comerciais que podem ser referenciados por ontologias.

5.1.5.1 Propriedades

- a) ***creationDate***: data aproximada da constituição do projeto;
- b) ***description***: texto livre para descrição do projeto, seus objetivos e artefatos produzidos a partir dele, e;
- c) ***name***: nome pelo qual o projeto é conhecido.

5.2 Descrições das relações do OMVR

Neste subcapítulo serão apresentadas as relações adicionadas ao modelo de metadados OMV *Core*.

- a) ***evaluatedOntology***: relaciona a ontologias com as suas respectivas avaliações;
- b) ***isBasedOn***: referência para outra ontologia usada como ponto de partida para construção da ontologia. Essa referência também permite identificar o grau de reúso que uma ontologia possui, isto é, se ela tem um alto nível de referência para construção de outras ontologias;
- c) ***hasEvaluator***: referência para grupo de pesquisa ou empresas responsáveis pela avaliação da ontologia;
- d) ***hasEvaluationCriteria***: referência para critério de avaliação associado à ontologia;
- e) ***hasSponsor***: referência para a instituição patrocinadora do projeto;
- f) ***usedByProject***: referência para o projeto que usa a ontologia;
- g) ***hasCreator***: referência para o grupo de pesquisa ou instituição que desenvolveu a métrica, e;
- h) ***usedMetric***: referência para a métrica associada à ontologia.

Cabe salientar que as classes ***Ontology*** e ***Party*** provêm do OMV Core original. Outro aspecto importante é que a notação gráfica utilizada no OMVR é a mesma adotada no diagrama do OMV Core [PAL08].

A partir da adaptação realizada, o OMVR passa a suportar, em sua estrutura, a descrição de metodologias para avaliação de ontologias propostas em [SIM09], [YU09] e [TAR07]. Outra contribuição da adaptação é que ela permite armazenar e agregar novas métricas sempre que for preciso. Essas métricas poderão ser utilizadas para avaliar uma ontologia sob diferentes aspectos. Por fim, espera-se que a extensão do OMV Core proporcione maior flexibilidade no processo de localização de ontologias, pois oferece aos usuários um conjunto adicional de informações sobre ontologias.

O próximo capítulo discorre sobre o Portal OntoLP, com uma breve descrição de seus objetivos e serviços de compartilhamento de recursos. Além disso, serão apresentados o protótipo de Portal OntoLP e os novos serviços de descrição e recuperação de ontologias, propostos nesse estudo, bem como as ferramentas desenvolvidas para recepção de ontologias e manipulação do OMVR.

6 PROTÓTIPO DA FERRAMENTA DE DESCRIÇÃO E RECUPERAÇÃO DE ONTOLOGIAS

Este capítulo apresenta o Portal OntoLP, bem como as funções desenvolvidas para manipulação do OMVR. Os objetivos de tais ferramentas são instanciar o modelo e recuperar informações das instâncias do OMVR. Cabe salientar que estas ferramentas serão acopladas ao Portal OntoLP, com o objetivo de estender as suas funcionalidades.

6.1 O Portal OntoLP

O Portal OntoLP (ver Figura 6) propõe-se a divulgar ontologias disponíveis na língua portuguesa (incluindo desde bases terminológicas, vocabulários controlados, até ontologias mais complexas do tipo OWL-DL), bem como ferramentas e recursos relacionados à pesquisa na área.



Figura 6 - Página inicial do Portal OntoLP.

O portal apresenta atualmente seis áreas diferentes: *Home*, *Recursos*, *Sobre*, *Contato*, *Oficina* e *Links*. Na área *Home* o objetivo do portal é apresentado, enquanto que a área denominada *Recursos*, demonstra a lista das ontologias e os trabalhos relacionados no portal. A área *Sobre* contém informações a respeito do grupo de pesquisa

envolvido neste projeto e seus integrantes. Através da área Contato (ver APÊNDICE A – PÁGINA DE CONTATO DO PORTAL ONTOLP) outros grupos, pesquisadores ou interessados no tema podem enviar seus trabalhos, suas sugestões e dicas. Finalmente, o portal oferece uma área de *links* onde estão elencados grupos de pesquisa e pesquisadores que colaboram direta ou indiretamente com o portal, bem como eventos relacionados com a área de ontologia.

6.2 Especificação das ferramentas de manipulação do OMVR

Este subcapítulo apresentará o detalhamento da ferramenta desenvolvida para manipulação do OMVR, a qual será referenciada como OMVR API. Essa API constitui a base de construção dos novos serviços do Portal OntoLP.

6.2.1 Organização dos componentes de infra-estrutura

Uma das principais tarefas da ferramenta de manipulação do OMVR é receber as ontologias, através da página de submissão de ontologias, e criar uma nova instância no OMVR, combinando as informações do formulário com as informações capturadas do arquivo OWL, fornecido pelo usuário. A Figura 7 apresenta um esquema do fluxo de informações na recepção de uma ontologia pelo Portal OntoLP.

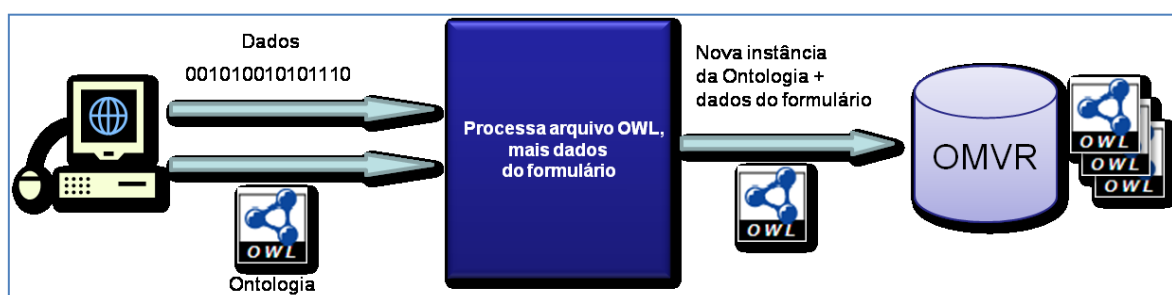


Figura 7 - Fluxo de armazenamento de uma ontologia no Portal OntoLP.

O OMVR API foi desenvolvido em linguagem de programação Java¹⁵ e utiliza extensivamente o *framework* Jena¹⁶ para manipulação de ontologias. A arquitetura do

¹⁵ <http://www.javasoft.com>

¹⁶ <http://jena.sourceforge.net/> - desenvolvido pela HP como um projeto *open source*.

Jena possui em sua API suporte para a criação, modificação e navegação em grafos descritos em RDF, RDF *schema* e OWL [CRA07] [IBM09]. A API também suporta diferentes tecnologias para armazenamento de grafos, além de oferecer interfaces para a implementação de código de escrita e leitura em diferentes linguagens para representação de grafos RDF [MCB02]. Além disso, grandes empresas multinacionais da indústria de software como, por exemplo, HP¹⁷ e Oracle¹⁸, vêm utilizando o Jena para construção de *softwares* para a *Web Semântica*. Além de empresas de vulto internacional, instituições acadêmicas também desenvolvem pesquisas e ferramentas sobre o *framework* Jena. Um exemplo prático do uso do desse *framework* é o editor de ontologias Protégé [SCB09], desenvolvido e mantido pelo Centro de Pesquisas em Informática Biomédica de Stanford, que até a sua versão 3.3.1 utiliza o Jena para manipulação de ontologias.

Além da capacidade de criar instâncias no OMVR, o OMVR API oferece outros serviços como, por exemplo, recuperação de informações das instâncias das ontologias (a partir da classe *Ontology*), possibilidade de criar e recuperar instâncias das avaliações de ontologias, bem como das métricas calculadas para cada ontologia armazenada no OMVR. A recuperação das informações das ontologias é realizada através de um conjunto de métodos que lêem os dados do OMVR e, logo após, gravam as informações solicitadas em arquivos no formato XML, para posterior apresentação na camada de apresentação do protótipo da aplicação. Cabe salientar que o cálculo das métricas, o qual é realizado na etapa de submissão da ontologia, é executado por uma biblioteca desenvolvida pela pesquisa de mestrado de [FRE09]. Essa biblioteca calcula automaticamente vinte métricas para a ontologia, onde os resultados obtidos são instanciados no OMVR para posterior consulta pelos usuários a partir da interface do Portal OntoLP.

Por fim, o OMVR API utiliza o motor de consultas SPARQL¹⁹ do Jena para implementar o serviço de pesquisa avançado do Portal OntoLP. A consulta é executada sobre o OMVR, especificamente sobre a classe *Ontology*, onde são selecionadas todas as ontologias que satisfaçam os critérios de seleção do usuário. O resultado obtido é então gravado em um arquivo XML, o qual é apresentado de forma apropriada pela

¹⁷ HP *Laboratories*, Bristol, UK

¹⁸ <http://www.oracle.com/technology/tech/>

¹⁹ <http://www.w3.org/TR/rdf-sparql-query/>

camada de apresentação do protótipo do Portal OntoLP. A Figura 8 apresenta um diagrama que representa as camadas de software que compõem o Portal OntoLP.

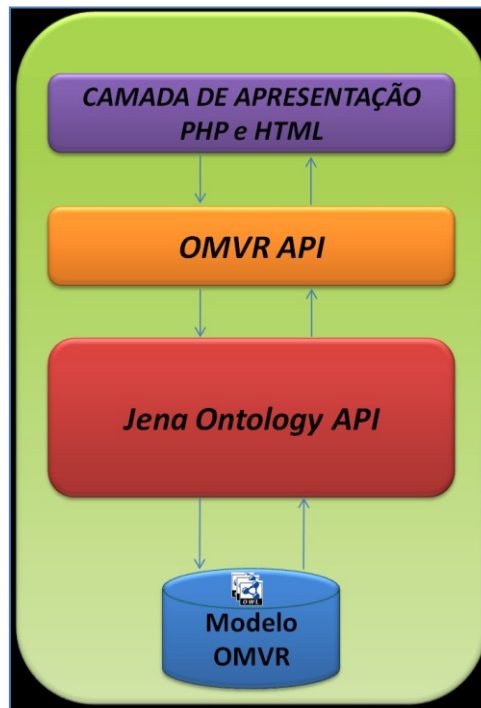


Figura 8 - Estrutura do OMVR API.

