

Portal de ontologias em língua portuguesa: métodos de descrição e acesso

Anderson Bestteti¹, Larissa A. de Freitas¹, Patrícia Pizzinato², Renata Vieira¹

¹ Faculdade de Informática – Pontifícia Universidade Católica do Rio Grande do Sul
Avenida Ipiranga, 6681 – 90619-900 – Porto Alegre – RS – Brazil

² Instituto de Informática – Unisinos
Avenida Unisinos, 950 – 93022-000 – São Leopoldo – RS – Brazil
{anderson.bestteti,larissa.freitas,renata.vieira}@pucrs.br
patricia.pizzinato@gmail.com

Resumo This paper presents the portal for Portuguese ontologies, OntoLP, developed at PUCRS.

1 Introdução

A pesquisa na área de ontologias ganhou novo fôlego com o Projeto da *Web Semântica* [1]. A comunidade científica tem mostrado grande interesse a respeito deste assunto, porém o desenvolvimento desta área no Brasil ainda é recente.

Consideramos que algumas áreas específicas do conhecimento podem se beneficiar de ontologias em língua portuguesa, como é o caso de governo eletrônico, saúde pública, cultura, entre outras, pois são áreas do conhecimento ligados ao idioma local.

A construção de ontologias é um processo custoso e demorado, que requer a presença de especialistas do domínio a ser modelado. Devido a esta dificuldade, pesquisadores têm estudado questões relacionadas ao reuso, construção automática e mapeamento de ontologias.

Em relação ao reuso, os passos envolvidos são, segundo [2]: encontrar as ontologias candidatas; avaliá-las e selecionar aquelas que serão usadas na ontologia alvo; analisá-las e compará-las; adaptar as ontologias escolhidas; traduzi-las em formas de representação e extrair sub-ontologias relevantes; integrar (ou fundir) as fontes adaptadas na ontologia alvo. Portanto, no processo do reuso, um dos primeiros problemas enfrentados é o de encontrar ontologias. Este problema é ainda mais crítico ao se buscar ontologias em língua portuguesa, pois ainda são raras.

A extração automática de textos e o mapeamento de ontologias são processos que envolvem, muitas vezes, tecnologias dependentes de língua. Para avançar nas pesquisas nesses temas é necessária a disponibilização de recursos em língua portuguesa.

O projeto OntoLP, desenvolvido no Programa de Pós-graduação em Ciência da Computação na FACIN-PUCRS, se propõe a colaborar no preenchimento destas lacunas, disponibilizando ontologias e sua descrição de modo a facilitar o

acesso a ontologias em língua portuguesa, bem como a outros recursos e ferramentas relacionadas a pesquisas na área. Este projeto é financiado pelo CNPq através do Edital Desafios da Computação no Brasil. O portal está em desenvolvimento e já abriga um conjunto de recursos e informações.

O conteúdo deste artigo é organizado como segue: na Seção 2 introduzimos o portal OntoLP e na Seção 3 expomos as considerações finais e os trabalhos futuros.

2 O portal OntoLP

Entre os objetivos do projeto OntoLP³ podemos citar os seguintes: realizar um levantamento de recursos existentes, organizá-los, desenvolver um portal de recursos para pesquisa em ontologias em língua portuguesa e as publicações relacionadas.

O portal apresenta atualmente uma área *home* ou principal onde podemos encontrar uma breve descrição do propósito do portal; uma área *recursos* onde estão armazenadas as ontologias e os trabalhos relacionados a essas; uma área *sobre* que contém informações a respeito do grupo de pesquisa envolvido neste projeto e seus integrantes; uma área *contato* na qual outros grupos, pesquisadores, interessados no tema podem enviar suas ontologias e suas sugestões; e finalmente, uma área *links* onde estão elencados grupos de pesquisa e pesquisadores que colaboram direta ou indiretamente com o portal e eventos que englobam o assunto ontologia.

Atualmente o portal disponibiliza as ontologias descritas a seguir.

Ecologia: essa ontologia faz parte do projeto *BLOC-Eco* [3] do ICMC USP que tem como foco principal o desenvolvimento de bases de dados ontológicos da ecologia para o português do Brasil.

Arte: essa ontologia foi criada no projeto *Ontology Aided Information Retrieval* (OnAIR) [4] do IME USP que envolve um sistema de recuperação de vídeos para uma coleção de vídeos Arte Contemporânea.

Música: a ontologia (Ontomusica engloba a história da música, incluindo gêneros musicais, obras e autores.

Telecomunicação: a ontologia *Smartphone* foi criada para apoiar a realização de pesquisas em conteúdos de blogs.

Emoção: essa ontologia foi criada no projeto *PersonaImp: Ferramenta para construção da personalidade e de cursos de ação para agentes improvisacionais voltados a interfaces* [5].

Para a organização das ontologias no portal, propomos o uso do modelo proposto em [6] e de métricas descritivas. Esses são temas de duas dissertações de mestrado em desenvolvimento na PUCRS.

A padronização de descrição de ontologias é importante para a construção de repositórios visando o reuso. Nesse sentido, foi criado o modelo OMV [7], que especifica um modelo de meta dados para ontologias. O OMV pretende transformar-se em um esquema para a *Web Semântica* [6].

