

Adaptação e Extensão do Modelo MER para Representação de Dados Semiestruturados

Abstract. *To build a solid database for any application, an abstraction of the data schema is required, which we call the ERM (Entity-Relationship Model). A well-designed data model can result in a cohesive database. The ERM is based on rules for data management and distribution, these rules known as normal forms. These normal forms define the structure of how data will be organized within the model and the actual database. In data modeling, we have a way to represent data and their relationships to later implement the database. Data modeling addresses various aspects such as relationships, data types, cardinality, keys (primary and/or foreign), generalizations, and specializations. Among these aspects, we focus on data types, proposing a method to adapt and extend the representation of semi-structured data types such as JSON/XML. These data, which are typically a collection of separate elements, are accommodated in relationships following a context that makes sense for modeling. The proposal to study, adapt, and extend the model to accommodate this scenario can introduce a new abstraction for understanding data, potentially facilitating database construction.*

Resumo. *Para a construção de uma boa base de dados para qualquer aplicação necessita de uma abstração do esquema de dados que nos nomeamos de MER (Modelo entidade relacionamento). Um bom modelo de dados pode gerar uma boa base coesa. O MER se baseia em regras para o gerenciamento e distribuição dos dados, essas regras e definições nomeamos de formas normais. Essas formas normais definem a estrutura de como os dados serão comportados no modelo e na base de dados real. No modelo de dados nos temos uma maneira de representar os dados e suas relações para posteriormente implementar a base de dados. A modelagem de dados aborda diversas situações como relações, tipo de dados, cardinalidade, chaves(primária e/ou estrangeira), generalizações e especializações. Dentro dessas situações atacaremos no tipo de dados propondo uma maneira de adaptar e estender a representação de tipo de dados semiestruturados como JSON/XML. Esses dados que por sua vez é um conjunto de dados, normalmente é separado e acomodado em relações seguindo um contexto que faz sentido para a modelagem. A proposta de estudar, adaptar e estender o modelo para que acomode essa situação, pode trazer uma nova abstração para o entendimento dos dados podendo facilitar a construção da base de dados.*

1. Introdução

Modelagem de dados segundo [THALHEIM 1991] é uma técnica de abstração e representação de dados para a construção de bases de dados e que desde a década de 1970 eles são utilizados para este fim. Segundo ele ainda, a modelagem é um facilitador de design de banco de dados sem que o usuário tenha total conhecimento de regras e formas para a construção.

Soluções que são proveniente de softwares devem se preocupar com a qualidade e integridade dos dados que compõe sua base. Nesse aspecto [MICROSOFT] aborda o conceito das formas normais para que tudo isso seja contemplado. Consolidando essa necessidade de ter qualidade e integridade dos dados [PRATT 1985] em seu artigo "*A relational approach to database design*" comenta a importância da estruturação dos dados de maneira correta seguindo as formas normais. O autor aponta que seguindo as definições teremos uma qualidade melhor na distribuição dos dados.

Porem temos uma definição de forma normal em sua primeira abordagem, que diz para identificar e solucionar dados e multivalorados, e que a possível solução é a distribuição destes dados em tabelas segundo cada contexto [MICROSOFT].

Profissionais deste seguimento que trabalham para gerar bases de dados a fim de solucionar problemas, tarefas ou qualquer tipo de necessidade do cotidiano se deparam com modelagem de dados. Normalmente trabalhos que propõe alguma solução que tem uma base de dados, recorre ao MER (Modelo Entidade-Relacionamento). Na modelagem de dados é identificado contextos podendo separar os dados para que a distribuição fique granulada e coesa de acordo com aspecto de cada dado. Cada dado faz parte de um contexto que integra outros contextos gerando as tabelas e suas relações.

Quando construímos os modelos de dados indiretamente utilizamos as definições de formas normais, já pensando no banco de dados real. Nossa proposta vem para a tentar contemplar a abstração de um dado que não é comum, o dado semiestruturado.

Baseando-se nessa abordagem nosso objetivo é adaptar e estender a modelagem para que dados semiestruturados possam ser representados de maneira fácil e objetiva, de forma que, o entendimento seja instantâneo.

Para que possamos representar dados semiestruturados na modelagem de dados temos que nos atentar aos prós e contras desta proposta, [LANDER 1994] investiga recursos de integridade em SQL para ser representados na modelagem. Como nossa proposta vem para a extensão do modelo queremos seguir as normas de integridade para que não haja nenhum problema com a questão ACID (Atomicidade, Consistência, Isolamento e Durabilidade). Por outro lado [THALHEIM 1991] que propõe a extensão da modelagem para melhor compreensão, segundo ele, o entendimento do modelo não necessita do conhecimento da normalização. [FONG 1995] mostra uma proposta de orientação a objetos aplicada na modelagem de banco de dados fazendo correlações de estruturas dentro de um problema de transição de bases.

A adaptação do MER proverá uma nova leitura para a construção da base de dados. Essa representação poderá ajudar as abstrações para o armazenamento dos dados de característica semiestruturado. Com a representação em modelo, o entendimento desses dados se tornam mais claros ajudando a implementação. Com essa nova maneira de representação pode-se dar mais amplitude a banco de dados relacionais para dados semiestruturados. O trabalho tende a priorizar o uso de base de dados relacionais para esses dados, sem a necessidade de uso de uma base de dados NoSQL(Not Only SQL).