Conforme a Figura 8, a camada OMVR API é formada por um conjunto de classes que oferecem os serviços necessários de manipulação do OMVR. Sobre a OMVR API existe a camada de apresentação do Portal OntoLP. Essa camada é formada por um conjunto de páginas HTML²⁰ dinâmicas, desenvolvidas em linguagem de script PHP²¹. A seguir são relacionadas as classes que compõem a API seguido de uma breve descrição das suas funcionalidades.

6.2.1.1 Classe OMVR

Essa é a principal classe da API. Ela provê todos os serviços necessários para criar instâncias das classes *Ontology*, *OntologyMetric* e *OntologyEvaluation* do OMVR. Ainda, ela oferece os serviços necessários para recuperar informações de metadados do OMVR para apresentação dos mesmos pela camada de interface com o usuário. Além disso, essa classe oferece um método para execução de consultas SPARQL.

²⁰ <http://www.w3.org/standards/Webdesign/htmlcss>

²¹ <http://php.net/index.php>

6.2.1.2 Classe *AdvancedSearchEngine*

Essa classe é responsável por criar uma consulta SPARQL a partir de um arquivo XML, que contém os critérios de pesquisa de ontologias avançado do Portal OntoLP. Uma vez criada a declaração SPARQL, a mesma é submetida para o motor de execução SPARQL do Jena. Os nomes das ontologias recuperadas pela consulta são gravados em um arquivo XML, o qual é utilizado pela camada de apresentação do Portal OntoLP para compor a árvore de navegação de ontologias descritas no OMVR.

6.2.1.3 Classe *OntologyEvaluationInformation*

Essa classe representa a estrutura de dados com todas as informações necessárias para criar uma instância da classe *OntologyEvaluation* do OMVR. Essa classe ainda oferece um conjunto de métodos *sets* e *gets* para atualizar e recuperar informações dos seus atributos. Ela é utilizada pela página de avaliação de ontologias do Portal OntoLP.

6.2.1.4 Classe *OntologyInformation*

Essa classe representa a estrutura de dados com todas as informações necessárias para criar uma instância da classe *Ontology* do OMVR. Essa classe ainda oferece um conjunto de métodos *sets* e *gets* para atualizar e recuperar informações dos seus atributos. Ela é utilizada principalmente pela página de submissão de ontologias do Portal OntoLP.

6.2.1.5 Classe *ReadSubmitPageOntologyEvaluationInformation*

Essa classe é responsável por carregar o arquivo XML com os dados de avaliação de uma ontologia, submetida por um usuário. Uma vez que o objeto do tipo *OntologyEvaluationInformation* inicializa todas as informações da avaliação, o mesmo é passado para a classe OMVR para ser criada uma nova instância da classe *OntologyEvaluation* do OMVR.

6.2.1.6 Classe *ReadSubmitPageOntologyInformation*

Essa classe é responsável por carregar o arquivo XML com os dados de uma ontologia, submetida por um usuário. Uma vez que o objeto do tipo *OntologyInformation* inicializa todas as informações da ontologia, o mesmo é passado para a classe OMVR, da API de programação, para ser criada uma nova instância da classe *Ontology* do metadados OMVR.

6.2.1.7 Classe *WriteOntologyEvaluationViewXML*

Essa classe é responsável por recuperar todas as avaliações de uma ontologia. As instâncias da classe *OntologyEvaluation*, referentes à ontologia selecionada pelo usuário, do OMVR serão gravadas em arquivo XML, o qual é utilizado pela camada de apresentação do Portal OntoLP para compor a interface de visualização das avaliações que uma ontologia possui.

6.2.1.8 Classe *WriteOntologyListXML*

Essa classe é responsável por recuperar todas as instâncias da classe *Ontology* do OMVR para gravá-las em arquivo XML, o qual é utilizado pela camada de apresentação do Portal OntoLP para compor a interface de navegação de ontologias descritas no OMVR.

6.2.1.9 Classe *WriteOntologyMetricViewXML*

Essa classe é responsável por recuperar todas as métricas instanciadas no OMVR para gravá-las em arquivo XML, o qual é utilizado pela camada de apresentação do Portal OntoLP para compor a interface de visualização das métricas calculadas para a ontologia selecionada pelo usuário.

6.2.1.10 Classe *WriteOntologyViewXML*

Essa classe é responsável por recuperar a instância da classe *Ontology* referente à ontologia selecionada pelo usuário para visualização das informações do OMVR. As informações sobre a ontologia selecionada são gravadas em arquivo XML, o qual é utilizado pela camada de apresentação do Portal OntoLP para compor a interface de visualização das informações da ontologia selecionada.

6.2.1.11 Classe *WriteOntologyXMLTree*

Essa classe é responsável por recuperar os nomes das ontologias, a partir das instâncias da classe *Ontology* do OMVR. As informações sobre a ontologia selecionada são gravadas em arquivo XML, o qual é utilizado pela camada de apresentação do Portal OntoLP para compor a árvore de navegação de ontologias descritas no OMVR. As classes do OMVR API, descritas anteriormente, são apresentadas no diagrama UML da Figura 9.

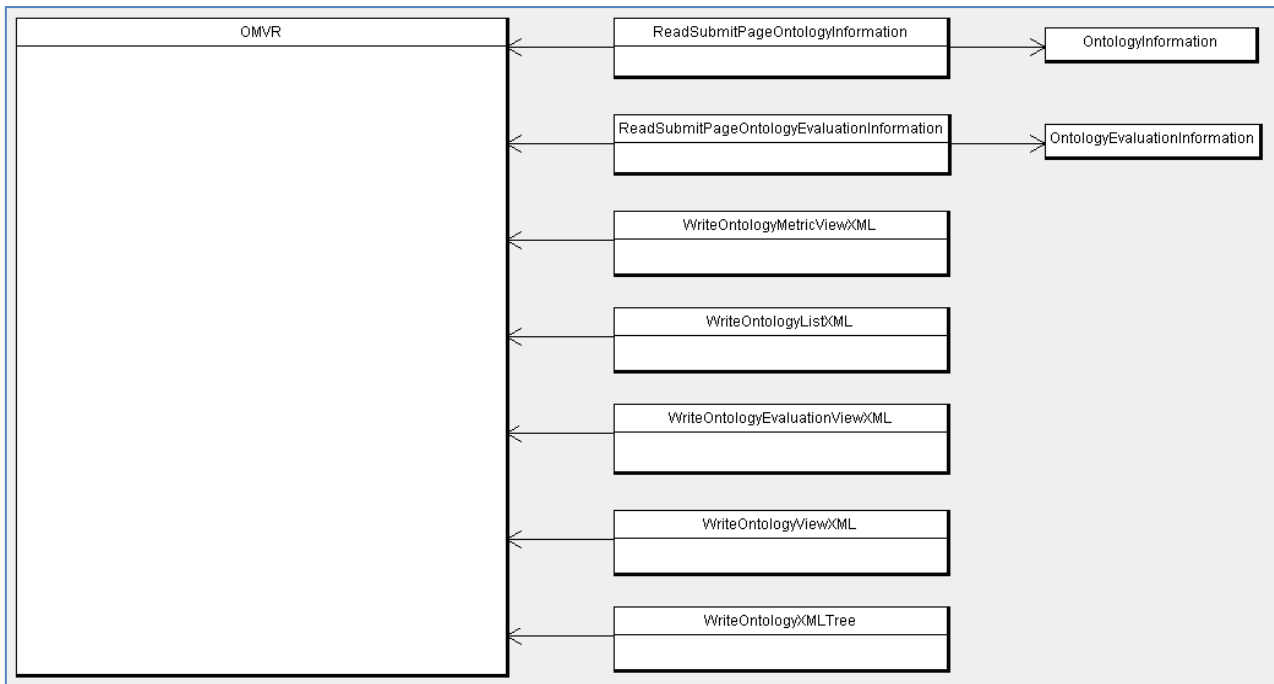


Figura 9 - Modelo UML do OMVR API

As próximas subseções apresentam as interfaces *Web* desenvolvidas para a implementação das novas funcionalidades do Portal OntoLP.

6.2.2 Especificação da Interface de Navegação sobre o OMVR

Uma vez que uma nova instância da classe *Ontology* do OMVR foi criada com as informações da ontologia, é necessário que o Portal OntoLP ofereça mecanismos para apresentação das informações das ontologia previamente instanciadas. Logo, o desenvolvimento de um conjunto de páginas para apresentação dos dados se faz necessário. Outro aspecto desejável é que o Portal OntoLP ofereça um mecanismo para execução de consultas avançadas sobre o conjunto pré-definido de propriedades oferecidas pelo OMVR. Tais consultas serão repassadas para o motor SPARQL, provido pelo Jena. Por fim, também é desejável que o Portal OntoLP ofereça páginas específicas para que os usuários possam realizar as avaliações das ontologias, bem como consultar as métricas calculadas, automaticamente, para cada ontologia descrita no OMVR.

