

PONTIFÍCIA UNIVERSIDADE CATÓLICA
DO RIO DE JANEIRO



Sandra Aparecida Dias

**Integração Semântica de Dados Através de Federação de
Ontologias**

Dissertação de Mestrado

Dissertação apresentada como requisito parcial para 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. Rubens Nascimento Melo

Rio de Janeiro
17 de março de 2006



Sandra Aparecida Dias

Integração Semântica de Dados Através de Federação de Ontologias

Dissertação apresentada como requisito parcial para obtenção do título de Mestre pelo Programa de Pós-Graduação em Informática da PUC-Rio. Aprovada pela Comissão Examinadora abaixo assinada.

Prof. Rubens Nascimento Melo

Orientador

Departamento de Informática – PUC-Rio

Prof. Antonio Luz Furtado

Departamento de Informática – PUC-Rio

Prof. Álvaro César Pereira Barbosa

Departamento de Informática – PUC-Rio

Prof. José Eugenio Leal

Coordenador Setorial do Centro

Técnico Científico – PUC-Rio

Rio de Janeiro, 17 de março de 2006

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

Sandra Aparecida Dias

Graduou-se em Matemática pelo IBILCE - Instituto de Biociências Letras e Ciências Exatas da UNESP - Universidade Estadual Paulista Júlio de Mesquita Filho. Obteve o Aperfeiçoamento em Matemática pelo IMPA - Instituto de Matemática Pura e Aplicada. Pós-graduou-se em Análise de Sistemas pelo CCE - Coordenação Central de Extensão da Puc-Rio. Foi Bolsista DTI do CNPq em projeto no CNPS - Centro Nacional de Pesquisa em Solos da EMBRAPA - Empresa Brasileira de Pesquisas Agropecuárias. Foi bolsista da CAPES durante o Mestrado. Realiza atividades de pesquisa no TecBD - Laboratório de Bancos de Dados da Puc-Rio.

Ficha Catalográfica

Dias, Sandra Aparecida

Integração Semântica de Dados Através de Federação de Ontologias / Sandra Aparecida Dias; orientador: Rubens Nascimento Melo. - Rio de Janeiro: PUC-Rio, Departamento de Informática, 2006.

83 f. ; 29,7 cm

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

Incluí referências bibliográficas.

1 - Informática - Teses. 2. Ontologia. 3. Banco de Dados Distribuídos Heterogêneos. 4. Federação de Dados. 5. Integração de Dados. I. Melo, Rubens Nascimento. II. Pontifícia Universidade Católica do Rio de Janeiro. Departamento de Informática. III. Título.

CDD: 004

Aos meus pais que não mediram esforços para que eu pudesse chegar até aqui.
Ao meu filho amado que me ensinou a viver!

Agradecimentos

Primeiro quero agradecer a Deus que me proporcionou estar aqui junto de tantas pessoas amadas que me estenderam as mãos em vários momentos muito importantes.

Sem dúvida alguma conseguir chegar ao final deste trabalho foi um deles. A jornada foi longa sem dúvida, mas o aprendizado ficou gravado para sempre na alma. Foram momentos especiais, pois no meio das dificuldades e conseqüentes conquistas fui crescendo e vivenciei lições preciosas que me fizeram ter certeza que estar aqui aprendendo é o melhor da vida. A vida é realmente bela e vive-la é uma dádiva de Deus.

Melhor ainda quando Ele nos proporciona encontrar companheiros de jornada como os que eu pude encontrar aqui. Sem querer ser injusta vou citar apenas os que estiveram comigo nesta última fase da dissertação, mas sem dúvida muitos me ajudaram para que eu pudesse construir o que construí na minha jornada.

Poder cursar o mestrado no departamento de Informática da PUC foi um presente. Conheci no departamento pessoas especiais que marcaram profundamente a minha vida, amigos de cursos, professores, funcionários.

Foi no curso de PAA, que conheci Celso Carneiro Ribeiro, uma pessoa que eu admiro muito e que, com sua dedicação ao bom trabalho que desempenha como professor e pesquisador, fez com que eu aprendesse lições que marcarão para sempre a minha vida profissional. Celso é dedicado, atencioso com os alunos, preza pela qualidade da nossa formação como pesquisadores e não foca somente no aprendizado da disciplina. Com ele aprendemos a pesquisar, raciocinar de forma diferente, crescer! Ainda que isto tenha tomado horas de bom sono, naqueles dias de noites inteiras no Lab-Pós plotando trabalhos infinitos!

O amor pelos estudos cresceu no curso de Lógica e Especificação. O professor Hermann Hausler também teve papel especial na minha formação e com certeza na dos meus colegas de curso. Eu particularmente pude recuperar o gosto pela

matemática, minha primeira formação, e aprender a pensar e a filosofar. É um profissional dedicado, amigo dos alunos, sempre presente para esclarecer quaisquer dúvidas, de uma boa vontade incomparável! Que os conhecimentos de lógica adquiridos possam nos acompanhar para sempre nas experiências da vida!

O amor pelas Ontologias veio com o curso do professor Marco Antonio Casanova, organizadíssimo, eficiente e prestativo. Aprendi no seu curso a agregar conhecimentos de lógica aos conhecimentos de bancos de dados. Apaixonei-me tanto pelo assunto que virei referência no Lab-Pós, depois evidentemente de virar piada, eu só falava nas Ontologias! Devo a ele a motivação deste trabalho.

Ao meu orientador a quem devo toda a minha formação em bancos de dados, e que esteve presente durante anos na minha vida (já são dez anos de amizade!). Agradeço especialmente por ter me aceitado como aluna e por ter sempre acreditado em mim, além de ter me aberto as portas sempre que precisei, desde o início da minha carreira. Obrigada pela amizade e orientação no curso e na vida, Guru!

Aos amigos de curso, que também estiveram ao meu lado meus agradecimentos e meu carinho. Sem querer ser injusta mencionando alguns nomes, agradeço especialmente à amiga Aletéia pelos momentos especiais e ao amigo Portela pela força e ajuda na leitura final do trabalho.

À grande amiga, Vera Syme Benzecry que sempre me estendeu as mãos nos momentos difíceis, que não foram poucos durante 20 anos de amizade, e muito especialmente na fase final deste trabalho. A você, Vera, minha eterna gratidão.

À família representada pelos funcionários do departamento agradeço o carinho e dedicação. Nunca em um ambiente de trabalho eu pude ver nada igual! Pessoas tão dedicadas, atenciosas e amigas de verdade. Não queria citar nomes com receio de ser injusta, mas não posso deixar de falar do carinho da Carmen com o cafezinho e a conversa animada; da atenção dada pelo Nelson e Vagner - o botafoguense, do carinho do pessoal da secretaria Alex, Ruth, Manu, Fátima e especialmente claro agradeço à minha maninha santista que me deu, além de todas

as atenções do mundo, sempre uma palavra de apoio e força - ela é realmente uma pessoa muito especial, obrigada Deborah, muito obrigada!

É claro que numa jornada tão especial, tinha que ter alguém muito especial para ser testemunha e porto seguro nos momentos difíceis. Acredito que Deus, pensando em me amparar, colocou Bruno Santana como meu colega de curso e anjo da guarda. Bruno foi a pessoa mais especial nesse processo e foi um amigo que conquistei para a vida toda. Companheiro de trabalhos, de descobertas, de noites sem dormir, de desespero e de alegrias; como crescemos tentando chegar até aqui. Sem a amizade dele eu não teria conseguido. Amigo, obrigada é uma palavra simples que talvez não possa conter tudo o que devo a você, mas saiba que você mora no meu coração! Desejo que sejas muito feliz!

Aos meus familiares agradeço o apoio incondicional. Ao meu irmão agradeço o apoio logístico fundamental, sempre presente na minha vida! Aos meus pais como disse, devo tudo! Sem eles eu não só não teria chegado aonde cheguei, mas, especialmente não teria concluído o meu trabalho, eles foram fundamentais e imprescindíveis, e me ensinaram com exemplos, a ser dedicada, abnegada, me amparando e amparando ao meu filho para que eu pudesse estudar com tranquilidade. Peço perdão pelos momentos em que exigi tanto deles não podendo inclusive estar presente, mas agradeço do fundo da alma a dedicação.

