
Operações de consulta por similaridade em
grandes bases de dados complexos

Maria Camila Nardini Barioni

SERVIÇO DE PÓS-GRADUAÇÃO DO ICMC-USP

Data de Depósito: 11/08/2006

Assinatura:

Operações de consulta por similaridade em grandes bases de dados complexos

Maria Camila Nardini Barioni

Orientador: Prof. Dr. Caetano Traina Júnior

Tese apresentada ao Instituto de Ciências Matemáticas e de Computação - ICMC-USP, como parte dos requisitos para obtenção do título de Doutor em Ciências - Ciências de Computação e Matemática Computacional.

**USP – São Carlos
Agosto de 2006**

Ao meu querido marido, Humberto.

Agradecimentos

A meu orientador, Prof. Dr. Caetano Traina Júnior, agradeço o apoio, o incentivo e a confiança que dispensou durante a elaboração deste trabalho.

Ao Humberto, meu querido marido e companheiro de todos os momentos, agradeço pelo incentivo, pelo carinho, pela paciência e, principalmente, pela colaboração em vários trabalhos desenvolvidos durante a realização desta tese.

A minha mãe, Juliana, sempre dedicada e atenciosa, agradeço pelo incentivo, amor e carinho. Sem seu apoio fundamental eu não teria chegado até aqui.

Ao meu pai, Carlos Alberto, por ter me mostrado que é preciso lutar por aquilo que se almeja.

Aos meus queridos avós Idalina e Dito que, apesar de não entenderem muito bem o que se faz em um programa de doutorado, sempre me incentivaram a continuar “estudando”.

Aos meus tios mineiros, Laura e Roberto, por vibrarem com cada uma de minhas conquistas.

Ao José Rubens, aos meus irmãos, Guilherme, Gabriel, Rebecca, Levy e Natan, ao meu querido afilhado, Ruan, e demais familiares, pela atenção e carinho.

Aos estimados amigos Thaty e Enzo pelo companheirismo e pelos bons momentos compartilhados.

Gostaria também de agradecer à Profa. Dra. Agma J. M. Traina, pela atenção e pelas sugestões críticas que muito contribuíram para o aprimoramento deste trabalho.

Aos colegas do Grupo de Bases de dados e Imagens, com os quais compartilhei todos esses anos.

Aos funcionários do ICMC, especialmente às secretárias da Seção de Pós-Graduação, Beth, Laura e Ana Paula, pelo apoio técnico e administrativo.

À CAPES, Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, pelo apoio financeiro.

Resumo

Os Sistemas de Gerenciamento de Bases de Dados (SGBD) foram desenvolvidos para armazenar e recuperar de maneira eficiente dados formados apenas por números ou cadeias de caracteres. Entretanto, nas últimas décadas houve um aumento expressivo, não só da quantidade, mas da complexidade dos dados manipulados em bases de dados, dentre eles os de natureza multimídia (como imagens, áudio e vídeo), informações geo-referenciadas, séries temporais, entre outros. Assim, surgiu a necessidade do desenvolvimento de novas técnicas que permitam a manipulação eficiente de tipos de dados complexos. Para atender às buscas necessárias às aplicações de base de dados modernas é preciso que os SGBD ofereçam suporte para buscas por similaridade – consultas que realizam busca por objetos da base similares a um objeto de consulta, de acordo com uma certa medida de similaridade. Outro fator importante que veio contribuir para a necessidade de suportar a realização de consultas por similaridade em SGBD está relacionado à integração de técnicas de mineração de dados. É fundamental para essa integração o fornecimento de recursos pelos SGBD que permitam a realização de operações básicas para as diversas técnicas de mineração de dados existentes. Uma operação básica para várias dessas técnicas, tais como a técnica de detecção de agrupamentos de dados, é justamente o cálculo de medidas de similaridade entre pares de objetos de um conjunto de dados. Embora haja necessidade de fornecer suporte para a realização desse tipo de consultas em SGBD, o atual padrão da linguagem SQL não prevê a realização de consultas por similaridade. Esta tese pretende contribuir para o fornecimento desse suporte, incorporando ao SQL recursos capazes de permitir a realização de operações de consulta por similaridade sobre grandes bases de dados complexos de maneira totalmente integrada com os demais recursos da linguagem.

BARIONI, M. C. N. *Operações de consulta por similaridade em grandes bases de dados complexos*. São Carlos, 2006. 145 p. Tese de Doutorado – Instituto de Ciências Matemáticas e de Computação – ICMC, USP.

Abstract

Database Management Systems (DBMS) were developed to store and efficiently retrieve only data composed by numbers and small strings. However, over the last decades, there was an expressive increase in the volume and complexity of the data being managed, such as multimedia data (images, audio tracks and video), geo-referenced information and time series. Thus, the need to develop new techniques that allow the efficient handling of complex data types also increased. In order to support these data and the corresponding applications, the DBMS needs to support similarity queries, i.e., queries that search for objects similar to a query object according to a similarity measure. The need to support similarity queries in DBMS is also related to the integration of data mining techniques, which requires the DBMS acting as the provider for resources that allow the execution of basic operations for several existing data mining techniques. A basic operation for several of these techniques, such as clustering detection, is again the computation of similarity measures among pairs of objects of a data set. Although there is a need to execute these kind of queries in DBMS, the SQL standard does not allow the specification of similarity queries. Hence, this thesis aims at contributing to support such queries, integrating to the SQL the resources capable to execute similarity query operations over large sets of complex data.

Gracias por visitar este Libro Electrónico

Puedes leer la versión completa de este libro electrónico en diferentes formatos:

- HTML(Gratis / Disponible a todos los usuarios)
- PDF / TXT(Disponible a miembros V.I.P. Los miembros con una membresía básica pueden acceder hasta 5 libros electrónicos en formato PDF/TXT durante el mes.)
- Epub y Mobipocket (Exclusivos para miembros V.I.P.)

Para descargar este libro completo, tan solo seleccione el formato deseado, abajo:

