Pontifícia Universidade Católica
do Rio de Janeiro

José Antonio Fernandes de Macêdo

Um modelo conceitual para biologia molecular

Tese de Doutorado

Tese apresentada ao Programa de Pós–graduação em Informática do Departamento de Informática da PUC–Rio como parte dos requisitos parciais para obetnção Do título de Doutor em Informática

Orientador: Prof. Edward Hermann Haeusler

Rio de Janeiro
Agosto de 2005

Pontifícia Universidade Católica
do Rio de Janeiro

José Antonio Fernandes de Macêdo

Um modelo conceitual para biologia molecular

Tese apresentada ao Programa de Pós–graduação em Informática do Departamento de Informática do Centro Técnico Científico da PUC–Rio como parte dos requisitos parciais para obetnção Do título de Doutor em InformáticaAprovada pela Comissão Examinadora abaixo assinada.

Prof. Edward Hermann Haeusler
Orientador
Departamento de Informática — PUC–Rio

Prof. Rubens Nascimento Melo
Departamento de Informática - PUC-Rio

Prof. Daniel Schwabe
Departamento de Informática - PUC-Rio

Prof. Marta L. Queirós Mattoso
Departamento de Informática - UFRJ

Prof. Antonio Basílio Miranda
Departamento de Bioquímica - Fiocruz

Prof. José Eugênio Leal
Coordenador Setorial do Centro Técnico Científico —
PUC–Rio

José Antonio Fernandes de Macêdo

Cursou Tecnólogo em Processamento de Dados na PUC-Rio em 1998. Atuou em empresas como Gerente de Informática, Analista de Sistemas, Administrador de Bancos de Dados, Programador de 1986 a 1991. Retornou ao mundo acadêmico para realizar Mestrado em Banco de Dados na PUC-Rio, encerrado em 2000, com publicações na área de Banco de Dados. Ministrou disciplinas relacionadas à Modelagem de Dados, Banco de dados, Linguagens de Programação em diversos cursos, inclusive no Bacharelado em Informática da PUC-Rio. Lecionou no curso de Administração e Tuning de Banco de Dados do CCE PUC-Rio. Possui interesse acadêmico e profissional nas áreas de Banco de Dados, Desenvolvimento de Aplicações baseadas em objetos e Engenharia de Software.(Rio de Janeiro, Brasil)

Ficha Catalográfica

Macêdo, José Antonio Fernandes de

Um modelo conceitual para biologia molecular/ José Antonio Fernandes de Macêdo; orientador: Edward Hermann Haeusler . — Rio de Janeiro : PUC–Rio, Departamento de Informática, 2005.

v., 93 f: il. ; 29,7 cm

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

Inclui referências bibliográficas.

1. Informática – Teses. 2. Banco de Dados. 3. Modelagem Conceitual. 4. Biologia Molecular. 5. Bioinformática. I. Haeusler, Edward Hermann. II. Pontifícia Universidade Católica do Rio de Janeiro. Departamento de Informática. III. Título.

CDD: 510

Aos meus pais, Lásaro e Maria Lúcia, que nunca mediram esforços para me oferecer a melhor educação. À minha esposa Janaína, pela seu amor, carinho, confiança e paciência, sem os quais teria sido muito difícil esta jornada. E ao meu filho Antonio que em breve nascerá e a quem aguardamos com muita ansiedade e amor.

Agradecimentos

Ao Professor Sérgio Lifschitz, orientador deste trabalho, pela motivação, pela disponibilidade e pela amizade que demonstrou ao longo de todo nosso convívio.

Aos Professores Fábio Porto e Philippe Picouet, pelas boas discussões e acolhida quando estive longe do Brasil.

Ao professor Edward Hermann Haeusler, pela ajuda na busca de soluções para os problemas encontrados neste trabalho.

Aos funcionários da secretaria, da biblioteca e do Lab-DI, pela atenção e presteza com que sempre atenderam às minhas solicitações.

Aos amigos da PUC, que tive a felicidade de conhecer durante esta minha jornada. Eles ajudaram a abrandar os momentos mais difíceis.

Resumo

Macêdo, José Antonio Fernandes de; Haeusler, Edward Hermann. **Um modelo conceitual para biologia molecular**. Rio de Janeiro, 2005. 93p. Tese de Doutorado — Departamento de Informática, Pontifícia Universidade Católica do Rio de Janeiro.

Projetos de genômica e biológica molecular estão gerando dados cujos volumes e complexidades jamais foram observados nesta área. Além disso, fontes de dados e de conhecimento são produzidas e utilizadas por grupos de pesquisa os quais utilizam terminologias diferentes (sinônimos, apelido e fórmulas), sintaxes diferentes (estrutura de arquivos e separadores) e semânticas diferentes (intra e interdisciplinares homônimos). Desta forma, o sucesso da pesquisa em biologia dependerá da correta representação e manipulação dos dados biológicos permitindo os cientistas criarem, gerenciarem, manipularem, integrarem e analisarem os dados de forma a gerar informação e conhecimento. Neste trabalho, estudamos os problemas para representação de dados biológicos apresentados nas principais linguagens de modelagem tradicionais. Em seguida, levantamos os requisitos para um novo modelo de dados conceitual para biologia molecular. Finalmente, propomos um novo modelo conceitual contendo construtores específicos para solucionar alguns dos problemas estudados. Além disso, formalizamos o modelo proposto usando lógica de primeira ordem e a utilizamos para realizar inferências, as quais possam auxiliar o projetista de banco de dados na criação de um esquema conceitual de dados.

Palavras–chave

Informática – Teses; Banco de Dados, Modelagem Conceitual, Biologia Molecular, Bioinformática

Abstract

Macêdo, José Antonio Fernandes de; Haeusler, Edward Hermann. **A conceptual model for molecular biology**. Rio de Janeiro, 2005. 93p. PhD. Thesis — Departamento de Informática, Pontifícia Universidade Católica do Rio de Janeiro.

Genomic and molecular biology projects are generating knowledge data whose volume and complexity are unparalleled in this research area. In addition, data and knowledge sources produced and used by research groups have terminological differences (synonyms, aliases and formulae), syntactic differences (file structure, separators and spelling) and semantic differences (intra- and interdisciplinary homonyms). In this context, data management techniques play a fundamental role for biological applications development because it offers adequate abstractions to design, implement, access and manage data, in order to generate knowledge. In this work, we study the representation problems presented in traditional languages. Following, we raise the main requirements for a new conceptual data model specially conceived for molecular biology. Finally, we propose a new conceptual data model with special types of constructor trying to solve some of the representation problems discussed before. In addition, we formalize our proposed model using first-order logic and we use this logical description to infer some properties that may help database designer during the elaboration of a conceptual database schema.

Keywords

Computer Science; Database Systems; Conceptual Modeling; Molecular Biology; Bioinformatics.

Conteúdo

Lista de Figuras

Lista de Tabelas