6.2.2.1 Submissão de ontologias

A interface de submissão de ontologias é responsável por coletar as informações de descrição da ontologia, tais como, dados gerais da ontologia (nome, versão, data de criação, etc), informação sobre tipo de licenciamento usado para distribuição da ontologia,

informações sobre procedência, dados sobre aplicabilidade da ontologia e informações sobre reuso. Além disso, a interface solicita ao usuário o arquivo OWL da ontologia, o qual será armazenado no servidor do Portal OntoLP e processado pelo OMVR API para recuperação de metadados do arquivo. As informações obtidas diretamente do arquivo OWL são combinadas com os dados do formulário para criar uma nova instância da classe *Ontology* do OMVR. Além da nova instância da classe *Ontology*, são calculadas automaticamente vinte métricas para a ontologia. Esse conjunto de métricas também cria uma nova instância da classe *OntologyMetric*, para cada métrica calculada. No APÊNDICE C – PÁGINA DE SUBMISSÃO DE ONTOLOGIAS DO PORTAL ONTOLP-OMVR é apresentada a página de submissão de ontologias na íntegra. A Figura 10 apresenta a interface de submissão de ontologias do Portal OntoLP.

The screenshot shows the 'OntoLP - OMVR PORTAL DE ONTOLOGIA' website. The main navigation bar includes 'Home', 'Recursos', 'Sobre', 'Contato', and 'Links'. A sidebar on the left lists 'SERVIÇOS' with options for 'Submissão de Ontologia', 'Pesquisa de Ontologias', and 'Questionário'. The main content area is titled 'OntoLP OMVR | Submissão de Ontologia' and contains a form titled 'DADOS GERAIS SOBRE A ONTOLOGIA'. The form fields are: 'Nome da ontologia *' (text input), 'Versão *' (text input), 'Data de criação *' (date picker with 'dd/mm/aa' format), 'Descrição *' (large text area), 'Documentação' (text input), 'Referências bibliográficas' (large text area), 'Classes chaves' (text input), and 'Idioma de construção' (text input).

Figura 10 - Interface de submissão de ontologias.

6.2.2.2 Exploração de ontologias

Com o objetivo de apresentar para o usuário o conjunto de ontologias, descritas no OMVR, foi desenvolvida a página que as organiza em estrutura de árvore. A interface é constituída de duas áreas verticais onde o lado esquerdo apresenta os nomes das ontologias e, no lado direito, as informações disponíveis no OMVR sobre a ontologia selecionada. Ainda, ao selecionar uma ontologia do lado esquerdo, é possível visualizar

dois sub-nodos: (i) **Avaliações** para visualizar e realizar a avaliação da ontologia; (ii) **Métricas** para visualização das métricas. A Figura 11 apresenta a interface de exploração de ontologias.

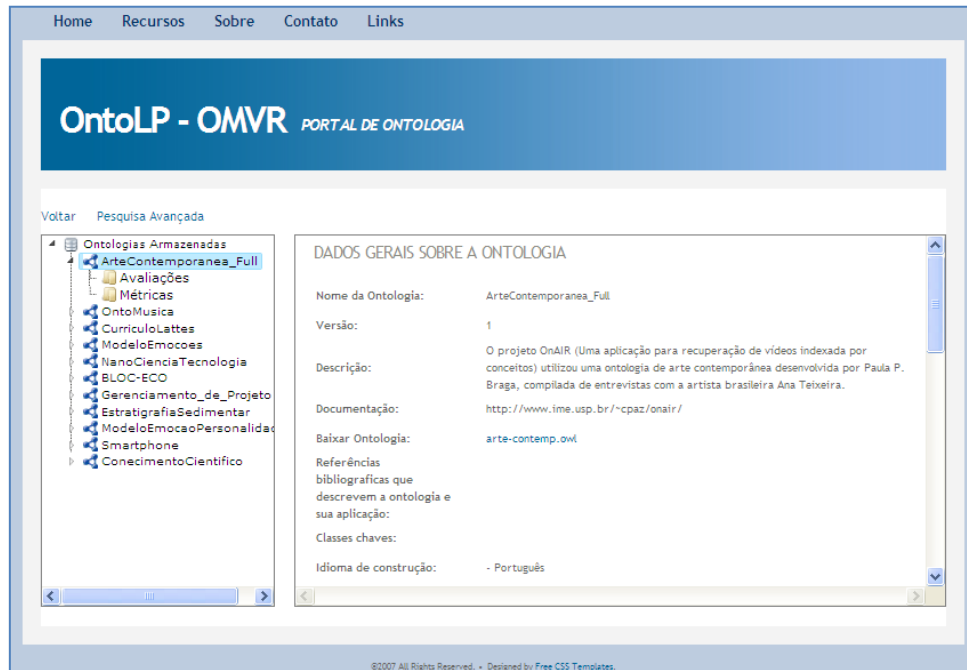


Figura 11 - Interface de exploração de ontologias.

6.2.2.3 Avaliação de ontologias

A avaliação de ontologia [YU09] permite ao usuário selecionar um conjunto de perguntas pré-definidas, as quais também estão instanciadas no OMVR (classe *EvaluationCriteria*). Cada avaliação submetida pela página cria uma nova instância da classe *OntologyEvaluation*, enriquecendo, dessa forma, as informações de descrição da ontologia selecionada. A Figura 12 apresenta a interface de avaliação de ontologias.

Home Recursos Sobre Contato Links

OntoLP - OMVR PORTAL DE ONTOLOGIA

Voltar Pesquisa Avançada

Ontologias Armazenadas

- ArteContemporanea_Full
- Avaliações**
- Métricas
- OntoMusica
- CurriculoLattes
- ModeloEmocoes
- NanoCienciaTecnologia
- BLOC-ECO
- Gerenciamento_de_Projeto
- EstratigrafiaSedimentar
- ModeloEmocaoPersonalidad
- Smartphone
- ConecimentoCientifico

Inserir nova avaliação

Questões de avaliação * A ontologia captura os conceitos do domínio?

Resposta textual

ou

Resposta numérica

Avaliador *

Instituição *

e-mail de contato *

Fone de contato *

Inserir

©2007 All Rights Reserved. • Designed by Free CSS Templates.

Figura 12 – Interface inserção de avaliação de ontologias.

O conjunto de avaliações da ontologia tem por objetivo oferecer aos usuários, interessados em reuso, informações que possam auxiliá-los na seleção da ontologia através de dados coletados sobre as avaliações da ontologia. Outra característica importante é que a interface de avaliação apresenta todas as avaliações que uma ontologia possui, conforme é apresentado na Figura 13.

Home Recursos Sobre Contato Links

OntoLP - OMVR PORTAL DE ONTOLOGIA

Voltar Pesquisa Avançada

Ontologias Armazenadas

- ArteContemporanea_Full
- Avaliações**
- Métricas
- OntoMusica
- CurriculoLattes
- ModeloEmocoes
- NanoCienciaTecnologia
- BLOC-ECO
- Gerenciamento_de_Projeto
- EstratigrafiaSedimentar
- ModeloEmocaoPersonalidad
- Smartphone
- ConecimentoCientifico

AVALIAÇÕES DA ONTOLOGIA

Questão de Avaliação: A ontologia tem instâncias claramente definidas?

Resposta textual: Sim, a ontologia tem instâncias bem definidas.

Resposta numérica:

Data da avaliação: 2009-12-17

Avaliado por: Renata Vieira

e-mail: renata.vieira@gmail.com

Instituição: PUCRS

Questão de Avaliação: A ontologia captura corretamente os conceitos do domínio?

Resposta textual: Sim, pela análise da ontologia ela captura de forma apropriada e abrangente o domínio que ela procura descrever.

Resposta numérica:

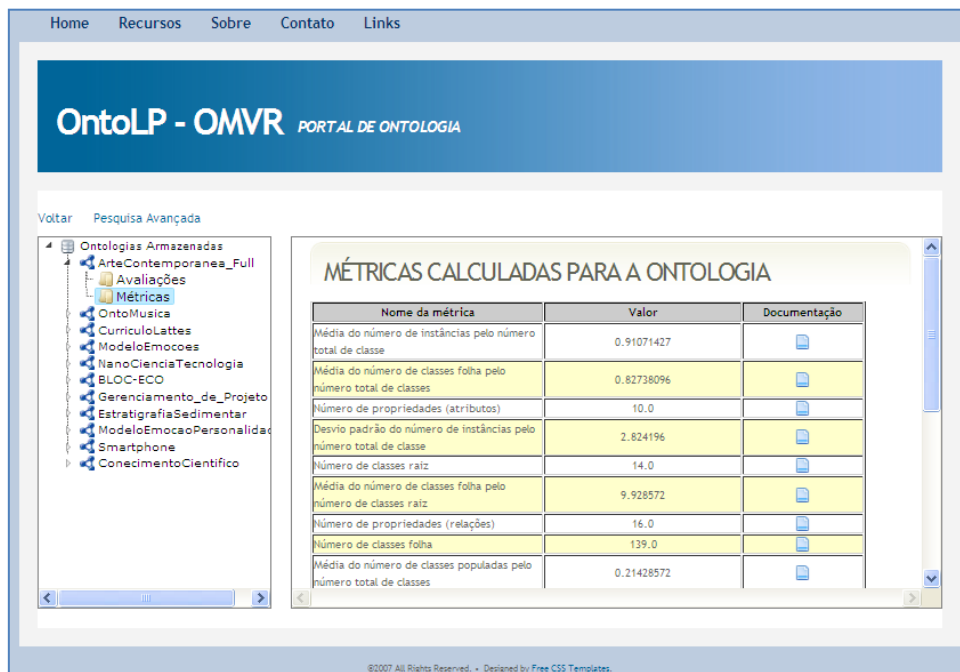
©2007 All Rights Reserved. • Designed by Free CSS Templates.

Figura 13 - Interface visualização das avaliações de ontologias.

Espera-se que com o tempo de uso do portal, as informações das avaliações possam auxiliar os usuários a selecionar ontologias a partir das observações de outros grupos de pesquisas e especialistas da área, conferindo às ontologias mais ou menos credibilidade a partir das avaliações feitas.

6.2.2.4 Visualização de métricas calculadas para as ontologias

A interface de visualização de métricas apresenta um conjunto de vinte métricas, calculadas automaticamente pelo portal no momento da sua submissão. As métricas calculadas são instâncias da classe *OntologyMetric* do OMVR. Assim como na avaliação de ontologias, as métricas calculadas oferecem ao usuário, interessado em reuso, a possibilidade de selecionar uma ontologia a partir de uma avaliação quantitativa e/ou qualitativa que as métricas oferecem [GAN06a] [YAN06] [CRO05]. A Figura 14 apresenta a interface de visualização de métricas calculadas para a ontologia selecionada.



The screenshot shows the 'OntoLP - OMVR PORTAL DE ONTOLOGIA' interface. On the left, there is a navigation menu with 'Ontologias Armazenadas' expanded to show 'ArteContemporanea_Full', 'Avaliações', 'Métricas', and 'OntoMusica'. The main content area is titled 'MÉTRICAS CALCULADAS PARA A ONTOLOGIA' and contains a table with three columns: 'Nome da métrica', 'Valor', and 'Documentação'. Each row in the table has a document icon in the 'Documentação' column.

Nome da métrica	Valor	Documentação
Média do número de instâncias pelo número total de classe	0.91071427	
Média do número de classes folha pelo número total de classes	0.82738096	
Número de propriedades (atributos)	10.0	
Desvio padrão do número de instâncias pelo número total de classe	2.824196	
Número de classes raiz	14.0	
Média do número de classes folha pelo número de classes raiz	9.928572	
Número de propriedades (relações)	16.0	
Número de classes folha	139.0	
Média do número de classes populadas pelo número total de classes	0.21428572	

Figura 14 - Interface de visualização das métricas calculadas para a ontologia.

Cabe salientar que cada métrica, apresentada pela interface, possui uma documentação associada, representado pelo ícone em forma de uma folha de papel. O objetivo da documentação é explicar ao usuário, de forma sucinta, qual o objetivo da métrica. A Figura 15 apresenta a caixa de texto com a explicação da métrica “Número de classes raiz”.

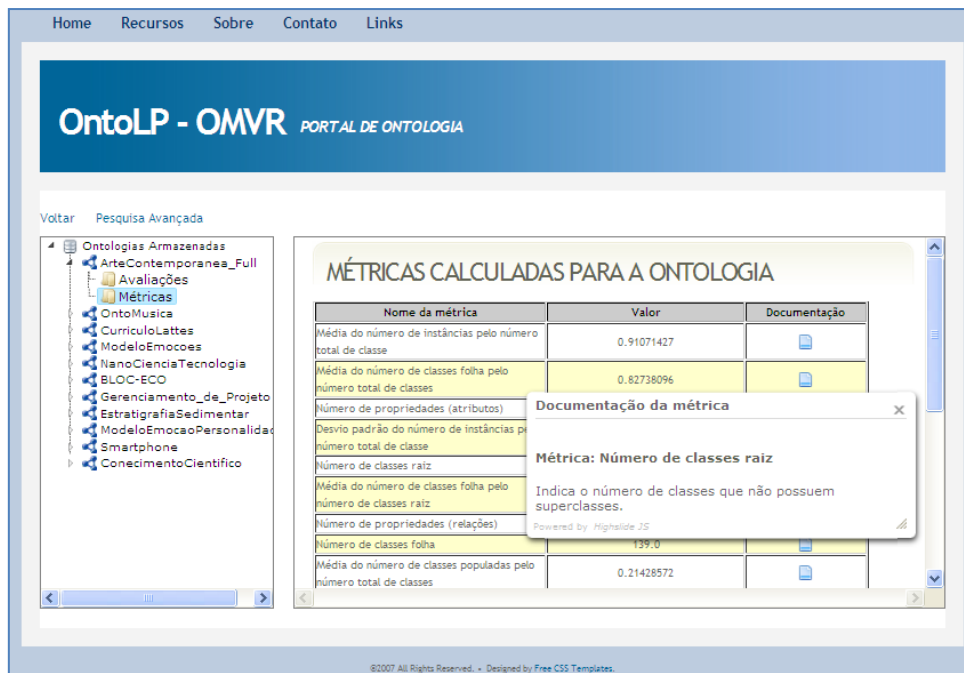


Figura 15 - Exemplo de visualização da documentação sobre a métrica selecionada.

Por fim, a interface de visualização de métricas oferece ao usuário a possibilidade de avaliar as métricas de forma gráfica. Para tanto, a página apresenta um *hyperlink*, identificado como “Visualizar gráfico”, para visualização do gráfico com as métricas calculadas, conforme apresentado na Figura 16.

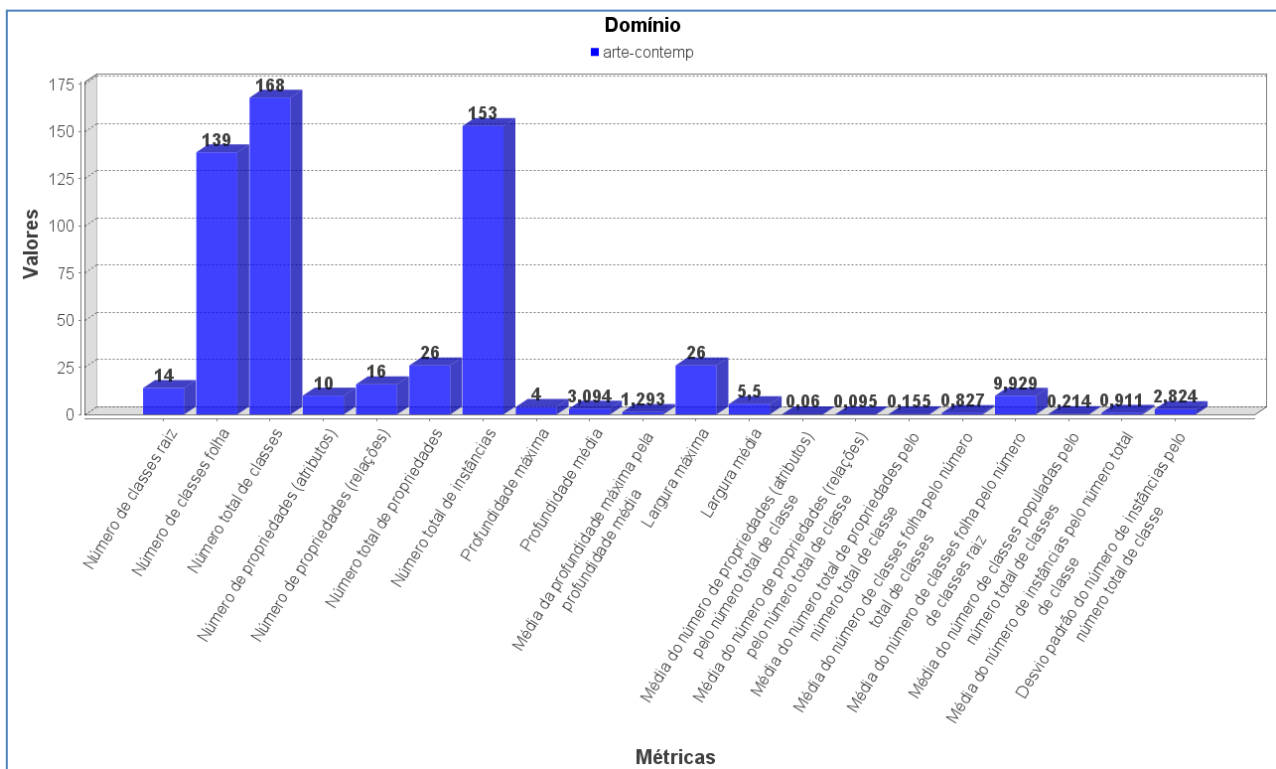


Figura 16 - Visualização do gráfico das métricas calculadas para a ontologia.

6.2.2.5 Interface de consultas avançada sobre o OMVR

A interface de consulta avançada permite ao usuário selecionar um conjunto de propriedades, pré-selecionadas, para executar a busca por ontologias, descritas no OMVR, que satisfaçam o critério de busca informado.

As opções de buscas são gravadas em um arquivo XML, o qual é usado como entrada pela rotina da OMVR API responsável pela execução da consulta. Essa rotina transforma os critérios de seleção em uma sentença SPARQL, a qual é repassada para o motor SPARQL do Jena. Por fim, o conjunto de ontologias selecionadas é gravado em um arquivo XML, o qual é apresentado pela camada de apresentação do Portal OntoLP. A Figura 17 apresenta a interface de consulta avançada.

Home Recursos Sobre Contato Links

OntoLP - OMVR PORTAL DE ONTOLOGIA

Pesquisa Avançada de Ontologias

Selecione os campos abaixo para compor um filtro de pesquisadas ontologias armazenadas no portal OntoLP-OMVR.
Após selecionar os campos de seu interesse, clique no botão Pesquisar para visualizar a página com as ontologias recuperadas pelo seu filtro de pesquisa.

Metodologia

- Cyg
- Grüninger Fox
- KACTUS
- Methontology
- On To Knowledge
- SENSUS
- Uschold King

Domínio

- Agricultura
- Arte
- Biologia
- Direito
- Ecologia
- Esportes
- Geologia
- Gerência de Projetos
- Medicina
- Música
- Nanociência
- Nanotecnologia
- Política
- Psicologia
- Telecomunicação
- Tecnologia da Informação

Projetos que utilizam a ontologia

Figura 17 - Interface de pesquisa avançada de ontologias.

6.3 Avaliação de uso do protótipo

Neste capítulo, será apresentada a avaliação das novas funções de descrição e recuperação de ontologias do Portal OntoLP propostas por este trabalho. Conforme descrito anteriormente no capítulo 4.2, o teste e avaliação do protótipo foram realizados através da aplicação de um questionário (ver APÊNDICE B – QUESTIONÁRIO), disponibilizado no Portal OntoLP. Os dados obtidos, ou seja, o perfil dos entrevistados e as respostas das perguntas, e as respectivas análises serão descritas nas subseções a seguir. Logo após, o próximo subcapítulo apresenta a cobertura da pesquisa realizada, a qual procurou atingir pesquisadores da área de ontologias, espalhados por todo o território brasileiro.

