PONTIFÍCIA UNIVERSIDADE CATÓLICA
DO RIO DE JANEIRO

Geórgia Regina Rodrigues Gomes

Integração de Repositórios de Sistemas de Bibliotecas Digitais e de Sistemas de Aprendizagem

Tese de Doutorado

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

Orientadores: Prof. Rubens Nascimento Melo
Prof. Sean Wolfgand Matsui Siqueira
Profª. Maria Helena Lima Baptista Braz

Rio de Janeiro

Setembro de 2006

Geórgia Regina Rodrigues Gomes

Integração de Repositórios de Sistemas de Bibliotecas Digitais e Sistemas de Aprendizagem

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. Rubens Nascimento Melo
Orientador
PUC-Rio

Profª. Maria Helena Lima Baptista Braz
Co-Orientador
Instituto Superior Técnico - Lisboa

Prof. Sean Wolfgand Matsui Siqueira
Co-Orientador
UNIRIO

Prof. Hugo Fuks
PUC-Rio

Prof. Emmanuel Piseces Lopes Passos
IME-RJ

Prof. Maria Carmen Romcy de Carvalho
Universidade Católica de Brasília - UCB

Prof. Luiz Antonio de Moraes Pereira
Banco Central do Brasil

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

Rio de Janeiro, 29 de setembro de 2006.

Geórgia Regina Rodrigues Gomes

Graduou-se em Matemática pela FAFITA em 1989. Obteve o grau de Mestre em Informática, pela Pontifícia Universidade Católica do Rio de Janeiro (PUC-Rio) em 1999. Trabalhou de 1992 a 2003 como cooordenadora da seção de automação da DBD/PUC-Rio. Desenvolveu junto com a PUCPR o sistema Pergamum – Sistema Integrado de Bibliotecas utilizado hoje com mais de 150 instituições no Brasil. Foi coordenadora acadêmica de dois cursos a distância na PUC-Rio. Atualmente é Professora Adjunta e Pesquisadora da Universidade Cândido Mendes em Campos dos Goytacazes, atuando principalmente nas seguintes áreas: Banco de Dados, Bibliotecas Digitais, Ensino a Distância, Integração de Dados, Padrões de Metadados e Recuperação da Informação.

Ficha Catalográfica

Gomes, Geórgia Regina Rodrigues

Integração de repositórios de sistemas de bibliotecas digitais e sistemas de aprendizagem / Geórgia Regina Rodrigues Gomes ; orientador: Rubens Nascimento ; co-orientadores: Sean Wolfgand Matsui Siqueira, Maria Helena Lima Baptista Braz. – 2006.

143 f. ; 30 cm

Tese (Doutorado em Informática)–Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, 2006.

Incluí referências bibliográficas

1. Informática – Teses. 2. Bibliotecas digitais. 3. Educação baseada na web. 4. Objetos de aprendizagem. 5. Integração de dados. 6. Mineração de texto. 7. Ontologia. 8. Banco de dados. I. Melo, Rubens Nascimento. II. Siqueira, Sean Wolfgand Matsui. III. Braz, Maria Helena Lima Baptista. IV. Pontifícia Universidade Católica do Rio de Janeiro. Departamento de Informática. V. Título.

CDD: 004

A Deus, pela graça alcançada.

Ao meu marido Joney e meu filho Diogo.

Aos meus pais Jorge e Aparecida.

À minha sogra e meu sogro.

Agradecimentos

A Deus, que ilumina meus caminhos e me deu forças a cada minuto para que chegasse até o fim, sem Ele seria impossível.

Ao meu marido Joney Junior, pelo apoio, ajuda e principalmente, pela compreensão e paciência ao longo da realização deste trabalho.

Ao meu filho Diogo, luz que ilumina minha vida, pela aceitação dos momentos que não pude estar ao seu lado, por causa deste trabalho.

Aos meus pais, Aparecida e Jorge, pelo amor, incentivo e carinho que sempre me dedicaram em todos os momentos da minha vida.

À minha sogra D. Lurdinha e meu sogro Sr. Joney, pela ajuda nos momentos mais importantes deste trabalho, cuidando do meu filho com carinho e amor na minha ausência e pela honra de tê-los como sogros.

As minhas irmãs, Mara, Alzira e Rita, por serem minhas irmãs.

Ao meu orientador, Rubens Nascimento Melo, pelos ensinamentos transmitidos durante todos estes anos, apoio e incentivo nos momentos de desanimo, e por ser este ser humano maravilhoso.