Ao meu filho amado agradeço por fazer parte da minha vida, pois foi somente pela força que você me passa que eu consegui chegar até aqui, minha vida mudou muito depois que você começou a fazer parte dela e eu te amo muito!

À CAPES agradeço o apoio financeiro no decorrer do curso pelo benefício da bolsa que me foi concedida.

Resumo

Dias, Sandra Aparecida; Melo, Rubens Nascimento. **Integração Semântica de Dados através de Federação de Ontologias**. Rio de Janeiro, 2006. 83p. Dissertação de Mestrado - Departamento de Informática, Pontifícia Universidade Católica do Rio de Janeiro.

O advento da WEB propiciou a disseminação de bases de dados distribuídas e heterogêneas. Por vezes, a resposta a uma consulta demanda o uso de várias destas bases. É necessário, então, algum nível de integração destas. A publicação dessas bases nem sempre segue um padrão semântico. Em função disso parece ser essencial existir um meio de relacionar os diferentes dados para satisfazer tais consultas. Este processo é comumente denominado de integração de dados. A comunidade de Banco de Dados tem conhecimento de métodos para dar conta desta integração no contexto de federações de Bancos de Dados heterogêneos. No entanto, atualmente existem descrições mais ricas e com mais possibilidades de semântica, tais como aquelas induzidas pelo conceito de ontologia. A comunidade de Banco de Dados tem considerado ontologias na solução do problema da integração de Banco de Dados. O alinhamento ou “merge” de ontologias são algumas das propostas conhecidas da comunidade de WEB semântica. Este trabalho propõe o uso de métodos de “merge” de ontologias como solução ao problema da construção de uma federação de ontologias como método integrador de fontes de dados. O trabalho inclui a implementação de um estudo de caso na ferramenta Protegé. Este estudo de caso permite discutir aspectos de escalabilidade e de aplicabilidade da proposta como uma solução tecnologicamente viável.

Palavras-chave

Ontologia; Banco de Dados Distribuídos Heterogêneos; Federação de Dados; Integração de Dados

Abstract

Dias, Sandra Aparecida; Melo, Rubens Nascimento. **Semantic Data Integration with an Ontology Federation**. Rio de Janeiro, 2002. 83p. MSc. Dissertation - Departamento de Informática, Pontifícia Universidade Católica do Rio de Janeiro.

The WEB has spread out the use of heterogeneous distributed databases. Sometimes, the answer to a query demands the use of more than one database. Some level of integration among these databases is desired. However, frequently, the bases were not designed according a unique semantic pattern. Thus, it seems essential to relate the different data, in the respective base, in order to provide an adequate answer to the query. The process of building this relationship is often called data integration. The Data Base community has acquired enough knowledge to deal with this in the context of Data Base Heterogeneous Federation. Nowadays, there are more expressive model descriptions, namely ontologies. The Data Base community has also considered ontologies as a tool to contribute as part of a solution to the data integration problem. The Semantic WEB community defined alignment or merge of ontologies as one of the possible solutions to the some of this integration problem. This work has the aim of using “merge” of ontologies methods as a mean to define the construction of a Federation of ontologies as a mean to integrate source of data. The dissertation includes a case study written in the Protegé tool. From this case study, a discussion follows on the scalability and applicability of the proposal as a feasible technological solution for data integration.

