Pontifícia Universidade Católica
do Rio de Janeiro

Carolina Howard Felicíssimo

Interoperabilidade Semântica na Web:
Uma Estratégia para o Alinhamento Taxonômico de Ontologias

Dissertação de Mestrado

Dissertação apresentada ao Programa de Pós-Graduação em Informática do Departamento de Informática da PUC-Rio como parte dos requisitos parciais para obtenção do título de Mestre em Informática.

Orientadores: Julio Cesar Sampaio do Prado Leite
Karin Koogan Breitman

Rio de Janeiro

Agosto de 2004

Pontifícia Universidade Católica
do Rio de Janeiro

Carolina Howard Felicíssimo

Interoperabilidade Semântica na Web: Uma Estratégia para o Alinhamento Taxonômico de Ontologias

Dissertação apresentada como requisito parcial para a obtenção do grau de Mestre pelo Programa de Pós-graduação em Informática do Departamento de Informática do Centro Técnico e Científico da PUC-Rio. Aprovada pela Comissão Examinadora abaixo assinada.

Prof. Julio Cesar Sampaio do Prado Leite
Orientador
Departamento de Informática – PUC-Rio

Profª. Karin Koogan Breitman
Co-orientadora
Departamento de Informática – PUC-Rio

Prof. Carlos José Pereira de Lucena
Departamento de Informática – PUC-Rio

Profª Simone Diniz Junqueira Barbosa
Departamento de Informática – PUC-Rio

Prof. Ricardo Choren Noya
Depto de Engenharia de Computação – IME

Prof. José Eugênio Leal
Coordenador Setorial do Centro
Técnico Científico – PUC-Rio

Rio de janeiro, 19 de agosto de 2004

Carolina Howard Felicíssimo

Graduou-se em Engenharia de Computação na PUC-Rio em 2001. É pesquisadora associada ao Laboratório de Engenharia de Software (LES) da PUC-Rio, atuando na área de Web Semântica da Engenharia de Software.

Ficha Catalográfica

Felicíssimo, Carolina Howard

Interoperabilidade semântica na Web : uma estratégia para o alinhamento taxonômico de ontologias / Carolina Howard Felicíssimo ; orientadores: Julio Cesar Sampaio do Prado Leite, Karin Koogan Breitman. – Rio de Janeiro : PUC, Departamento de Informática, 2004.

180 f. : il. ; 30 cm

Dissertação (mestrado) – Pontifícia Universidade Católica do Rio de Janeiro, Departamento de Informática.

Inclui referências bibliográficas.

1. Informática – Teses. 2. Web semântica. 3. Interoperabilidade semântica. 4. Ontologias. 5. Interoperabilidade de ontologias. 6. Alinhamento. 7. Agentes de software. 8. Engenharia de software. I. Leite, Julio Cesar Sampaio do Prado. II. Breitman, Karin Koogan. III. Pontifícia Universidade Católica do Rio de Janeiro. Departamento de Informática. IV. Título.

CDD: 004

A todos aqueles que, de uma forma ou de outra,
ajudaram a fazer este trabalho.

Agradecimentos

A Deus, por todos os caminhos iluminados decisivos na minha vida.

À minha família pelo apoio, carinho, suporte e encorajamento. Em especial, aos meus pais e irmã, à Izalea, ao meu afilhado Jordan e a sua irmã Giulia.

Ao meu orientador, Professor Julio Cesar Sampaio do Prado Leite que, mesmo quando eu ainda estava na graduação, me ofereceu a oportunidade de fazer mestrado sob sua orientação. Sua preocupação com os detalhes ajudou muito para a qualidade deste trabalho.

À minha co-orientadora, Professora Karin Koogan Breitman que me apresentou ao tema de pesquisa da dissertação e teve certeza, desde o início, deste tratar-se de um bom tema. Sua paciência, apoio, atenção e didática contribuíram para que, com o tempo, eu conseguisse caminhar sozinha no desenvolvimento do trabalho com tranqüilidade e confiança.

À minha amiga Karin, por me propiciar tantos momentos de descontração com sua família e amigos. Pelo otimismo e confiança na minha capacidade.

Aos meus amigos, em especial, Gustavo Robichez, Miriam Sayão, Lyrene Fernandes e Roberto Martins por todos os ensinamentos profissionais e pessoais, e por estarem sempre presentes em tantos momentos difíceis e bons da minha vida.

Aos Professores Carlos José Pereira de Lucena, Simone Barbosa e Ricardo Choren por participarem da Comissão Examinadora. Ao professor Ulf Bergmann pela ajuda e por todos os recursos disponibilizados, essenciais para a execução do trabalho no período devido. A todos os professores e funcionários do Departamento pela ajuda. Ao Luís Fernando pelo seu suporte.

À CAPES, à FAPERJ e ao Departamento de Informática da PUC-Rio, pelos auxílios concedidos, sem os quais este trabalho não poderia ter sido realizado.