6.3.1 Perfil dos entrevistados

Da amostra obtida de 19 respondentes, 18 residem em diferentes estados brasileiros, sendo mais de 47.37% no Rio Grande do Sul e outros 36.84% na Região Sudeste, e 1 dos respondentes localizado na França, pesquisador vinculado ao INRIA (*Institut National de Recherche em Informatique et en Automatique*) Grenoble. A Figura 18 apresenta a distribuição de respondentes por estado (exceto FR que representa a França).

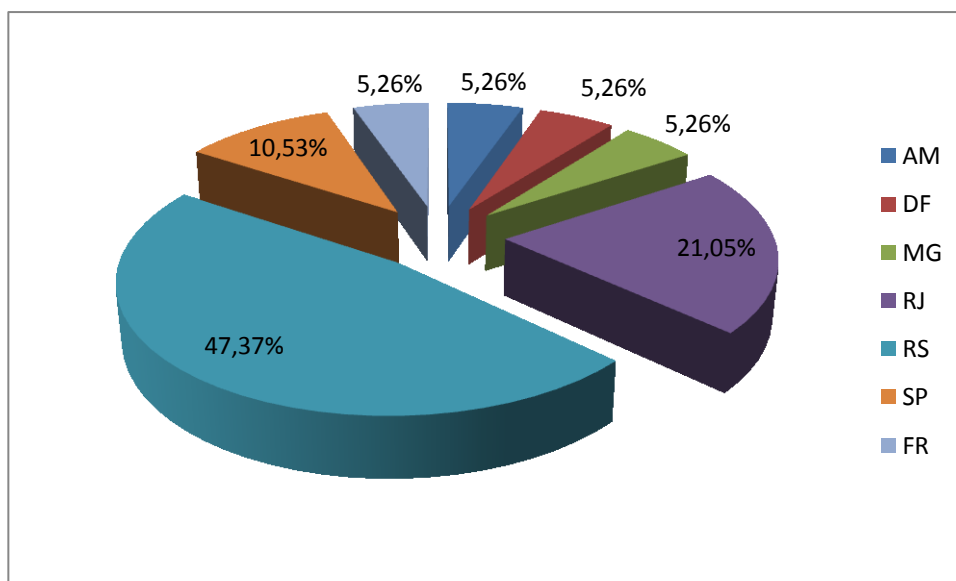


Figura 18 - Gráfico de distribuição de respondentes.

Todos os participantes da pesquisa atuam em universidades (ver Figura 19), mais de 31% nesta universidade, além disso, estão vinculados a cursos de Doutorado, Mestrado e Graduação, respectivamente, 36.84%, 47.37%, 15.79% (ver Figura 20).

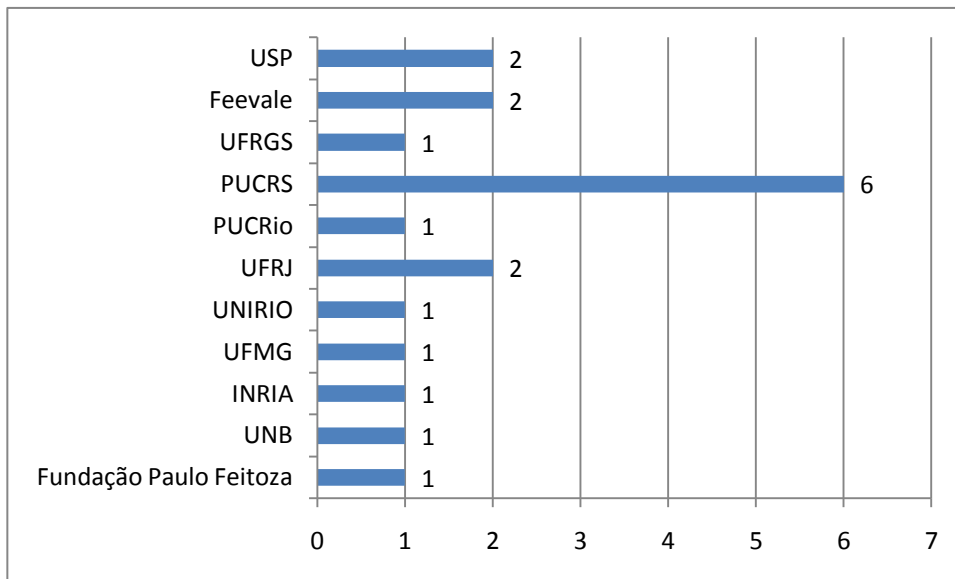


Figura 19 - Número de respondentes por universidade.

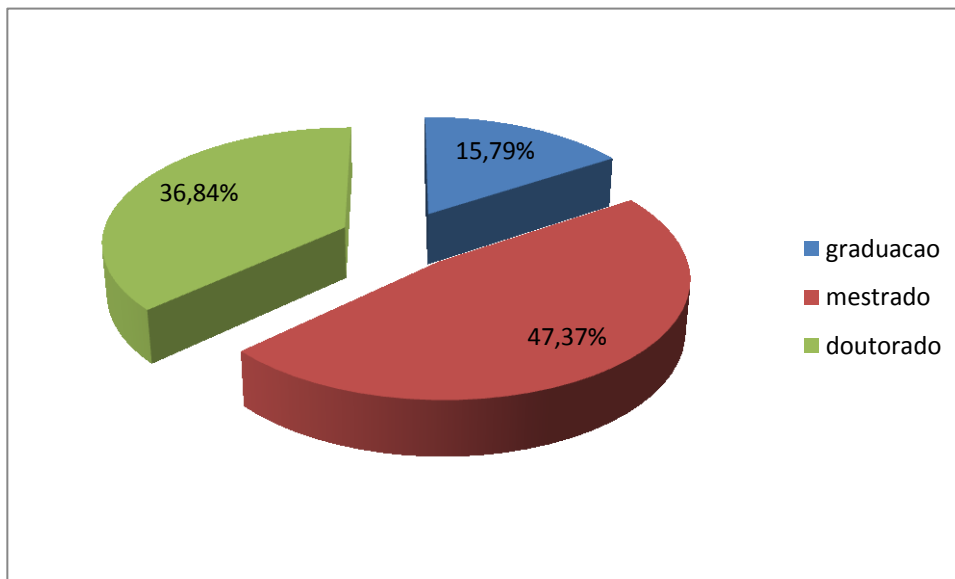


Figura 20 - Percentual de respondentes por educação formal.

Este subcapítulo apresentou o perfil dos participantes da *survey*, onde são identificados os estados, universidades e nível de educação formal dos respondentes. No próximo subcapítulo são apresentados os resultados e análise dos dados coletados.

6.3.2 Resultados da Survey

Conforme método apresentado no subcapítulo 4.2, os respondentes foram orientados a visitarem o Portal OntoLP e a avaliarem os novos serviços de submissão e pesquisa de ontologias, comparando-os com a versão vigente do portal. Ao serem questionados se esses novos serviços facilitaram o processo de busca por ontologias, aproximadamente 58% dos respondentes concordaram, 26% mantiveram-se neutros e pouco mais de 15% discordaram. O gráfico da Figura 21 apresenta a distribuição das respostas referentes à primeira pergunta do questionário: “*Os dados apresentados pelos serviços de submissão e pesquisa do novo Portal OntoLP-OMVR facilitaram o processo de busca de ontologias.*”.

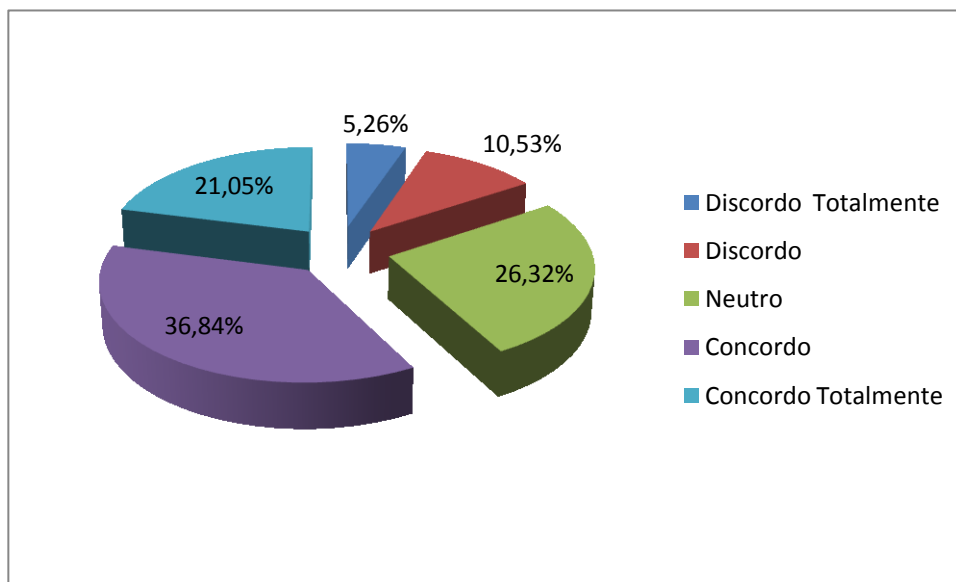


Figura 21 - Avaliação dos novos serviços de submissão e busca do Portal OntoLP.

Apesar de alguma discordância sobre a facilidade de busca por ontologias do Portal OntoLP, cerca de 95% dos respondentes acreditam que os novos serviços de submissão e busca de ontologias melhoraram em relação ao que era disponível anteriormente, como mostra o gráfico da Figura 22, que apresenta distribuição das respostas referentes à segunda pergunta do questionário: “*Os serviços de submissão e pesquisa do novo Portal OntoLP-OMVR são melhores do que os serviços disponíveis no Portal OntoLP atual.*”.

