

Jairo Francisco de Souza

**Uma abordagem heurística uni-objetivo para
calibragem em meta-alinhadores de
ontologias**

Tese de Doutorado

Tese apresentada ao Programa de Pós-graduação em Informática
do Departamento de Informática da PUC-Rio como requisito
parcial para obtenção do título de Doutor em Informática

Orientador : Prof. Carlos José Pereira de Lucena
Co-Orientador: Prof. Rubens Nascimento Melo

Rio de Janeiro
Junho de 2012



Jairo Francisco de Souza

**Uma abordagem heurística uni-objetivo para
calibragem em meta-alinhadores de
ontologias**

Tese apresentada ao Programa de Pós-graduação em Informática do Departamento de Informática do Centro Técnico Científico da PUC-Rio como requisito parcial para obtenção do título de Doutor em Informática. Aprovada pela comissão examinadora abaixo assinada.

Prof. Carlos José Pereira de Lucena

Orientador

Departamento de Informática — PUC-Rio

Prof. Rubens Nascimento Melo

Co-Orientador

Departamento de Informática — PUC-Rio

Prof. Karin Koogan Breitman

Departamento de Informática – PUC-Rio

Prof. Antonio Luz Furtado

Departamento de Informática – PUC-Rio

Prof. Sean Wolfgang Matsui Siqueira

Departamento de Informática Aplicada – UNIRIO

Prof. Jonice de Oliveira Sampaio

Departamento de Ciência da Computação – UFRJ

Prof. José Eugenio Leal

Coordenador do Centro Técnico Científico — PUC-Rio

Rio de Janeiro, 28 de Junho de 2012

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

Jairo Francisco de Souza

Graduou-se em Informática na Universidade Federal de Juiz de Fora. Concluiu o Mestrado em Engenharia de Sistemas e Computação na área de Banco de Dados na Universidade Federal do Rio de Janeiro. Concursado em 2009 como professor do Departamento de Ciência da Computação na Universidade Federal de Juiz de Fora, onde é coordenador do MBA em Gestão de Projetos de TI e pesquisador do Núcleo de Pesquisa em Engenharia do Conhecimento (NEnC). Ministrou aulas em cursos de extensão da PUC-Rio, participou de projetos de pesquisa na PUC-Rio e no Núcleo de Pesquisa e Prática em Tecnologia (NP2Tec) da UNIRIO, publicou artigos e capítulos de livros.

Ficha Catalográfica

Souza, Jairo Francisco de

Uma abordagem heurística uni-objetivo para calibragem em meta-alinhadores de ontologias / Jairo Francisco de Souza; orientador: Carlos José Pereira de Lucena; co-orientador: Rubens Nascimento Melo. — Rio de Janeiro : PUC-Rio, Departamento de Informática, 2012.

v., 105 f: il. ; 29,7 cm

1. Tese (Doutorado em Informática) - Pontifícia Universidade Católica do Rio de Janeiro, Departamento de Informática.

Inclui referências bibliográficas.

1. Informática – Tese. 2. Alinhamento de ontologias. 3. Meta-alinhamento. 4. Calibragem de pesos. 5. Algoritmo genético. I. Lucena, Carlos José Pereira de. II. Melo, Rubens Nascimento. III. Pontifícia Universidade Católica do Rio de Janeiro. Departamento de Informática. IV. Título.

CDD: 004

Agradecimentos

Ao meu orientador, professor Carlos Lucena, por aceitar esta orientação.

Aos meus co-orientadores, professores Sean Siqueira e Rubens de Melo, por estimularem o meu trabalho e por serem exigentes e críticos para que eu obtivesse o melhor resultado.

Ao CNPq e à PUC-Rio, pelos auxílios concedidos, sem os quais este trabalho não poderia ter sido realizado.

Ao professor Stênio Sã, por disponibilizar várias horas do seu tempo para ensinar e discutir esta pesquisa.

Aos professores Antonio Furtado, Jonice Oliveira, Karin Breitman e Stênio Sã, por aceitarem participar da banca de defesa deste trabalho.

Aos meus colegas de trabalho e familiares, que incentivaram e torceram pela conclusão deste trabalho.

Resumo

Souza, Jairo Francisco de; Lucena, Carlos José Pereira de; Melo, Rubens Nascimento. **Uma abordagem heurística uni-objetivo para calibragem em meta-alinhadores de ontologias**. Rio de Janeiro, 2012. 105p. Tese de Doutorado — Departamento de Informática, Pontifícia Universidade Católica do Rio de Janeiro.

Ontologias são vistas como uma solução para o problema de interoperabilidade em muitas aplicações como integração de banco de dados, comércio eletrônico, serviços web e redes sociais. Contudo, ao utilizar diferentes ontologias, persiste o problema de incompatibilidade semântica entre os modelos. Dentre os desafios para o campo de alinhamento de ontologia, está a seleção de alinhadores e a auto-configuração desses alinhadores. Este trabalho tem por objetivo tratar o problema de meta-alinhamento de ontologias utilizando métodos heurísticos para calibragem de alinhadores escolhidos sem detrimento do desempenho do sistema. Para avaliar esta proposta em comparação a outras propostas da literatura, optou-se por utilizar um benchmark para alinhamentos de ontologias amplamente utilizado por pesquisadores desse campo. A comparação entre as propostas é realizada com base em medidas de qualidade de alinhamentos sugeridas pelo benchmark. Além da avaliação quantitativa proposta pelo benchmark, é realizada uma comparação qualitativa em relação às características desejáveis de meta-alinhadores propostas na literatura.