Resumo

Felicíssimo, Carolina Howard; Leite, Julio Cesar Sampaio do Prado; Breitman, Karin Koogan. Interoperabilidade Semântica na Web: Uma Estratégia para o Alinhamento Taxonômico de Ontologias. Rio de Janeiro, 2004. 180p. Dissertação de Mestrado - Departamento de Informática, Pontifícia Universidade Católica do Rio de Janeiro.

Com a evolução da Web atual para a Web Semântica, acredita-se que as informações disponíveis estarão estruturadas de forma a permitir o processamento automático de seu conteúdo por máquinas. Além do processamento individual, deseja-se uma melhor troca de informações entre aplicações Web. Para estes propósitos, são necessários mecanismos que garantam a interoperabilidade semântica, i.e., identificação e compatibilidade de informações. Neste sentido, ontologias são utilizadas como um recurso para disponibilizar um vocabulário estruturado e livre de ambigüidades. Ontologias fornecem um padrão bem definido para a estruturação da informação e promovem um formalismo passível de processamento automático. Neste trabalho, propomos uma estratégia para interoperabilidade de ontologias. O Componente para Alinhamento Taxonômico de Ontologias – CATO, resultado da implementação desta estratégia proposta, alinha automaticamente as taxonomias de ontologias comparadas. O alinhamento realizado é obtido em três etapas executadas seqüencialmente. A primeira etapa compara lexicalmente os conceitos das ontologias entradas e usa um mecanismo de poda estrutural dos conceitos associados como condição de parada. A segunda etapa compara estruturalmente as hierarquias das ontologias identificando as similaridades entre suas sub-árvores comuns. A terceira etapa refina os resultados da etapa anterior classificando os conceitos identificados como *similares* em *bem similares* ou *pouco similares*, de acordo com um percentual de similaridade pré-definido.

Palavras-chave

Web Semântica, Interoperabilidade Semântica, Ontologias, Interoperabilidade de Ontologias, Alinhamento, Agentes de Software, Engenharia de Software.

Abstract

Felicíssimo, Carolina Howard; Leite, Julio Cesar Sampaio do Prado; Breitman, Karin Koogan. Semantic Web Interoperability: One strategy for the Taxonomic Ontology Alignment. Rio de Janeiro, 2004. 180p. Master Dissertation - Departamento de Informática, Pontifícia Universidade Católica do Rio de Janeiro.

With the Web evolving towards a Semantic Web, it is believed that the available information will be presented in a meaningful way to allow machines to automatically process its content. Besides the individual processing, a better information exchange among Web applications is desired. For this purpose, mechanisms are called for guarantee the semantic interoperability, that is, the identification and compatibility of information. In this direction, ontologies are used as one resource to make available a structured vocabulary, free of ambiguities. Ontologies provide a well-defined standard to structure the information and to promote formalism for automatic processing. In this work, we propose one strategy for ontology interoperability. The Ontology Taxonomic Alignment Component – CATO, which is the result of the implementation of this proposed strategy, provides an automatic taxonomic ontologies alignment. In this way, the alignment is obtained by a three-step process. The first step is the lexical comparison between the concepts from the entries ontologies. It uses a trimming mechanism of the related associated concepts as a stop condition. The second step is the structural comparison of the ontologies structures used to identify the similarities between common sub-trees. The third step refines the results of the previous step, classifying the *similar* identified concepts as *very similar* or *little similar*, according to a pre-defined similarity measurement.

Keywords

Semantic Web, Semantic Interoperability, Ontologies, Ontologies Interoperability, Alignment, Software Agents, Software Engineering.

Sumário

Lista de figuras

Lista de Abreviaturas

API – *Application Programming Interface*

ATLAS – *Agent Transaction Language for Advertising Services*

CATO – Componente para Alinhamento Taxonômico de Ontologias

CMU – *Carnegie Mellon University*

DAML – *Darpa Agent Markup Language*

OIL – *Ontology Inference Layer*

DAML+OIL – *Darpa Agent Markup Language* + *Ontology Inference Layer*

DOM – *Document Object Model*

HTML – *HyperText Markup Language*

IEEE – *Institute of Electrical and Electronics Engineers*

ITM – *Intelligent Topic Manager*

LAL – Léxico Ampliado da Linguagem

OWL – *OWL Web Ontology Language*

PUC-RIO – Pontifícia Universidade Católica do Rio de Janeiro

RDF – *Resource Description Framework*

RDF Schema – *RDF Vocabulary Description Language*

SUMO – *Suggested Upper Merged Ontology*

UdI – Universo de Informação

URI – *Uniform Resource Identifier*

URL – *Uniform Resource Locator*

XML – *Extensible Markup Language*

W3C – *World Wide Web Consortium*

Web – *World Wide Web*