³ <http://www.inf.pucrs.br/~ontolp/>

A definição do OMV está disponível em seu *site* ⁴. Neste artigo apresentamos apenas uma breve descrição do modelo. A classe principal é a classe *Ontology*, suas propriedades são nome, sigla, língua em que é escrita, data de criação, entre outras. Essa classe relaciona-se com outra classe identificando seu criador, que pode ser uma pessoa (*Person*) ou uma organização (*Organization*). Essas duas classes (pessoa e organização) são agrupadas sobre a classe genérica *Party*. Os editores de ontologias utilizados na criação da ontologia são definidos pela classe *OntologyEngineeringTool*. As sintaxes, linguagens de paradigmas de representação são dadas pelas classes *OntologySyntax*, *OntologyLanguage* e *RepresentationParadigm*. O modelo OMV ainda possui a classe *OntologyEngineeringMethodology*, responsável por identificar que metodologias foram utilizadas na construção da ontologia. O modelo possui a classe *LicenseModel* para que os proprietários de uma ontologia indiquem quais licenças se aplicam a(s) ontologia(s) desenvolvida(s). O modelo prevê ainda uma série de métricas básicas associadas à ontologia.

A descrição das ontologias disponibilizadas no portal, através de um conjunto de métricas básicas, é apresentada na **Tabela 1**.

	Ecologia	Arte	Música	Telecomunicação	Emoção
Número de Classes	59	170	19	123	68
Número de Propriedades	16	26	21	97	15
Número de Indivíduos	60	151	282	78	0
Número de Restrições	0	0	0	35	17

Tabela 1. Métricas das ontologias disponibilizadas no portal OntoLP.

Para esse conjunto de ontologias podemos observar que a ontologia com maior número de conceitos é a da Arte. Em relação às instâncias, a ontologia mais representativa é a da Música. A que apresenta maior número de propriedades e de restrições é a da Telecomunicação. A visualização dos dados dessas métricas pode auxiliar um pesquisador na busca de um recurso, de acordo com as suas necessidades.

As pesquisas em desenvolvimento irão implementar e avaliar o uso do modelo e de métricas como recursos descritivos para o portal.

Um terceiro trabalho realizado pelo grupo refere-se à extração automática de ontologias de texto, sendo esse o tema de uma tese de doutorado. Etapas como extração de termos simples e compostos já apresentaram resultados, descritos no portal.

⁴ <http://ontoware.org/projects/omv/>

3 Considerações finais e trabalhos futuros

Neste artigo apresentamos o portal OntoLP em desenvolvimento na PUCRS com o objetivo de disponibilizar recursos para a pesquisa em ontologias no Brasil e que envolvam a língua Portuguesa.

O modelo OMV é proposto para a descrição de ontologias do portal. Por meio de um formulário, o usuário deverá cadastrar uma ontologia e informar seus dados de acordo com o modelo (seu objetivo, grupo de pesquisa, instituição, tipo de licença, e outros). Além dessas informações, o usuário poderá fazer o *upload* da ontologia. A partir do momento que a ontologia é recebida, ela poderá ser processada para a criação de uma nova instância do modelo, combinando as informações provenientes do formulário e outras informações extraídas automaticamente do arquivo enviado.

O próximo passo será implementar os serviços de busca para auxiliar os usuários no processo de acesso e análise das ontologias disponíveis.

Outras métricas propostas na literatura serão analisadas e incorporadas ao portal, proporcionando assim uma descrição quantitativa dos recursos para pesquisas neste campo no Brasil. Consideramos que as métricas permitem uma avaliação global rápida das ontologias e também auxiliam a monitorar as suas evoluções.

Planejamos oferecer uma ferramenta para visualização gráfica das ontologias armazenadas. Pretende-se ainda continuar o levantamento de ontologias disponíveis em língua portuguesa, e ampliar o portal em termos de recursos disponibilizados.

Referências

1. Berners-Lee, T., Hendler, J., Lassila, O.: The semantic web. *Scientific American* **284**(5) (2001) 34–43
2. Pinto, H.S., Martins, J.P.: A methodology for ontology integration. In: K-CAP '01: Proceedings of the 1st international conference on Knowledge capture, New York, NY, USA, ACM (2001) 131–138
3. Zavaglia, C., Kasama, D.Y.: O delineamento de uma ontologia com vistas ao tratamento computacional: uma proposta para o subdomnio de ecologia de comunidades. *Palavra (PUCRJ)* **12**(1) (2004) 56–71
4. Paz-Trillo, C., Wassermann, R., Braga, P.P.: An information retrieval application using ontologies. *JBCS* **11**(2) (2005) 17–31
5. Wilkens, R.S., Moraes, M.C.: Proposta de uma ontologia para representar e gerenciar emoções em agentes. WESAAC - Workshop e Escola de Sistemas de Agentes para Ambientes Colaborativos, UCPel (2007) <http://www.inf.pucrs.br/petinf/>.
6. Hartmann, J., Bontas, E.P., Palma, R., Gomez-Perez, A.: Demo - design environment for metadata ontologies. In SpringerLink, ed.: *The Semantic Web: Research and Applications*, Springer, Springer Berlin / Heidelberg (2006) 427–441
7. Palma, R., Hartmann, J., Gomez-Perez, A.: Towards an ontology metadata standard. 3rd European Semantic Web Conference (June 2006)