Palavras-chave

Alinhamento de ontologias; Meta-alinhamento; Calibragem de pesos; Algoritmo genético;

Abstract

Souza, Jairo Francisco de; Lucena, Carlos José Pereira de (Advisor); Melo, Rubens Nascimento (Co-advisor). **A single goal heuristic approach to solve tuning in ontology meta-matching**. Rio de Janeiro, 2012. 105p. DSc Thesis — Departamento de Informática, Pontificia Universidade Católica do Rio de Janeiro.

Ontologies are seen as a solution for many applications such as database integration, e-commerce, web services and social networks. However, the use of distinct ontologies does not solve the semantic interoperability problem among models. Matcher selection and self-configuring are challenges from the ontology matching field. This work aims to treat the ontology meta-matching problem using heuristic methods to fast tune a set of matchers. To evaluate this proposal, a well-known benchmark for ontology matching is used. The comparison among meta-matching approaches includes the benchmark evaluation metrics and qualitative metrics for meta-matching proposed in the literature.

Keywords

Ontology matching; Meta-matching; Self-tuning; Genetic Algorithm;

Sumário

1	Introdução	11
1.1	Motivação	13
1.2	Objetivos	16
1.3	Enfoque de solução e método de avaliação da pesquisa	16
1.4	Organização da tese	18
2	Meta-alinhamento de ontologias	19
2.1	Definições	19
2.2	Algoritmos genéticos	24
2.3	Meta-alinhadores	27
2.4	Abordagens heurísticas para meta-alinhamento	29
3	Abordagem proposta para calibragem de alinhadores	35
3.1	Descrição do problema	36
3.2	Representação da solução	36
3.3	Construtor das soluções iniciais	39
3.4	Operadores de reprodução e mutação	41
3.5	Intensificação da solução	44
3.6	Construção da nova geração	45
3.7	Demonstração do comportamento da abordagem	46
4	Implementação do meta-alinhador	53
4.1	Arquitetura da ferramenta	53
4.2	Módulo de composição	54
4.3	Módulo analisador	59
4.4	Módulo de persistência	63
5	Avaliação do meta-alinhador	64
5.1	Descrição dos testes	64
5.2	Configuração utilizada para realização dos testes	67
5.3	Resultados obtidos	69
6	Análise dos resultados	75
6.1	Considerações quanto aos meta-alinhadores	75
6.2	Considerações quanto à base de teste	78
6.3	Considerações finais	79
7	Conclusões	83
7.1	Contribuições	83
7.2	Limitações	84
7.3	Trabalhos futuros	85
	Referências Bibliográficas	88

A	Algoritmos para medidas de similaridade de texto	99
A.1	Igualdade	99
A.2	Hamming	100
A.3	Levenshtein	100
A.4	Jaro-Winkler	101
A.5	Damerau-Levenshtein	101
A.6	TFIDF	102
B	Arquivos XML utilizados para configuração do sistema	103

Lista de figuras

1.1	Quantidade de artigos publicados por ano	14
2.1	Método de ponto de cruzamento único, cruzamento de dois pontos e cruzamento uniforme	26
3.1	Representação da equação $(F_1P_4 + F_2P_5)P_1 + F_4P_2 + F_5P_3 = K_1$	37
3.2	Cromossomo com destaque para a posição de cada nível	40
3.3	Exemplo de cromossomo com valores válidos para funções do primeiro nível	40
3.4	Exemplo de um cromossomo com valores válidos	40
3.5	Técnica da roleta para seleção de indivíduos (Gudwin & von Zuben, 1998)	42
3.6	Exemplo de <i>crossover</i>	43
3.7	Como uma nova população é gerada	45
3.8	Demonstração da convergência do AG para a equação 3-10	49
3.9	Gráfico ampliado da convergência do AG para a equação 3-10	49
3.10	Demonstração da convergência do AG para a equação 3-11	50
3.11	Gráfico ampliado da convergência do AG para a equação 3-11	51
3.12	Demonstração da convergência do AG para a equação 3-11	52
4.1	Arquitetura da ferramenta	54
4.2	Exemplo de documento XML para descrição de funções compostas	55
4.3	Exemplo de documento XML com alinhamentos de entrada	56
5.1	Gráfico triangular com cobertura e precisão totais	71
6.1	Comparação entre os meta-alinhadores heurísticos	81

Lista de tabelas

2.1	Processo de evolução versus resolução em um problema de otimização, adaptado de (Talbi, 2009)	25
4.1	Algoritmos de distância de edição	57
4.2	Funções de similaridade disponíveis no GNoSIS+	58
4.3	Conjunto de tuplas	61
5.1	Tabela com o resumo dos testes realizados	66
5.2	Resultados encontrados no <i>benchmark</i>	70
5.3	Média harmônica dos resultados encontrados no <i>benchmark</i>	71
5.4	Comparação dos resultados com o meta-alinhador GOAL	72
5.5	Comparação dos resultados com o meta-alinhador MaSiMe	73
5.6	Comparação entre a média harmônica dos meta-alinhadores GNoSIS+, GAOM, GOAL e MaSiMe	74