Keywords

Ontology, Heterogeneous Data Bases, Data Federation, and Data Integration

Sumário

1	INTRODUÇÃO	13
1.1	MOTIVAÇÃO	13
1.2	ORGANIZAÇÃO DA DISSERTAÇÃO	17
2	INTEGRAÇÃO DE DADOS	19
2.1	INTEROPERABILIDADE SEMÂNTICA	19
2.2	HETEROGENEIDADES ENTRE BANCOS DE DADOS	20
2.2.1	<i>Heterogeneidade quanto a SGBDs</i>	20
2.2.2	<i>Heterogeneidade Estrutural</i>	20
2.2.3	<i>Heterogeneidade Sintática</i>	21
2.2.4	<i>Heterogeneidade Semântica</i>	21
2.3	ONTOLOGIAS	23
2.4	INTEGRAÇÃO DE ONTOLOGIAS	25
2.5	DIFERENÇAS ENTRE ONTOLOGIAS	25
2.5.1	<i>Diferenças no Nível da Linguagem</i>	26
2.5.2	<i>Diferenças no nível da Ontologia</i>	27
2.6	MAPEAMENTO, ALINHAMENTO E TRADUÇÃO	29
2.6.1	<i>Mapeamento/Alinhamento</i>	29
2.6.2	<i>Merging/Integração</i>	29
2.7	BANCOS DE DADOS FEDERADO	29
2.7.1	<i>Arquitetura de cinco níveis</i>	30
2.8	ABORDAGEM DO TRABALHO	32
3	MERGING DE ONTOLOGIAS PARA UMA FEDERAÇÃO DE ONTOLOGIAS	34
3.1	FEDERAÇÃO DE ONTOLOGIAS	34
3.2	APROXIMAÇÕES PARA UMA FEDERAÇÃO DE ONTOLOGIAS	37
3.3	MERGING DE ONTOLOGIAS PARA DEFINIR UM SISTEMA FEDERADO AUTÔNOMO	38
3.4	UMA ARQUITETURA PARA FEDERAÇÃO DE ONTOLOGIAS	39
4	ESTUDO DE CASO	41
4.1	CONTEXTUALIZAÇÃO DO PROBLEMA	41
4.2	DA FEDERAÇÃO DE BANCOS DE DADOS PARA A FEDERAÇÃO DE ONTOLOGIAS	44
4.3	DESCRIÇÃO DO ESTUDO DE CASO	46
4.3.1	<i>O editor de ontologias Protégé</i>	46
4.3.2	<i>Definição das ontologias</i>	48
4.3.3	<i>A ontologia Resultante do Merging de Ontologias – O Modelo Global</i>	56
4.3.4	<i>O processo de integração das ontologias</i>	60
4.3.5	<i>A ontologia para a camada de aplicação</i>	65

5	CONCLUSÃO E TRABALHOS FUTUROS	72
5.1	RETROSPECTIVA	72
5.2	AVALIAÇÃO DA ESTRATÉGIA USADA.....	73
5.2.1	<i>Vantagens do uso de ontologias</i>	<i>75</i>
5.3	CONTRIBUIÇÕES.....	75
5.3.1	<i>Aplicação da solução.....</i>	<i>76</i>
5.4	TRABALHOS FUTUROS.....	77
6	REFERÊNCIAS BIBLIOGRÁFICAS.....	79

Lista de Figuras

Figura 1 : Arquitetura de 5 níveis para uma federação de bancos de dados.	32
Figura 2 : Arquitetura para uma federação de ontologias.	40
Figura 3: Modelo de dados Telefonia que representa a base de dados BD1.	42
Figura 4: Modelo de dados Campanha que representa a base de dados DB2.	43
Figura 5: Hierarquia de classes da ontologia Telefonia.	49
Figura 6: Classes e propriedades do domínio telefonia.	50
Figura 7: Definição da classe Produto.	51
Figura 8: Exemplo de propriedade declarada como <i>objectProperty</i>	52
Figura 9: Definição da propriedade Data_vencimento.	52
Figura 10: Resultado do teste de consistência da ontologia Telefonia usando o providor <i>Pellet</i>	53
Figura 11: Hierarquia das classes da ontologia Campanha.	54
Figura 12: Classes e propriedades do domínio Campanha.	55
Figura 13: Resultado de teste de consistência da ontologia Campanha usando o provedor <i>Pellet</i>	56
Figura 14: O conjunto de ferramentas Prompt (retirado de Noy & Musen, 2003).	58
Figura 15: Tela do PromptTab.	61
Figura 16: Situação de conflito identificada pelo iPrompt.	63
Figura 17: Hierarquia da ontologia resultante do <i>merging</i>	64
Figura 18: Resultado do teste de consistência da ontologia global.	65
Figura 19: Mapeamento do acesso dos usuários.	67
Figura 20: Hierarquia da ontologia de controle de acesso.	68
Figura 21: Tela mostrando o usuário criado para realizar inferência.	69
Figura 23: Tela mostrando o resultado da classificação realizada pelo raciocinador <i>Pellet</i>	70
Figura 24: Propriedade de inferência em DL.	71