Ao Sean Wolfgand Matsui Siqueira, meu co-orientador, pela grande determinação e ajuda durante todo o processo de desenvolvimento desta tese, sem ele seria quase impossível.

A Maria Helena Lima Baptista Braz, minha co-orientadora, pela força e atenção dispensados a mim, principalmente no mês da entrega deste trabalho, sem ela seria muito difícil.

À minha amiga e irmã de coração Diva de Souza e Silva Rodrigues, que em

todos os momentos difíceis desde o mestrado, estava ao meu lado me apoiando e incentivando.

Aos amigos Roberto Rodrigues (marido de minha amiga Diva) e Elizabeth Vitória (minha grande amiga e irmã), pelo tempo dispensado e a grande boa vontade de revisar o texto final da tese, que Deus os mantenha sempre assim.

Aos amigos do TecBD, Simone Leal de Moura, Carlos Eduardo Portela, Álvaro César Pereira Barbosa, Carolina de Lima Aguilar, Luiz Antônio de Moraes Pereira, Fábio André Machado Porto, Fernanda Lima, Julita Glória Machado Cravo, Paulo Sérgio Simões de Araujo, Cássia Blondet Baruque, Sandra Dias de Souza e Fábio Coutinho.

Aos Professores Hugo Fuks, Emmanuel Piseces Lopes Passos, Maria Carmen Romcy de Carvalho, Luiz Antonio de Moraes Pereira, por aceitarem a participar desta banca.

A todos os Professores do Departamento de Informática pelos valiosos ensinamentos ministrados.

Aos amigos, colegas, professores e funcionários da PUC-Rio, que, a seu modo, mesmo que às vezes sem saber, ajudaram direta ou indiretamente na realização deste trabalho.

À PUC-Rio pela bolsa de isenção concedida até o penúltimo semestre deste curso.

À UCAM-Campos, que me abriu as portas num momento que eu precisava muito, isto também teve influência no término deste trabalho.

Ao meu aluno Igor, pela ajuda na implementação da aplicação de extração de informação.

A todas as pessoas que contribuíram direta ou indiretamente para a realização deste trabalho.

Resumo

Gomes, Geórgia R. R.. **Integração de Repositórios de Sistemas de Bibliotecas Digitais e Sistemas de Aprendizagem.** PUC-Rio, 2006.143p. Tese de Doutorado – Departamento de Informática, Pontifícia Universidade Católica do Rio de Janeiro.

Com o uso generalizado das tecnologias de informação no apoio ao ensino, é comum disponibilizar conteúdos digitais, seja através de Sistemas de Bibliotecas Digitais (DLMS) ou de Sistemas de Gerência de Aprendizagem (LMS). No entanto, estes sistemas funcionam de forma independente, têm características diferentes e manipulam tipos diferentes de materiais, sendo seus repositórios com dados e metadados heterogêneos e distribuídos. Os conteúdos destes repositórios seriam melhor aproveitados se estivessem integrados a um ambiente comum, ou fossem acessados de modo integrado a partir dos ambientes de de DLMS e LMS. Nesta tese é apresentada uma visão homogênea dos conteúdos de DLMS e LMS. Para esta homogeneização utilizou-se uma extensão da arquitetura de mediadores e tradutores que trata a integração de metadados, assim como ontologias para tratamento semântico. Foram consideradas ontologias locais para descrever os metadados de cada repositório e uma ontologia global para a integração. No entanto, os documentos dos repositórios dos DLMS tendem a ser monolíticos e não têm um enfoque na reutilização(reuso). Assim, foram definidas regras para extração dos conteúdos mais importantes destes documentos, o que possibilita a reutilização. Esta extração envolve técnicas de mineração de texto e utiliza regras para descobrir as definições contidas nos documentos. Foi desenvolvido um protótipo que demonstra a viabilidade do processo. Para facilitar o entendimento do trabalho, é apresentado um estudo de caso que utiliza a técnica proposta e o protótipo desenvolvido. O trabalho facilita e enriquece o desenvolvimento de materiais de aprendizagem, uma vez que torna os conteúdos de documentos das bibliotecas digitais reutilizáveis e integrados aos Objetos de Aprendizagem (LO) existentes.