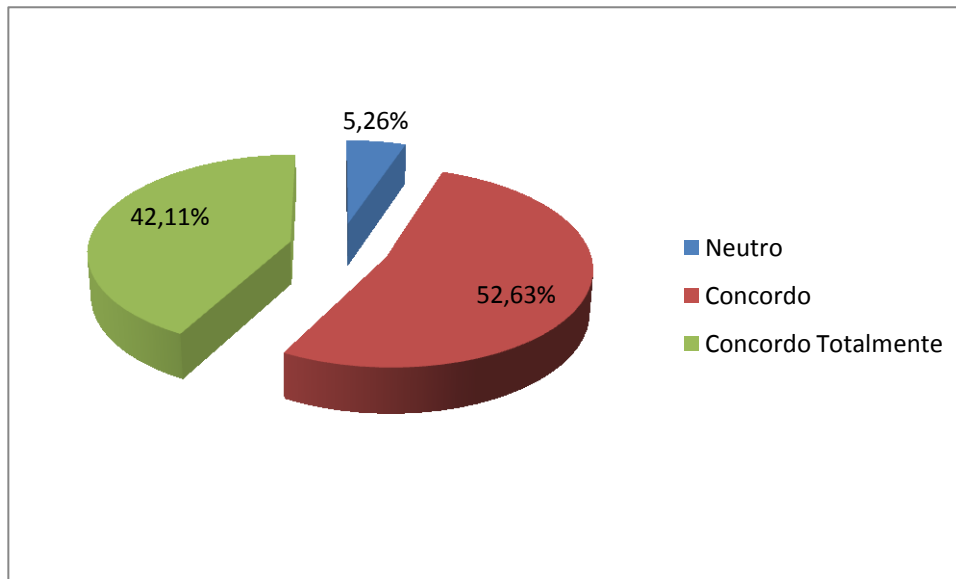


Figura 22 - Avaliação das novas funções em relação aos serviços vigentes do portal OntoLP.

Além das duas questões fechadas quanto à satisfação e à melhoria em relação à submissão e à busca por ontologias no Portal OntoLP, os respondentes também foram orientados a relacionar outras informações a serem incluídas no OMVR. Para tanto os respondentes deveriam responder duas questões abertas. A primeira questão procurou obter dos participantes a identificação de informações que poderiam ser incluídas no OMVR.

Apenas 26.32% dos respondentes contribuíram com sugestões; 32.58% alegaram que as informações contidas no modelo eram suficientes ou razoáveis ou ainda que não souberam identificar alguma informação que pudesse contribuir com o OMVR; e 36.84% não responderam. O gráfico da Figura 23 apresenta a distribuição das respostas para a pergunta: “*Que tipo de informação você acredita que falta no metadados do OMVR?*”

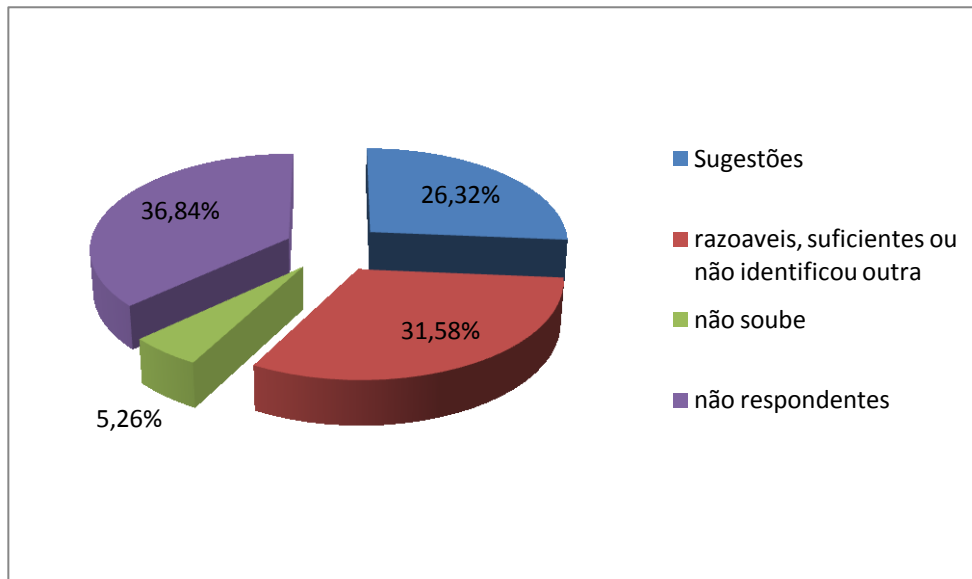


Figura 23 – Sugestões sobre a inclusão de novas informações no OMVR.

Pode-se inferir pela análise dos dados que o alto grau de não respostas resulta da não identificação de outras informações além das contidas no OMVR. Corroboram com essa observação os relatos dos seguintes respondentes:

- a) **2**: “Acredito que o conjunto de informações é suficiente para pesquisa de ontologias”;
- b) **9**: “Acho que as apresentadas são razoáveis.”;
- c) **14**: “Não detectei nenhuma informação importante que estivesse faltando.”, e;
- d) **16**: “Até o presente momento, as informações disponibilizadas cobrem as necessidades básicas.”.

As sugestões descritas pelos respondentes envolveram, na sua maioria, que a informação sobre o domínio da ontologia é um dos dados mais úteis para localização. Além disso, observa-se que são relacionados novos atributos e serviços a serem disponibilizados pelo Portal OntoLP. Um dos respondentes encontrou dificuldade em localizar o *hyperlink* do *download* da ontologia, conforme depoimento: “Onde eu posso fazer o download das ontologias disponíveis? Isso não está fácil de encontrar ao contrário do antigo Portal OntoLP” (Respondente 8).

Outro respondente alegou a necessidade da hierarquização e de relacionamento das ontologias no portal, dizendo: “Em princípio diria que em algum momento precisaremos de uma hierarquização (quando haja mais de 20 ou 30 ontologias) e algum mecanismo para relacioná-las (meta-ontologia)” (Respondente 13).

Além disso, também foi sugerido o uso da metodologia OntoUML, conforme escreveu o respondente 12:

Creio que no item sobre metodologia, deveria haver algo para contemplar o desenvolvimento baseado na UFO e nas ferramentas o ontoUML, pois diversos trabalhos no Brasil e no exterior estão agora considerando estas diretrizes e os padrões do OntoUML (Respondente 12).

Os respondentes ainda foram questionados sobre quais são as informações mais úteis para localização de ontologias. Os 16 respondentes identificaram 27 diferentes tipos de informações (ver Figura 24), dos quais o mais citado foi o **domínio** da ontologia, relacionada por 14 respondentes; seguido do **paradigma** de representação de conhecimento usado para construção da ontologia, relatado por 5 respondentes; do **tipo**, da **linguagem** e da **metodologia** de construção da ontologia, descritos por 3 respondentes; e as informações sobre a **classificação**, o **tema** e a **instituição apoiadora** das ontologias, cada uma delas citadas por 2 respondentes.

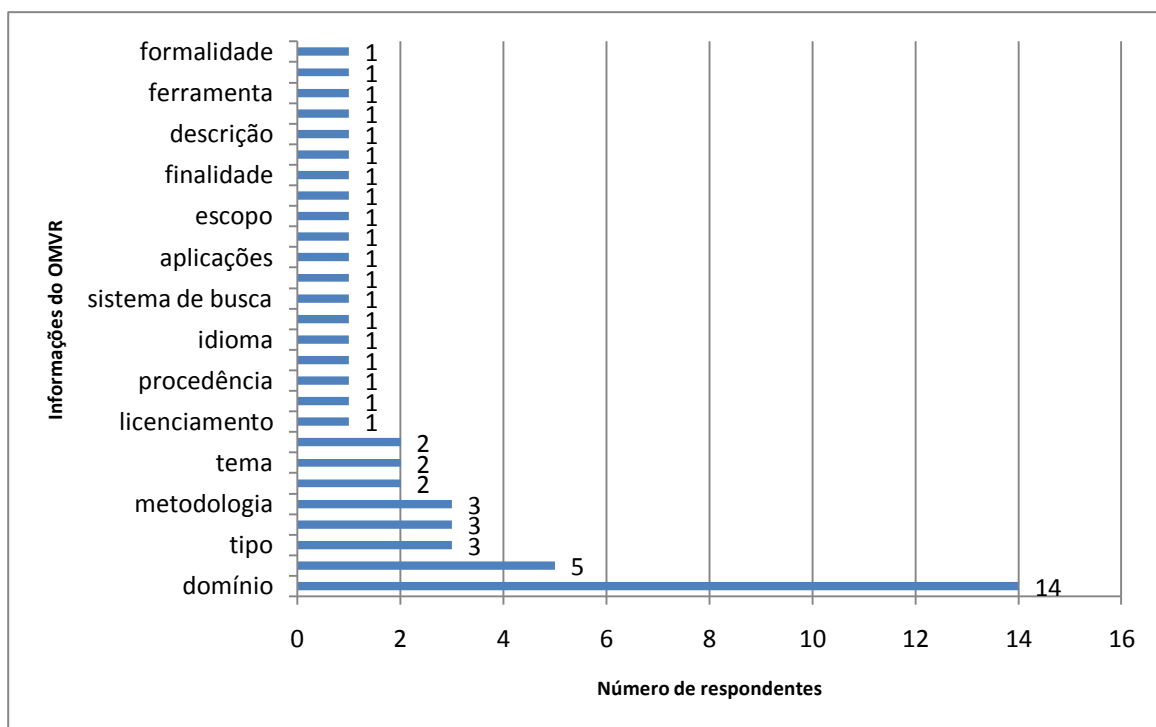


Figura 24 - Ranking das informações mais úteis para localização de ontologias, segundos os respondentes.

