



Raphael do Vale Amaral Gomes

**Matchmaking – Uma infraestrutura para alinhamento de
esquemas**

Dissertação de mestrado

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.

Orientador: Prof. Marco Antonio Casanova

Rio de Janeiro

Abril de 2010



Raphael do Vale Amaral Gomes

**Matchmaking – Uma infraestrutura para alinhamento de
esquemas**

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. Marco Antonio Casanova

Orientador

Departamento de Informática - PUC-Rio

Prof. Prof. Antonio L. Furtado

Departamento de Informática - PUC-Rio

Prof. Luiz André P. Paes Leme

UFF

Prof. José Eugenio Leal

Coordenador Setorial do Centro

Técnico Científico – PUC-Rio

Rio de Janeiro, 05 de Abril de 2010.

Todos os direitos reservados. É proibida a reprodução total ou parcial do trabalho sem autorização da universidade, do autor e do orientador.

Raphael do Vale Amaral Gomes

Graduou-se em Bacharelado em Informática pela Pontifícia Universidade Católica do Rio de Janeiro (PUC-Rio) em Agosto de 2006. Tem experiência na área de Ciência da Computação, com ênfase em Banco de Dados e Engenharia de Software. Tem trabalhado em análise de sistemas desde 2006.

Ficha Catalográfica

Gomes, Raphael do Vale Amaral

Matchmaking – uma infraestrutura para alinhamento de esquemas / Raphael do Vale Amaral Gomes ; orientador: Marco Antonio Casanova. – 2010.

78 f. : il. ; 30 cm

Dissertação (Mestrado)–Pontifícia Universidade Católica do Rio de Janeiro, Departamento de Informática, Rio de Janeiro, 2010.

Inclui bibliografia

1. Informática – Teses. 2. Alinhamento de esquemas. 3. Funções de similaridade. 4. OWL. I. Casanova, Marco Antonio. II. Pontifícia Universidade Católica do Rio de Janeiro. Departamento de Informática. III. Título.

CDD: 004

Para Mariana. Por fazer parte disso tudo.

Agradecimentos

Primeiramente, gostaria de agradecer a minha namorada Mariana. Sem ela, tenho certeza que nada disso teria acontecido. Sua força, energia e carinho me ajudaram demais a atingir esse objetivo.

Agradeço aos meus pais, Fátima e Walter. Sem toda a criação que me deram e o esforço dedicado em toda a minha vida, eu jamais teria chegado aqui. Agradeço ao meu irmão, Gabriel, por estar sempre ao meu lado.

Agradeço ao meu orientador, Professor Marco Antonio Casanova, pelo seu conhecimento técnico, carinho, paciência e por conseguir sempre me ajudar quando foi preciso.

Agradeço ao professor Luiz André P. Paes Leme pela ajuda e solicitude. Sua contribuição foi fundamental para a conclusão deste trabalho.

À CAPES, pelo financiamento desta pesquisa.

Resumo

Gomes, Raphael do Vale Amaral; Casanova, Marco Antonio. **Matchmaking - Uma infraestrutura para alinhamento de esquemas**. Rio de Janeiro, 2010. 78p. Dissertação de Mestrado - Departamento de Informática, Pontifícia Universidade Católica do Rio de Janeiro.

Um esquema conceitual de banco de dados, ou simplesmente um esquema, é a descrição em alto nível de como os conceitos de um banco de dados estão organizados. O alinhamento de um esquema origem S em um esquema destino T define conceitos em T nos termos dos conceitos de S. Esse trabalho descreve uma ferramenta de software que auxilia a implementação de técnicas de alinhamento de esquemas OWL baseadas em instâncias, que dependem da definição de funções de similaridade para avaliar a proximidade semântica dos elementos de dois esquemas diferentes. A ferramenta foi projetada para permitir a utilização de diferentes funções de similaridade e a troca dos algoritmos de alinhamento, facilitando assim a experimentação com diferentes configurações de alinhamento.

Palavras-chave

Alinhamento de esquemas; funções de similaridade; OWL.