Palavras-chave

Bibliotecas Digitais; Educação Baseada na *Web*; Objetos de Aprendizagem; Integração de Dados; Mineração de Texto; Ontologia; Banco de Dados

Abstract

Gomes, Geórgia R. R.. **Integration of Repositories of Digital Library Systems and Learning Management Systems.** PUC-Rio, 2006.143p. PhD. Thesis – Computer Science Department, Pontifical Catholic University of Rio de Janeiro, Brazil

With the widespread use of Information Technology for teaching support, it is usual to made digital content available through Digital Library Systems (DLMS) or Learning Management Systems (LMS).These systems, however, work independently, have different characteristics and manipulate different types of materials, and their data and metadata repositories are heterogeneous and distributed. The content of repositories would be better used if it was integrated in the same environment or accessed in an integrated way from DLMS and LMS. This thesis presents a homogeneous view of DLMS and LMS content. In order to provide such homogenization, it is proposed an extension of the mediator and wrapper architecture for dealing with metadata integration and ontologies for treating semantics. Local ontologies are used for describing each metadata repository, and a global ontology for the integration. As documents of DLMS repositories tend to be monolithic and not to follow a reuse approach, rules for extracting the most important content from these documents were developed in order to make them reusable. This extraction includes text mining techniques as well as rules for discovering definitions embedded in the documents. A prototype was developed which implements the extraction and proves the feasibility of this approach. In order to make the work easier to understand, it is presented a case study that uses the proposed technique and the prototype. The work described in this thesis facilitates and enriches the development of learning material by making the content of digital library documents reusable and integrated to existing learning objects.

Keywords

Digital Library; Web-Based Education; Learning Objects; Data Integration; Text Mining; Ontology; Database

Sumário

Lista de figuras

Lista de tabelas

Abreviaturas e Siglas

ADL	Iniciativa da Secretaria de Defesa dos EUA no sentido de estabelecer um ambiente distribuído de aprendizagem - *Advanced Distributed Learning*
ARIADNE	*Alliance of Remote Instructional Authoring and Distribution Networks for Europe*
DC	*Dublin Core*
DD	Documento Digital
DL	*Digital Library* (Biblioteca Digital)
DLF	*Digital Library Federation*
DLMS	*Digital Library Management Systems*
DLO	*Digital Library Object* (Objeto de Bibliotecas Digitais)
DLOMS	*Digital Library Object Management Systems*
EI	Extração da Informação
F	Medida padrão combinando as métricas P e R
IEEE	*Institute of Electrical and Electronics Engineers*
IMS	*Information Management Systems* (Sistemas de Gerência de Informação)
IR	*Information Retrieval*
KDT	*Text Mining ou Knowledge Discovery from Texts*
LB	*Library of Congress* (Biblioteca do Congresso Americano)
LCMS	*Learning Content Management Systems* (Sistemas de Gerência de Conteúdo de Aprendizagem)
LMS	*Learning Management Systems* (Sistemas de Gerência da Aprendizagem)
LO	*Learning Object* (Objeto de Aprendizagem)
LOM	*Learning Objects Metadata* (Padrão de metadados proposto pelo IEEE para descrição de LOs)
LTSC	*Learning Technology Standards Committee* (Comitê do IEEE responsável pela padronização da tecnologia de aprendizagem)
MARC	*Machine Readable Cataloging*
METS	*Metadata Encoding and Transmission Standard*
OWL	*Web Ontology Language* (Linguagem para definição/especificação de ontologias para a *Web*)
P	*Precision (Precisão)*
PLN	*Processamento de Linguagem Natural*
PGL	*Partnership in Global Learning*
R	*Recall (Abrangência)*
RDA	*Remote Data Access*
RIO	*Reusable Information Objects (Objetos de Informação Reutilizáveis)*
RLO	*Reusable Learning Object*
SCORM	*Sharable Content Object Reference Model*
SBD	*Sistema de Banco de Dados*
SBDH	*Sistema de Banco de Dados Heterogêneos*
SGBD	*Sistema de Gerência de Banco de Dados*
SGBDH	*Sistema de Gerência de Banco de Dados Heterogêneos*
SQL	*Structured Query Language*
TecBD	Laboratório de Tecnologia em Banco de Dados do Departamento de Informática da PUC-Rio
VIO	*Very Important Object*
XML	*Extensible Markup Language*
W3C	*World Wide Web Consortium*
Web	*World Wide Web*

"Quando pensares em desistir, lembre-se que Deus está ao seu lado, para te ajudar a prosseguir, e com Ele, o sonho se torna realidade"

Geórgia Gomes