

Tanara Lauschner

Modelagem de restrições de esquemas mediados

Tese de Doutorado

Tese apresentada como requisito parcial para obtenção do título de Doutor pelo Programa de Pós-Graduação em Informática da PUC-Rio.

Orientador: Prof. Marco Antonio Casanova

Rio de Janeiro
Agosto de 2010



Tanara Lauschner

Modelagem de restrições de esquemas mediados

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

Prof. Marco Antonio Casanova

Departamento de Informática - PUC-Rio (Orientador)

Prof. Antonio Luz Furtado

Departamento de Informática - PUC-Rio (Co-orientador)

Prof^a. Karin Koogan Breitman

Departamento de Informática - PUC-Rio

Prof. Altigran Soares da Silva

Departamento de Ciência da Computação - UFAM

Prof^a. Vânia Maria Ponte Vidal

Departamento de Computação - UFC

Prof. José Antônio F. de Macedo

Departamento de Computação – UFC

Luiz André P. Paes Leme

Instituto de Computação – UFF

Rio de Janeiro, 09 de agosto de 2010

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

Tanara Lauschner

Mestre em Ciência da Computação pela UFMG(Universidade Federal de Minas Gerais) em fevereiro de 2002. Graduiu-se em Engenharia Elétrica pela UFAM (Universidade Federal do Amazonas) em março de 1998.

Ficha Catalográfica

Lauschner, Tanara

Modelagem de restrições de esquemas mediados/
Tanara Lauschner; orientador: Marco Antonio Casanova. -
Rio de Janeiro: PUC-Rio, Departamento de Informática,
2010.

v., 80 f.: il. ; 29,7 cm

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

Inclui bibliografia

1. Informática – Teses 2. Restrições de esquemas
mediados. 3. Esquema mediado. 4. Lógica de Descrição.
5. Bancos de Dados. I. Casanova, Marco Antonio. II.
Pontifícia Universidade Católica do Rio de Janeiro.
Departamento de Informática. III.

Resumo

Lauschner, Tanara; Casanova, Marco Antonio. **Modelagem de restrições de esquemas mediados**. Rio de Janeiro, 2010. 80p. Tese de Doutorado - Departamento de Informática, Pontifícia Universidade Católica do Rio de Janeiro.

Integração de dados refere-se ao problema de combinar dados que estão armazenados em diferentes fontes, fornecendo ao usuário uma visão unificada dos dados. As consultas são então expressas em um esquema global ou esquema mediado, que deve incluir restrições de integridade que contribuam para um entendimento correto sobre o que a semântica das fontes de dados do ambiente de mediação tem em comum. Esta tese endereça o problema de modelar as restrições de um esquema mediado a partir das restrições dos esquemas importados e dos mapeamentos de esquemas. Argumenta-se que as restrições devem ser modeladas como o ínfimo das restrições dos esquemas exportados, depois de traduzidos para um vocabulário comum. Desta forma, assegura-se que os usuários do esquema mediado interpretarão os resultados das consultas corretamente. Para uma família expressiva de restrições, esta tese mostra como computar eficientemente o ínfimo de conjuntos de restrições.

Palavras-chave

Restrições de esquemas mediados, esquemas mediados, lógica de descrição.

Abstract

Lauschner, Tanara; Casanova, Marco Antonio. **Modeling the Mediated Schema Constraints**. Rio de Janeiro, 2010. 80p. DSc. Thesis - Departamento de Informática, Pontifícia Universidade Católica do Rio de Janeiro.

Data integration refers to the problem of combining data stored in different sources, providing users with a unified view of the data. Queries are then expressed in terms of a global or mediated schema, which should include integrity constraints that contribute to a correct understanding of what the semantics of the data sources have in common. This thesis addresses the problem of modeling the constraints of a mediated schema from the imported schemas constraints and mappings. It argues that the constraints should be modeled as the greatest lower bound of the constraints of the export schemas, after appropriate translation to a common vocabulary. This assures that users of the mediated schema will correctly interpret query results. For a rich family of constraints, it shows how to efficiently compute the greatest lower bound of sets of constraints.

Keywords

Integrity constraint, mediated schema, description logics.

Sumário

1	Introdução	11
1.1.	Integração de dados e ambiente de mediação	11
1.2.	O problema de revisar as restrições de um esquema mediado	13
1.3.	Contribuições	15
1.4.	Trabalhos relacionados	16
1.5.	Organização da tese	18
2	Esquemas conceituais em lógica de descrição	20
2.1.	Uma família de linguagens atributivas	20
2.2.	Esquemas elementares	23
2.3.	Esquemas ultra elementares	27
2.4.	Ambiente de mediação	28
2.5.	Conclusões do capítulo	34
3	Problema de decisão em esquemas conceituais	35
3.1.	Inclusão de consultas e problemas de decisão correlatos	36
3.2.	Extensão do tableau para esquemas ultra elementares	39
3.3.	Um procedimento de decisão rápido	43
3.4.	Conclusões do capítulo	48
4	Construção das restrições do esquema mediado	50
4.1.	Passos básicos do processo de revisão das restrições	50
4.2.	Cálculo do ínfimo de dois conjuntos de restrições	62
4.3.	Conclusões do capítulo	72
5	Conclusões e trabalhos futuros	74
6	Referências bibliográficas	76

Lista de figuras

Figura 1 – Componentes de um ambiente de mediação	12
Figura 2 – Definição informal do esquema da <i>Amazon</i>	26
Figura 3 – Definição informal do esquema do <i>eBay</i>	26
Figura 4 – Definição formal das restrições do esquema da <i>Amazon</i>	27
Figura 5 – Definição formal de algumas das restrições do esquema do <i>eBay</i> ...	27
Figura 6 – Vocabulário do esquema mediado <i>Sales</i>	30
Figura 7 – Vocabulário do esquema importado da <i>Amazon</i>	30
Figura 8 – Restrições do esquema importado da <i>Amazon</i>	31
Figura 9 – Mapeamento local do esquema da <i>Amazon</i>	31
Figura 10 – Vocabulário do esquema importado <i>eBay</i>	32
Figura 11 – Restrições do esquema importado <i>eBay</i>	32
Figura 12 – Mapeamento local do esquema <i>eBay</i>	32
Figura 13 – Restrições do esquema mediado <i>Sales</i>	32
Figura 14 – Mapeamento de mediação do esquema <i>Sales</i>	32
Figura 15 – Ambiente de mediação <i>Sales</i> com o esquema da <i>Amazon</i>	33
Figura 16 – Novo Mapeamento Local γ_a'	38
Figura 17 – Regras adicionais do tableau.....	40
Figura 18 - Exemplo de um tableau	42
Figura 19 - Procedimento <i>IMPLIES</i>	44
Figura 20 – Grafo de dependência estendido	46
Figura 21 – Criação do Novo Esquema Mediado.....	51
Figura 22 – Mapeamento de mediação revisado	56
Figura 23 – Vocabulário do esquema exportado <i>BN</i>	59
Figura 24 – Restrições do esquema exportado <i>BN</i>	59
Figura 25 – Vocabulário do esquema mediado <i>Sales/BN</i>	59
Figura 26 – Vocabulário do esquema mediado <i>Sales</i>	59
Figura 27 – Vocabulário do Esquema Importado <i>BN</i>	60
Figura 28 – Mapeamento local do esquema <i>BN</i>	60
Figura 29 – Mapeamento de mediação do ambiente de mediação <i>Sales/BN</i>	61
Figura 30 – Restrições do esquema importado de <i>BN</i>	61
Figura 31 – Restrições do esquema mediado revisado <i>Sales/BN</i>	62
Figura 32 – O grafo $G(\Sigma)$ que representa Σ	67
Figura 33 – O grafo $G(\Phi)$ que representa Φ	68
Figura 34 – Construção de Γ que gera o ínfimo de Σ e Φ	72

Lista de exemplos

Exemplo 2-1 – Esquemas <i>Amazon</i> e <i>eBay</i>	26
Exemplo 2-2 – Ambiente Mediado <i>Sales</i>	30
Exemplo 3-1 – Inclusão de consulta	38
Exemplo 3-2 – Consulta Eficaz	38
Exemplo 3-3 – Restrições de subconjunto do esquema importado	39
Exemplo 3-4 – Resolução das restrições de subconjunto com tableau	42
Exemplo 3-5 – Aplicação do procedimento <i>IMPLIES</i> ao Exemplo 3-4	46
Exemplo 3-6 – Problema de inclusão de consulta com <i>IMPLIES</i>	46
Exemplo 4-1 – Definição das restrições do esquema mediado.	55
Exemplo 4-2 – Restrições que geram resultado vazio.	63
Exemplo 4-3 – Geração do grafo $G(\Sigma, \Omega)$ que representa Σ	66
Exemplo 4-4 – Construção do grafo $G(\Phi)$ que representa Φ	67
Exemplo 4-5 – Formalização do Passo C do. Exemplo 4-1	70

Lista de símbolos

\mathcal{L}	Linguagem atributiva
\mathcal{A}	Alfabeto da linguagem atributiva
\top	Conceito/papel universal ou supremo (<i>universal concept/role</i>)
\perp	Conceito/papel vazio ou ínfimo (<i>bottom concept/role</i>)
$(p \circ q)$	Composição
p^{-}	Inversa
$\neg e$	Negação
$e \sqcap f$	Interseção
$e \sqcup f$	União
$\exists p$	Quantificação existencial restrita (<i>restricted existential quantification</i>)
$\exists p.e$	Quantificação existencial irrestrita (<i>full existential quantification</i>)
$\forall p.e$	Restrição de valor ou quantificação universal irrestrita
$(\leq n p)$	Restrição de máximo (<i>at-most restriction</i>)
$(\geq n p)$	Restrição de mínimo (<i>at-least restriction</i>)
s	Interpretação para os símbolos do alfabeto \mathcal{A}
Δ^s	Domínio da interpretação s
$u \sqsubseteq v$	Inclusão
$u \mid v$	Disjunção
$u \equiv v$	Equivalência
$s \models \sigma$	Interpretação s satisfaz a fórmula σ
$s \models \Sigma$	Interpretação s satisfaz todas as formulas do conjunto de fórmulas Σ
$\Sigma \models \sigma$	Conjunto de fórmulas Σ implica logicamente uma fórmula σ
$Th(\Sigma)$	Teoria induzida por Σ
M	Ambiente de mediação
MV	Vocabulário de M
MC	Conjunto de restrições de M
E_i	Esquema exportado da fonte de dados i
EV_i	Vocabulário de E_i
EC_i	Conjunto de restrições de E_i

I_i	Esquema importado referente ao E_i
IV_i	Vocabulário de I_i
IC_i	Conjunto de restrições de I_i
γ	Mapeamento de mediação (<i>mediated mapping</i>)
$\bar{\gamma}$	Função induzida por γ
γ_k	Mapeamento local (local mapping)
$\bar{\gamma}_k$	Função induzida por γ_k
$\Phi_1 \triangle \Phi_2$	Ínfimo ou maior limite inferior (<i>greatest lower bound - g.l.b.</i>)
$\Phi_1 \nabla \Phi_2$	Supremo ou menor limite superior (<i>least upper bound - l.u.b.</i>)