Extending the Entity-Relationship Model for a High-Level, Theory-Based Database Design

Bernhard Thalheim

Computer Science Department
University of Rostock
Albert-Einstein-Str. 21
D-O-2500 Rostock, FRG

Abstract

Database design methodologies should facilitate database modeling, effectively support database processing and transform a conceptual schema of the database to a high-performance database schema in the model of the corresponding DBMS. Since the late 1970's, various methodologies for database design have been introduced. Most of them, however, are dependent on the knowledge, comprehension and experience of the database analyst and their knowledge in normalization theory. The proposed methodology does not require the user to understand the theory, the implementational restrictions and the programming problems in order to design a database scheme. A novice designer can create a database design successfully using this method. The Entity-Relationship Model is extended to the **Higher-order Entity-Relationship** Model (HERM) by relationships of higher degrees and relationships of relationships. This model is used for a high-level database design system DBDB (DataBase Design by Beta). The HERM supports an efficient translation to nested relational, relational, network and hierarchical schemes. The model has the expressive power of semantic models and possesses the simplicity of the entity-relationship model.

1 Introduction

A large number of conceptual data models have been proposed. However, actual experience with the use of these models as a basis for implementing a generalized DBMS is very scant. Nearly all early commercial DBMS implementations were based on the hierarchical model such as IMS and SYSTEM-2000 or the network model such as IDS and IDMS or the relational model such as INGRES, DB2. The relational data model was proposed as a simple and theoretically well-founded representation of data, and it has soon become the most important model for database systems (see for example [PDG89, Ull89, Vos87]). The primary virtues of the model are its rigorous mathematical foundation and the correspondence of a relation with the notion of a table. Research efforts have highlighted drawbacks to the relational model. Rather than abandon the relational paradigm because of these considerations, we are interested in extending relational languages in a way that incorporates useful ideas from alternative language paradigms but allows the retention of most, if not all, of the advantages of the relational approach. A rich harvest of theoretical results [Tha90] have resulted from the study of relational models and should be used in other models. The classical Relational Model deals only with flat relations. It is not aware of any distinction between entity relations and relationship relations. In contrast, models like the network model, the hierarchical model and the entity-relationship model make distinctions between these two types of relations.

A conceptual scheme is a global description of the database that hides the details of physical storage structures and concentrates on describing entities, data types, relationships, and constraints. A high-level datamodel can be used at this level. Recent research in database design methods has developed the idea of using two distinct data models at the conceptual level. An enhanced conceptual model would provide an effective means of describing the database application environment. A representational data model would be used for efficient translation of a scheme into physical data structures. For instance, the relational data model could be employed as a representational data model. One of the most accepted enhanced conceptual models has been the entity-relationship model (ERM) of Chen [Che76, Che83]. It has been recognized as an excellent tool for high level database design because of its many convenient facilities for the conceptual modeling of reality. Its basic version deals with more static structural properties, such as entities, attributes and relationships. The shortcomings known for the relational model can be listed also for the ERM. One reason is that this model is unable to represent hierarchical and higher-order relationships. Other problems known for the entity-relationship model are the following:

- Only first-order relationships can be modeled. But the relational design leads already to higher-order relationships.
- Is-a-relationships cannot be modeled naturally.
- The concept of weak entities is not theoretically based.
- The classic entity-relationship model does not use n-ary relationships.
- The solution [Teo89] to define new entities as clusters of entities and relationships leads to a loss of information.
- Sets, sequences, null-valued relationships can not be simply represented.

It will be shown in this paper that the Higher-order Entity-Relationship Model (HERM) generalizes the solutions for the presented shortcomings. The extension in HERM may directly represent the above semantic concepts. We accomplish these extensions by introducing one additional generalization of the constructs of the entity-relationship model: the concept of higher-order relationships. The concept of higher-order relationships handles both the multiple entity set participation in a given role of the relationship, as well as the subclass, superclass and clustering concepts. The proposed concept generalizes furthermore the proposed concepts of data abstraction. Since the model incorporates