Abstract

Gomes, Raphael do Vale; Casanova, Marco Antonio (Advisor). **MatchMaking – An Infrastructure to Match Schemas**. Rio de Janeiro, 2010. 78p. MSc. Dissertation - Departamento de Informática, Pontifícia Universidade Católica do Rio de Janeiro.

A database conceptual schema, or simply a schema, is a high level description of how database concepts are organized. The schema matching from a source schema S into a target schema T defines concepts in T on terms of the concepts in S . This work describes a software tool that helps implement instance-based schema matching techniques for OWL dialects that depend on the definition of similarity functions to evaluate the semantic proximity of elements from two different schemas. The tool is designed to accommodate different similarity functions and distinct matching algorithms, thereby facilitating experimenting with alternative matching configurations.

Keywords

Schema matching; similarity functions; OWL

Sumário

1	Introdução	12
1.1.	Alinhamento de esquemas	12
1.2.	Trabalhos relacionados	13
1.3.	Organização do trabalho	17
2	Fundamentos	18
2.1.	RDF/OWL	18
2.2.	A técnica para alinhamento de esquemas	21
3	MatchMaking – Uma infraestrutura para alinhamento de esquemas	23
3.1.	Detalhes da arquitetura do Matchmaking	23
3.2.	Implementação do Matchmaking	28
4	Testes do protótipo	49
4.1.	MatchmakingGUI – a interface gráfica do Matchmaking	49
4.2.	Four Step Matcher – o algoritmo de alinhamento	58
4.3.	MatchmakingBatch – alinhamento de esquemas em lote	65
5	Conclusões e trabalhos futuros	67
6	Bibliografia	70
	Apêndice A. Lista de comandos	74
	Apêndice B. Configuração do Matchmaking	77

Lista de imagens

Figura 1 - Conjunto de propriedades de Classe1.propriedade1, Classe2.propriedade2 e Livro.autor.	14
Figura 2 - Modelo conceitual do banco de dados.	25
Figura 3 - Divisão da arquitetura.	27
Figura 4 - Matchmaking sendo acessado por várias aplicações.	28
Figura 5 - Implementação dos comandos	31
Figura 6 – seqüência genérica de operações de um comando	32
Figura 7 - Representação de dados.	34
Figura 8 - seqüência genérica da geração de uma representação de dados	35
Figura 9 - Funções de similaridade.	36
Figura 10 - Seqüência genérica do calculo de uma função de similaridade.	37
Figura 11 - Algoritmos de alinhamento.	40
Figura 12 - Seqüência genérica de um algoritmo de alinhamento.	42
Figura 13 – esquema	43
Figura 14 - DAO	46
Figura 15 - Modelo físico do banco de dados (1).	47
Figura 16 - Modelo físico do banco de dados (2).	48
Figura 17 - Casos de uso do matchmakingGUI.	50
Figura 18 - Tela: listar esquema.	50
Figura 19 - Tela: cadastrar esquema.	51
Figura 20 - Tela: novo <i>dataset</i> .	52
Figura 21 - Tela: listar algoritmos.	52
Figura 22 – Tela: cadastrar algoritmo de alinhamento.	53
Figura 23 - Tela: ver detalhes de um algoritmo.	54
Figura 24 - Tela: iniciar operação de alinhamento.	55
Figura 25 - Tela: exibir execuções.	56
Figura 26 - tela: exibir histórico (1)	57
Figura 27 – Tela: exibir histórico (2).	58

Figura 28 - Diagrama de atividades do <i>FourStepMatcher</i> .	59
Figura 29 - Seqüência simplificada de operações do <i>PropertyMatcher</i> .	63
Figura 30 - Exemplo de alinhamento em lote.	66
Figura 31 - Arquivo de configuração <i>matchmaking.properties</i> .	78

Lista de tabelas

Tabela 1 - Exemplo de um alinhamento de vocabulário	22
Tabela 2 - parâmetros de <i>tokens of properties</i>	60
Tabela 3 - parâmetros de contrast model	61
Tabela 4 - Parâmetros de <i>MatchablePropertiesOfClasses</i>	63
Tabela 5 - lista de comandos possíveis	74