PONTIFÍCIA UNIVERSIDADE CATÓLICA
DO RIO DE JANEIRO

Susana Rosich Soares Velloso

SQLLOMining: Obtenção de Objetos de Aprendizagem utilizando técnicas de Aprendizado de Máquina

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: Rubens Nascimento Melo

Rio de Janeiro, Julho de 2007

PONTIFÍCIA UNIVERSIDADE CATÓLICA
DO RIO DE JANEIRO

Susana Rosich Soares Velloso

SQLLOMining: Obtenção de Objetos de Aprendizagem utilizando técnicas de Aprendizado de Máquina

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.

Rubens Nascimento Melo
Orientador
PUC-Rio

Ruy Luiz Milidiú
Departamento de Informática - PUC-Rio

Sérgio Lifschitz
Departamento de Informática - PUC-Rio

José Eugênio Leal
Coordenador(a) Setorial do Centro Técnico Científico - PUC-Rio

Rio de Janeiro, 20 de julho de 2007

Susana Rosich Soares Velloso

Graduou-se em Engenharia Mecânica na PUC-Rio em 1994. Atuou como consultora em Banco de Dados para diversas empresas. Possui interesse acadêmico e profissional nas áreas de Inteligência Artificial e Banco de Dados.

Ficha Catalográfica

Velloso, Susana Rosich Soares

SQLLOMining: Obtenção de objetos de aprendizagem utilizando técnicas de aprendizado de máquina / Susana Rosich Soares Velloso; orientador: Rubens Nascimento Melo. – 2007.

118 f.; 30 cm

Dissertação (Mestrado em Informática)–Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, 2007.

Inclui bibliografia

1. Informática – Teses. 2. Educação baseada na Web. 3. Objetos de aprendizagem. 4. Ontologia. 5. Banco de dados para e-Learning. 6. Aprendizado de máquina. 7. Classificação de textos. 8. Naive-Bayes. I. Melo, Rubens Nascimento. II. Pontifícia Universidade Católica do Rio de Janeiro. Departamento de Informática. III. Título.

CDD: 004

Este trabalho é dedicado ao meu marido
pelo amor e companheirismo.

Agradecimentos

À PUC-Rio, ao departamento de informática e ao CAPES pela oportunidade.

Ao meu orientador, Rubens Nascimento Melo, pela motivação, paciência e ajuda.

Aos meus companheiros e companheiras de TecBD, pela motivação e ajuda.

A todos os professores, funcionários do Departamento de Informática pelo apoio dado quando precisei.

Ao professor Ruy Luiz Milidiú pelo apoio e pelo curso "Text Mining", grande inspiração para a minha dissertação.

Ao professor Eduardo Sany Laber pelo curso de PAA e pela carta de recomendação para o meu ingresso no departamento.

Às minhas amigas pelos momentos de pura diversão e pelo incentivo nos momentos difíceis.

Aos meus familiares pela educação e pelo apoio.

Aos professores que participaram da banca examinadora.

Resumo

Velloso, Susana Rosich Soares. **SQLLOMining: Obtenção de Objetos de Aprendizagem utilizando técnicas de Aprendizado de Máquina.** Rio de Janeiro, 2007. 118p. Dissertação de Mestrado - Departamento de Informática, Pontifícia Universidade Católica do Rio de Janeiro.

Objetos de Aprendizagem ou Learning Objects (LOs) são porções de material didático tais como textos que podem ser reutilizados na composição de outros objetos maiores (aulas ou cursos). Um dos problemas da reutilização de LOs é descobri-los em seus contextos ou documentos texto originais tais como livros, e artigos. Visando a obtenção de LOs, este trabalho apresenta um processo que parte da extração, tratamento e carga de uma base de dados textual e em seguida, baseando-se em técnicas de aprendizado de máquina, uma combinação de EM (Expectation-Maximization) e um classificador Bayesiano, classifica-se os textos extraídos. Tal processo foi implementado em um sistema chamado "SQLLOMining", que usa SQL como linguagem de programação e técnicas de mineração de texto na busca de LOs.

Palavras-chave

Educação Baseada na Web; Objetos de Aprendizagem; Ontologia; Banco de Dados para e-Learning; Aprendizado de Máquina; Classificação de Textos; Naive-Bayes

Abstract

Velloso, Susana Rosich Soares. **SQLLOMining: Finding Learning Objects using machine learning methods.** Rio de Janeiro, 2007. 118p. MSc. Dissertation - Departamento de Informática, Pontifícia Universidade Católica do Rio de Janeiro.

Learning Objects (LOs) are pieces of instructional material like traditional texts that can be reused in the composition of more complex objects like classes or courses. There are some difficulties in the process of LO reutilization. One of them is to find pieces of documents that can be used like LOs. In this work we present a process that, in search for LOs, starts by extracting, transforming and loading a text database and then continue clustering these texts, using a machine learning methods that combines EM (Expectation-Maximization) and a Bayesian classifier. We implemented that process in a system called "SQLLOMining" that uses the SQL language and text mining methods in the search for LOs.

Keywords

Web Based Education; Learning Objects; Ontology; Database; e-Learning; Machine Learning; Text Classification; Naive-Bayes; Internet.

Sumário

Lista de figuras

Lista de tabelas

O conhecimento é a pequena porção da ignorância que arrumamos e classificamos.

Ambrose Bierce