Para validar a hipótese, será desenvolvida uma adaptação do Modelo Entidade-Relacionamento (MER) que permita a representação de dados semiestruturados. Essa adaptação será avaliada considerando a capacidade de representar tais dados de forma clara e compreensível, mantendo as normas de integridade e qualidade exigidas por bancos de dados relacionais. Além disso, realizaremos testes práticos de modelagem e implementação para verificar a viabilidade da proposta, analisando sua aplicabilidade na

estruturação e armazenamento de dados semiestruturados sem a necessidade de utilizar bases NoSQL.

2. Trabalhos Relacionados

Segundo [FONG 1995] em seu artigo “Mapping Extended Entity Relationship Model to Object Modeling Technique” propõe modelagem de dados com orientação a objetos. Ele destaca que o objetivo principal é demonstrar como o mapeamento entre esses dois modelos, modelo Entidade-Relacionamento Estendido (EER) para o modelo de Técnica de Modelagem de Objetos (OMT), pode facilitar a transição de bancos de dados tradicionais baseados em registros para bancos de dados orientados a objetos. Na estratégia apresentada ele pontua as seguintes correlações:

Entidades → Classes: Cada entidade do modelo EER é mapeada para uma classe no modelo OMT;

Relacionamentos → Associações: Os relacionamentos entre entidades no EER são convertidos em associações e links no OMT;

Generalização → Métodos: Entidades semelhantes podem ser unificadas em uma classe geral, com subclasses herdando suas propriedades;

Categorização → Herança múltipla: Superclasses independentes podem ser unificadas em subclasses que herdam características de múltiplas classes;

Relacionamento "é-um" (isa) → Herança: Classes especializadas herdam propriedades e operações de uma classe mais geral;

Entidades fracas → Classes componentes: Dependências entre entidades são mantidas no OMT como classes componentes;

Agregação → Objetos compostos: Grupos de entidades relacionadas no EER são convertidos em classes compostas no OMT.

[LAENDER 1994] em seu trabalho nomeado “An Analysis of SQL Integrity Constraints from an Entity-Relationship Model Perspective” analisa as restrições de integridade do SQL sob a ótica do modelo entidade-relacionamento (ER). Ele investiga como os recursos de integridade do SQL podem ser usados para representar corretamente esquemas ER e propõe melhorias para otimizar essa conversão.

O estudo destaca que, enquanto o SQL fornece um suporte flexível para a integridade referencial, algumas de suas limitações tornam certas representações do modelo ER complexas e custosas. O artigo sugere pequenas extensões ao SQL para facilitar a conversão e otimização de esquemas ER para SQL.

[THALHEIM 1991] em seu trabalho “Extending the Entity-Relationship Model for a High-Level, Theory-Based Database Design” propõe uma extensão do modelo entidade-relacionamento (ERM) para um modelo mais avançado, chamado Higher-order Entity-Relationship Model (HERM).

O objetivo é facilitar o design de bancos de dados sem exigir que o usuário tenha conhecimento profundo de teoria da normalização ou restrições de implementação. O artigo destaca que, desde os anos 1970, várias metodologias de design de banco de dados foram desenvolvidas, mas muitas dependem do conhecimento e experiência do analista de banco de dados. O modelo proposto, HERM, busca simplificar esse processo e permitir que mesmo designers iniciantes possam criar esquemas eficientes.

O HERM amplia o modelo entidade-relacionamento tradicional ao permitir relacionamentos de ordem superior, ou seja, relações entre relações, suporte para hierarquias e subclasses, algo que o modelo ER clássico não representa de forma natural e representação de conjuntos, sequências e valores nulos, que são difíceis de modelar no ERM tradicional.

O autor propõe uma melhora significativa na modelagem conceitual de bancos de dados, corrigindo as limitações do modelo ER tradicional. Ele oferece uma abordagem mais expressiva e flexível, mantendo a simplicidade do ERM e permitindo um design mais acessível para usuários sem conhecimento avançado em teoria de bancos de dados.

Referências

PRATT, Philip J. A relational approach to database design. SIGCSE Bulletin, v. 17, n. 1, p. 184–201, mar. 1985. Disponível em: <https://doi.org.ez31.periodicos.capes.gov.br/10.1145/323275.323315>. Acesso em: 4 jan. 2025.

FONG, Joseph. Mapping extended entity relationship model to object modeling technique. SIGMOD Record, v. 24, n. 3, p. 18–22, set. 1995. DOI: 10.1145/211990.212007. Disponível em: <https://doi.org.ez31.periodicos.capes.gov.br/10.1145/211990.212007>. Acesso em: 30 jan. 2025.

LAENDER, A.H.F.; CASANOVA, M.A.; DE CARVALHO, A.P.; RIDOLFI, L.F.G.G.M. An analysis of SQL integrity constraints from an entity-relationship model perspective. Information Systems, v. 19, n. 4, p. 331–358, 1994. DOI: 10.1016/0306-4379(94)90018-3. Disponível em: <https://www.sciencedirect.com/science/article/pii/0306437994900183>. Acesso em: 30 jan. 2025.

THALHEIM, B. Extending the entity-relationship model for a high-level, theory-based database design. In: SCHMIDT, J.W.; STOGNY, A.A. (eds.). Next Generation Information System Technology. EWDW 1990. Lecture Notes in Computer Science, v. 504. Berlin, Heidelberg: Springer, 1991. DOI: 10.1007/3-540-54141-1_10. Disponível em: https://doi.org/10.1007/3-540-54141-1_10. Acesso em: 30 jan. 2025.

MICROSOFT. Descrição da normalização de banco de dados: o que é normalização? Disponível em: <https://learn.microsoft.com/pt-br/office/troubleshoot/access/database-normalization-description>. Acesso em: 1 fev. 2025.