Outro ponto importante a observar é que alguns respondentes atribuíram uma ordem de importância das informações sobre ontologias. O respondente 4 descreveu o “domínio e o paradigma, como os principais, seguidos de instituição, autores, ano e metodologia, para verificar a confiança/credibilidade na ontologia”. Por sua vez, o respondente 19 relacionou as informações na seguinte “ordem de importância (do mais

importante para o menos importante): licenciamento, idioma, tipo, domínio, ferramenta, nível de formalidade, paradigma”.

Pode-se, ainda, atribuir um grau de importância de acordo com a ordem em que os tipos de informações sobre ontologias foram citadas, cabendo dizer que cada respondente identificou de 1 a 7 tipos de informações diferentes. Assim, ao agrupar as respostas dos participantes da *survey*, as informações foram citadas na seguinte ordem:

- 1º. domínio, licenciamento, palavras-chave, paradigma, procedência, tipo;
- 2º. classes, classificação, idioma, instâncias, linguagem, metodologia, paradigma, sistema de busca²², tamanho, tema;
- 3º. expressividade, finalidade, instituição, metodologia, paradigma, tipo;
- 4º. aplicações, autores, descrição, domínio, linguagem;
- 5º. ano, ferramenta, instituição, métrica;
- 6º. formalidade, metodologia; e,
- 7º. paradigma.

Esse subcapítulo apresentou a análise dos dados coletados na *survey*. Foram apresentados os gráficos de cada uma das quatro perguntas que compõem o questionário sobre a avaliação dos novos serviços de descrição e recuperação de ontologias do Portal OntoLP, bem como os resultados obtidos de cada análise realizada. No próximo capítulo serão apresentadas as considerações finais deste trabalho

²² A busca avançada já está implementada no Portal OntoLP.

7 CONSIDERAÇÕES FINAIS E TRABALHOS FUTUROS

A utilização de um modelo de metadados consolidado para descrição e recuperação de ontologias tem uma importância significativa no sentido de promover o reúso de ontologias. Um dos benefícios que pode ser ressaltado é que o emprego de um modelo de metadados possibilita aos engenheiros de ontologias centralizarem o armazenamento da descrição das mesmas em um ambiente organizado e controlado.

Este trabalho apresentou uma adaptação do modelo de metadados OMV para suportar mecanismos de descrição e recuperação de ontologias em língua portuguesa que foram incorporados no Portal OntoLP. A adaptação do OMV visou complementá-lo com novas informações para descrição e recuperação de ontologias. Com base em [SIM09], [YU09], [TAR07], [GAN06a], [YAN06], [FRE09], foram identificados um conjunto informações relevantes para complementação do OMV, a saber: (i) Método de avaliação de ontologias; e (ii) Métricas. Além disso, como requisito do projeto do Portal OntoLP, considerou-se a rastreabilidade do uso de ontologias em projetos. Por esse motivo, o novo modelo, chamado de OMVR, oferece a possibilidade de que sejam instanciados projetos, assim como quais ontologias são utilizadas por esses projetos. A idéia por trás disso é determinar o grau de maturidade de uma ontologia em função do número de projetos que as utilizam. Por fim, no que diz respeito à avaliação de ontologias, espera-se que com o tempo de uso do Portal OntoLP as informações das avaliações venham a se tornar úteis para a comunidade, pois também constitui um instrumental importante para auxiliar no processo de seleção de ontologias.

Uma vez concluída a construção do OMVR, foram desenvolvidos os programas para manipulação da nova ontologia. Esses programas reunidos formam o OMVR API, que utiliza o *framework* Jena para manipular adequadamente o OMVR. Além disso, o OMVR API também serve de suporte para camada de apresentação do Portal OntoLP, a qual reúne, em um ambiente integrado, os novos serviços de descrição e recuperação de ontologias, quais sejam: submissão de ontologias, pesquisa simples e avançada de ontologias, avaliações e métricas calculadas para cada ontologia descrita no modelo de metadados OMVR. Especificamente sobre a pesquisa avançada, cabe salientar que o OMVR API utiliza o motor de execução da linguagem SPARQL para consultas em arquivos OWL.

Com o objetivo de avaliar o protótipo do novo Portal OntoLP, foi realizada uma pesquisa do tipo *survey*, onde foram elencadas algumas questões específicas sobre os novos serviços, bem como questões que pudessem identificar se houve algum ganho, em termos de facilidade de localização de ontologias em língua Portuguesa, quando comparado o novo portal com o portal vigente. Foram selecionados para participar do *survey*, pesquisadores, especialistas, doutorandos, mestrandos e graduandos ligados à área de ontologias.

Conforme análise apresentada no subcapítulo 6.3.2, pode-se concluir que os novos serviços do Portal OntoLP contribuem no processo de descrição e recuperação de ontologias. Além disso, outra contribuição são as novas informações disponíveis no OMVR, porque auxiliam os usuários a identificar ontologias através das avaliações e das métricas instanciadas para cada ontologia descrita no modelo. Por fim, pode-se dizer que esse novo conjunto de informações e serviços também favorece o reúso de ontologias.

Ainda sobre a análise, identificaram-se algumas informações que poderiam ser agregadas ao OMVR como, por exemplo, palavras chaves e hierarquização das ontologias. Sobre os serviços, também existem sugestões interessantes como a possibilidade de fazer buscas de ontologias por palavras chaves.

Este trabalho procurou contribuir para construção de mecanismos de descrição e recuperação de ontologias, que tiveram uma aplicação prática no portal de divulgação de ontologias em língua Portuguesa. Para tanto, foram executadas as seguintes atividades:

- a) Identificação de um modelo de metadados robusto sobre ontologias;
- b) Revisão da literatura sobre formas de avaliar ontologias, armazenamento em repositórios e métricas;
- c) Adaptação do modelo de metadados OMV baseado nas práticas estudadas durante a revisão da literatura;
- d) Implementação de uma API de programação para manipulação do novo modelo de metadados, chamado OMVR;
- e) Implementação de um mecanismo para avaliação das ontologias descritas no portal, e;
- f) Implementação de um protótipo do Portal OntoLP para prova de conceito.

Com base nas lições aprendidas durante o desenvolvimento do OMVR e do arcabouço de software desenvolvido, identificam-se alguns pontos para futuras pesquisas, a saber:

- a) Adaptação do OMVR API para trabalhar com o modelo OMVR armazenado em banco de dados relacional, a fim de obter escalabilidade em número de acessos ao portal e armazenamento das descrições de novas ontologias;
- b) Proposta de construção de serviços *Web* para que o Portal OntoLP possa ser utilizado em aplicações distribuídas que tenham como objetivo a descoberta de ontologias;
- c) Melhora da interface de pesquisa avançada de ontologias, permitindo que o usuário possa dispor de uma interface que o oriente a selecionar propriedades para execução da pesquisa;
- d) Proposta de construção de mecanismos para visualização das ontologias descritas no modelo de metadados OMVR;

Por fim, ainda há oportunidades de pesquisas no que diz respeito à padronização de metadados para descrição de ontologias, bem como o desenvolvimento de softwares que possam auxiliar os pesquisadores e engenheiros da área a localizarem ontologias de forma econômica e eficiente.

REFERENCIAL BIBLIOGRÁFICO

- [BEC08] Bechhofer, S. "OWL Web Ontology Language Reference". Capturado em: <http://www.w3.org/TR/2004/REC-owl-ref-20040210/#Intro>, Maio 2008.
- [BER01] Berners-Lee, T.; Hendler, J.; Lassila, O. "The Semantic Web". *Scientific America*, vol. 284, 2001, pp. 34-43.
- [BRE05] Breitman, K. K. "Web Semântica: A Internet do Futuro". Rio de Janeiro: LTC, 2005, 212p.
- [BON05] Bontas, E. P.; Mochol, M.; Tolksdorf, R. "Case studies on ontology reuse". In: Proceedings of I-KNOW, Graz, Austria, 2005, pp.345-353.
- [CAN07] Cantador, I.; Fernández, M.; Castells, P. "Improving Ontology Recommendation and Reuse in WebCORE by Collaborative Assessments". In: 16th International World Wide Web Conference, Banff, Alberta, Canada, 2007, 10p.
- [CRA07] Cranefield, S.; Pan, J. "Bridging the Gap Between the Model-Driven Architecture and Ontology Engineering". *International Journal of Human-Computer Studies*, ed. Elsevier, vol. 65-7, Jul. 2007, pp. 595-609.
- [CRO05] Cross, V.; Pal, A. "Metrics for ontologies". In: Fuzzy Information Processing Society, Annual Meeting of the North America, 2005, pp. 448-453.
- [DIN04] Ding, L.; Finin, T.; Joshi, A.; Pan, R.; Cost, R. S.; Peng, Y.; Reddivari, P.; Doshi, V.; Sachs, J. "Swoogle: A Semantic Web Search and Metadata Engine". In: Proceedings of the thirteenth ACM international conference on Information and knowledge management, Washington, D.C., USA, 2004, pp. 652-659.
- [FEN06] Fensel, D.; Lausen, H.; Polleres, A.; Bruijn, J. de; Stollberg, M.; Roman, D.; Domingue, J. "Enabling Semantic Web Services – The Web Service Modeling Ontology". Springer, 2006, 191p.
- [FRE09] Freitas, L. A. "Métricas para Ontologias: Revisão Sistemática e Aplicação ao Portal OntoLP", Dissertação de Mestrado, Programa de Pós-Graduação em Ciência da Computação, PUCRS, 2009.

