Pontifícia Universidade Católica
do Rio de Janeiro

André Luiz Almeida Marins

Modelos Conceituais para Proveniência

Dissertação de Mestrado

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

Orientador: Prof. Marco Antonio Casanova

Rio de Janeiro,
Março de 2008

PUC-Rio - Certificação Digital Nº 0611888/CA

Pontifícia Universidade Católica
do Rio de Janeiro

André Luiz Almeida Marins

Modelos Conceituais para Proveniência

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. Marco Antonio Casanova
Orientador
Departamento de Informática – PUC-Rio

Prof. Antonio Luz Furtado
Departamento de Informática – PUC-Rio

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

Profª. Melissa Lemos Cavaliére
Departamento de Informática – PUC-Rio

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

Rio de Janeiro, 18 de março de 2008

PUC-Rio - Certificação Digital Nº 0611888/CA

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

André Luiz Almeida Marins

Graduado em Engenharia de Computação pela Pontifícia Universidade Católica do Rio de Janeiro (1996), com especialização em Gestão Executiva no Programa de Desenvolvimento Gerencial (PDG Senior Exec) pelo Instituto Brasileiro de Mercado de Capitais (2001), Atualmente é pesquisador do Laboratório Tecgraf e atua nas áreas de Banco de Dados, Proveniência e Rastreabilidade.

Ficha Catalográfica

Marins, André Luiz Almeida

Modelos Conceituais para Proveniência / André Luiz Almeida Marins; orientador: Marco Antonio Casanova. – 2008.
189 f.; 30 cm

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

1. Informática – Teses. 2. Proveniência. 3. Rastreabilidade. 4. Modelagem Conceitual. 5. Alinhamento de Ontologias. I. Casanova, Marco Antonio. II. Pontifícia Universidade Católica do Rio de Janeiro. Departamento de Informática. III. Título.

CDD: 004

PUC-Rio - Certificação Digital Nº 0611888/CA

Agradecimentos

Aos patrocinadores CNPq e PUC-Rio pelas bolsas de fomento e isenção, obrigado. À Noemi, Hugo, Paulo, Costa, Gustavo, Fábio, Rosana e Eduardo pelas referências creditadas e a eterna amizade, muito obrigado!

Um provérbio árabe que adaptei diz que há quatro tipos de pessoas: aquelas que não sabem e não identificam o que não sabem (tolas); algumas que não sabem e reconhecem que não sabem (humildes); outras que sabem, mas não compreendem o que sabem (aprendizes); e as que sabem e conhecem o que sabem (mestres).

Às vezes, precisei me perder para me encontrar. Percebi que não sabia e, doutores-filósofos me indicaram a direção. Lutei, persisti e corrigi o percurso. Ao Marco (Casa) - professor, orientador e amigo - agradeço os ensinamentos práticos e teóricos, pessoais e profissionais.

(Re)conheci pessoas incríveis que levarei comigo por toda a vida. Elegi duas que representam com louvor todas as demais não mencionadas. Daniela e Ricardo, vocês são demais.

Nas horas mais difíceis assegure. Nas fáceis ratifique. Em ambas, reifique: o amor, o afeto, a amizade. O produto vira serviço exclusivo, terra fértil e água fresca para felicidade que encanta produtor e consumidor. Também registre, registre muito, registre tudo. Para que na falta de memória, a história seja preservada. Para aprender e ensinar depois de apreender. Para sorrir ou chorar, mas viver. Obrigado a minha família, minha base. Renata, você é única, voemos juntos. Rosely e Hélio, minha genética e educação tem raízes nobres em vocês. Rachel e Zuleika, duas gerações diferentes que complementam a minha e, com elas sou muito, mas muito, muito mais feliz.

A todos que contribuíram para o sucesso desta pesquisa reitero: tenho a certeza que tudo é efêmero. Não obstante, as experiências que vivi, estão em meu coração e as levarei sempre comigo. Obrigado por tudo! Vamos em frente!

PUC-Rio - Certificação Digital Nº 0611888/CA

Resumo

Marins, André Luiz Almeida; Casanova, Marco Antonio. **Modelos Conceituais para Proveniência.** Rio de Janeiro, 2008. 189p. Dissertação de Mestrado - Departamento de Informática, Pontifícia Universidade Católica do Rio de Janeiro.

Sistemas de informação, desenvolvidos para diversos setores econômicos, necessitam com maior freqüência de capacidade de rastreabilidade dos dados. Para habilitar tal capacidade, é necessário modelar a proveniência dos dados. A proveniência permite testar conformidade com a legislação, repetição de experimentos e controle de qualidade, entre outros. Habilita também a identificação de participantes (determinantes ou aderentes) como pessoas, organizações, agentes de software entre outros e, permite associá-los a atividades, eventos ou processos. Pode ser utilizada para estabelecer níveis de confiança para as transformações dos dados. Esta dissertação propõe um modelo genérico de proveniência criado com base no alinhamento de recortes de ontologias de alto nível, padrões internacionais e propostas de padrões que tratam direta ou indiretamente de conceitos relacionados à proveniência. As contribuições da dissertação são, portanto em duas direções: um modelo conceitual para proveniência - bem fundamentado – que facilita a interoperabilidade e a aplicação da estratégia de projeto conceitual baseada em alinhamento de ontologias.

PUC-Rio - Certificação Digital Nº 0611888/CA

Palavras-chave

Proveniência; rastreabilidade; modelagem conceitual; alinhamento de ontologias.

Abstract

Marins, André Luiz Almeida; Casanova, Marco Antonio. **Provenance Conceptual Models.** Rio de Janeiro, 2008. 189p. MSc. Dissertation - Departamento de Informática, Pontifícia Universidade Católica do Rio de Janeiro.

Information systems, developed for several economic segments, increasingly demand data traceability functionality. To endow information systems with such capacity, we depend on data provenance modeling. Provenance enables legal compliance, experiment validation, and quality control, among others . Provenance also helps identifying participants (determinants or immanents) like people, organizations, software agents among others, as well as their association with activities, events or processes. It can also be used to establish levels of trust for data transformations. This dissertation proposes a generic conceptual model for provenance, designed by aligning fragments of upper ontologies, international standards and broadly recognized projects. The contributions are in two directions: a provenance conceptual model - extensively documented – that facilitates interoperability and the application of a design methodology based on ontology alignment.

PUC-Rio - Certificação Digital Nº 0611888/CA

Keywords

Provenance; traceability; conceptual modeling; ontology alignment.

Sumário

PUC-Rio - Certificação Digital Nº 0611888/CA

PUC-Rio - Certificação Digital Nº 0611888/CA

PUC-Rio - Certificação Digital Nº 0611888/CA

Lista de figuras

PUC-Rio - Certificação Digital Nº 0611888/CA

PUC-Rio - Certificação Digital Nº 0611888/CA

PUC-Rio - Certificação Digital Nº 0611888/CA

Lista de tabelas

PUC-Rio - Certificação Digital Nº 0611888/CA

PUC-Rio - Certificação Digital Nº 0611888/CA