- [GAN06a] Gangemi, A.; Catenacci, C.; Ciaramita, M.; Lehmann, J. "Modelling Ontology Evaluation and Validation". In: *Proceedings Semantic Web: Research and Applications*, Budva, Montenegro, 2006, pp. 140-154.
- [GAN06b] Ganjisaffar, Y.; Abolhassani, H.; Neshati, M.; Jamali, M. "A Similarity Measure for OWL-S Annotated Web Services". In: IEEE Computer Society. *Proceedings of the 2006 IEEE/WIC/ACM International Conference on Web Intelligence*, Hong Kong, China, 2006, pp. 621-624.
- [GAN06c] Gannod, G. C. "Facilitating the Specification of Semantic Web Services Using Model-Driven Development". *International Journal of Web Services Research*, ed. Idea Group Inc., vol. 3, Jul-Set 2006, pp. 61-81.
- [GIB07] Gibson, A.; Wolstencroft, K.; Stevens, R. "Promotion of Ontological Comprehension: Exposing Terms and Metadata with Web 2.0". In: *16th International World Wide Web Conference*, Banff, Alberta, Canada, 2007, 9p.
- [GRU08] Gruber, T. R. "What is an Ontology?". Capturado em: <http://www.ksl.stanford.edu/kst/what-is-an-ontology.html>, Junho 2008.
- [HAR05] Hartmann, J.; Palma, R.; Sure, Y.; Suarez-Figueroa, M. C.; Haase, P.; Gómez-Pérez, A.; Studer, R. "Ontology Metadata Vocabulary and Applications", *On the Move to Meaningful Internet Systems 2005: Otm 2005 Workshops, Proceedings*, ed. Springer, vol. 3762, 2005, pp. 906-915.
- [HAR06] Hartmann, J.; Bontas, E. P.; Palma, R.; Gómez Pérez, A. "Demo - Design Environment for Metadata Ontologies". In: *The Semantic Web: Research and Applications*, Budva, Montenegro, 2006, pp. 427-441.
- [IEE08] IEEE. "Standard Upper Ontology Working Group (SUO WG)". Capturado em: <http://suo.ieee.org/>, Junho 2008.
- [IMB09] IBM. "Introduction to Jena". Capturado em: <http://www.ibm.com/developerworks/java/library/j-jena/>, Junho 2009.
- [MAE01] Maedche, A., Staab, S. "Ontology Learning for the Semantic Web". *IEEE Intelligent Systems*, ed. IEEE Journals, vol. 16, 2001, pp. 72-79.
- [MAR95] March, S.; Smith G. "Design and Natural Science Research on Information Technology". *Decision Support System*, vol. 15-4, 1995, pp. 251-266.
- [MCB02] McBride, B. "Jena: A Semantic Web Toolkit", *IEEE Internet Computing*, ed. IEEE Journals, vol. 6, 2002, pp. 55-59.

- [MCG08] McGuinness, D. L. "OWL Web Ontology Language Overview". Capturado em: <http://www.w3.org/TR/2004/REC-owl-features-20040210/>, Maio 2008.
- [MIN74] Minsky, M. "A Framework for Representing Knowledge", Relatório Técnico, Massachusetts Institute of Technology AI Laboratory, 1974, 82p.
- [NUN91] Nunamaker, J.F.; Chen, M.; Purdin, T. D. M. "Systems Development in Information Systems Research". *Journal Management Information Systems*, ed. M. E. Sharpe Inc., vol.7-3, Dec-Jan 1990-1991, pp. 89-106.
- [SCB09] Stanford Center for Biomedical Informatics Research. "The Protégé Ontology Editor and Knowledge Acquisition System". Capturado em: <http://protege.stanford.edu/>, Junho 2009.
- [NOY04] Noy, N. F.; Rubin, D. L.; Musen, M. A. "Making Biomedical Ontologies And Ontology Repositories Work", *IEEE Intelligent Systems*, ed. IEEE Journals, Nov-Dec 2004, pp. 78-81.
- [OAT06] Oates, B. J. "Researching Information Systems and Computing". London: SAGE Publications Ltd, 2006, 326p.
- [PAL06] Palma, R.; Hartmann, J.; Gómez-Pérez, A. "Towards an Ontology Metadata Standard". In: 3rd European Semantic Web Conference, Budva, Montenegro, 2006, 2p.
- [PAL08] Palma, R.; Hartmann, J.; Haase, P. "OMV Report 2.4 - Ontology Metadata Vocabulary for the Semantic Web", Technical Report, 2008, OMV Consortium, 94p.
- [SIM09] Simperl, E. "Reusing Ontologies on the Semantic Web: A Feasibility Study". *Data & Knowledge Engineering*, ed. Elsevier, vol. 68-10, Out. 2009, pp. 905-925.
- [YAN06] YANG, Z.; Zhang, D.; YE, C. "Evaluation Metrics for Ontology Complexity and Evolution Analysis". In: e-Business Engineering IEEE International Conference, Shanghai, China, 2006, pp. 162-170.
- [YU09] Yu, J.; Thom, J.A.; Tam, A. "Requirements-Oriented Methodology for Evaluating Ontologies". *Information Systems*, ed. Elsevier, vol. 34-8, Dec. 2009, pp. 766-791.

- [TAR07] Tartir, S.; Arpinar, I.B. "Ontology Evaluation and Ranking Using OntoQA". In: International Conference on Semantic Computing, Irvine, California, USA, 2007, pp. 185-192.

APÊNDICE A – PÁGINA DE CONTATO DO PORTAL ONTOLP

OntoLP PORTAL DE ONTOLOGIA

Contato

Nome *

E-mail *

Grupo

Instituição*

Projeto

Site

Possui algum material que gostaria de disponibilizar no portal? Sim Não

Tipo do recurso:

- Ontologia
- Ferramenta
- Outro

DETALHES DA ONTOLOGIA:

Título

Domínio*

Idioma

Especialista do Domínio*

Linguagem

Aplicação

Última Atualização (mm/aaaa)*

Ferramenta Utilizada

URL

ENVIO DA ONTOLOGIA:

DETALHES DA FERRAMENTA:

Nome

Linguagem

Propósito

Licença

URL

ENVIO DA FERRAMENTA:

DETALHES DO RECURSO:

Descrição do Recurso

URL do Recurso

ENVIO DO RECURSO:

REFERÊNCIAS

Existe alguma referência bibliográfica que gostaria de associar aos recursos disponibilizados?

Comentários

*campos obrigatórios

APÊNDICE B – QUESTIONÁRIO

Questionário Sobre os Novos Serviços do Portal OntoLP

Este questionário tem como objetivo avaliar a facilidade de uso e a abrangência dos novos serviços de descrição e recuperação de ontologias do portal OntoLP.

Esses novos serviços são baseados em uma ontologia chamada OMV (*Ontology Metadata Vocabulary*) que foi projetada para ser um vocabulário para descrição de ontologias. Ainda, o grupo de pesquisa de Processamento da Linguagem Natural (PLN) da PUCRS estendeu o OMV (extensão denominada OMVR) com o objetivo de enriquecer o conjunto de informações de metadados sobre ontologias, tais como métricas, avaliações e projetos que utilizam as ontologias armazenadas no Portal OntoLP-OMVR. (ver [descrição do metadados](#))

Para tanto, responda as questões abaixo sobre os serviços de **Submissão** e **Pesquisa** de ontologias do novo portal OntoLP OMVR, comparando-o com os serviços, anteriormente, disponíveis nos links **Contato** e **Recursos** do Portal OntoLP.

Responda as questões 1 e 2, atribuindo um valor de 1 a 5 conforme o grau de concordância com as afirmativas abaixo.

1. Os dados apresentados pelos serviços de submissão e pesquisa do novo Portal OntoLP-OMVR facilitaram o processo de busca de ontologias.

Níveis de relevância:

1 Discordo totalmente; 2 Discordo; 3 Neutro; 4 Concordo; 5 Concordo totalmente.

1 2 3 4 5

2. Os serviços de submissão e pesquisa do novo Portal OntoLP-OMVR são melhores do que os serviços disponíveis no Portal OntoLP atual.

Níveis de relevância:

1 Discordo totalmente; 2 Discordo; 3 Neutro; 4 Concordo; 5 Concordo totalmente.

1 2 3 4 5

Utilize as caixas de texto para responder as questões 3 e 4.

3. Que tipo de informação você acredita que falta no metadados do OMVR?

4. Quais informações são mais úteis para busca de ontologias?

5. Dados pessoais

País:

Estado:

Instituição:

Grau de instrução:

APÊNDICE C – PÁGINA DE SUBMISSÃO DE ONTOLOGIAS DO PORTAL ONTOLP-OMVR

Home Recursos Sobre Contato Links

OntoLP - OMVR PORTAL DE ONTOLOGIA

SERVIÇOS

- Submissão de Ontologia
- Pesquisa de Ontologias
- Questionário

OntoLP OMVR | Submissão de Ontologia

DADOS GERAIS SOBRE A ONTOLOGIA

Nome da ontologia *

Versão *

Data de criação *

Descrição *

Documentação

Referências bibliográficas

Classes chaves

Idioma de construção

INFORMAÇÃO SOBRE DISPONIBILIDADE DA ONTOLOGIA

Tipo de licenciamento

INFORMAÇÕES DE PROCEDÊNCIA DA ONTOLOGIA

Concebido por *

Colaborador(es)

Instituição apoiadora

Paradigma Lógica de Descrição
 Frames
 Outro

Metodologia Cyc
 Grüninger Fox
 KACTUS
 Methontology
 On To Knowledge
 SENSUS
 Uschold King
 Outra

Ferramenta Altova Semantic Works
 IsisViz
 OEd
 Onto Builder
 Onto Studio
 Protégé
 SWOOD
 WebODE
 WSMO Studio
 Outra

INFORMAÇÕES SOBRE APLICABILIDADE

Nível de formalidade

Tipo

Projetada para Tarefas de Anotação
 Tarefas de Configuração
 Tarefas de Filtro
 Tarefas de Indexação
 Tarefas de Integração
 Tarefas de Matching
 Tarefas de Mediação
 Tarefas de Personalização da Informação
 Tarefas de Formulação de Consultas
 Tarefas de Busca
 Outra Tarefa

Domínio

Projetos que a utilizam

INFORMAÇÕES SOBRE REUSO

Versão anterior

Compatível com

Baseada em quais ontologias?

INFORME O NOME DO ARQUIVO OWL/RDFS PARA ENVIO