Chapter III
Survey of Cardinality Constraints in Snapshot and Temporal Semantic Data Models

Faiz Currim
University of Iowa, USA

Sudha Ram
University of Arizona, USA

ABSTRACT

Cardinality captures necessary semantics in conceptual data modeling and determines how constructs are translated into relations. Business policies in a variety of domains like healthcare, education, supply chain management and geographic systems are often expressible in terms of cardinality. The knowledge about cardinality constraints is also useful during schema integration, in query transformation for more efficient search strategies, and in database testing. Practically every conceptual modeling grammar provides support for this kind of constraint, and in an effort to resolve the variations in semantics past research has studied the different types of cardinality constraints. None have been so far comprehensive, and further there has been very little coverage of the concept in temporal domain even though it provides some interesting extensions to the concept. This study considers existing work in snapshot and temporal cardinality and suggests some areas for future work.

INTRODUCTION

The last three decades have seen active research in the area of database design and modeling. A number of modeling grammars and implementation techniques have been proposed, including popular standards like the Entity Relationship (ER) model and the Unified Modeling Language (UML). Both ER and UML were designed as general-purpose models, and we have seen the development of model extensions to capture the semantics in specialized domains (e.g., for scientific, healthcare, and temporal applications). In various forms, these models all address important design needs of documenting and communicating the database schema, and are consequently popular in industry and academia. One would be hard-pressed to find a database textbook that did not include some conceptual model variant, and likewise most database CASE tools incorporate them in as well.

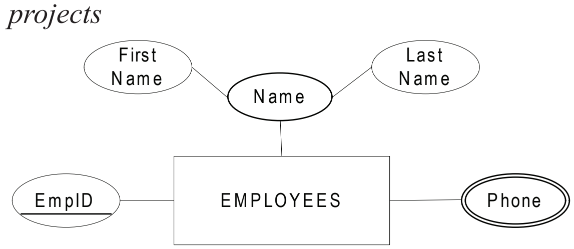
A number of grammars have been developed for snapshot and temporal data. Their popularity and importance can also be measured via a surrogate of the number of surveys and research commentaries developed for conceptual modeling (Gregersen & Jensen, 1999; Hull & King, 1987; Parent *et al.*, 1999; Peckham & Maryanski, 1988; Tryfona & Jensen, 1999; Wand & Weber, 2002). An important aspect of such models is the expression of data constraints (Ram & Khatri, 2005). The visible representation of rules helps organizations in a number of ways including better capturing of semantics, as an aid to translation of the schema, in search and query strategies.

Most conceptual models capture business policies that determine cardinality. However, there is a wide variation in how grammars treat the semantics of cardinality and how many different types of cardinality constraints they represent. Some consider cardinality as applied to relationships, while others also take into account attributes and classes. Cardinality for attributes is often integrated into the semantic model constructs by use of special symbols such as shading mandatory attributes (i.e., minimum cardinality of 1) or using some symbolic construct like a double-lined oval for a multi-valued attributes (maximum cardinality $\geq$ 2). Other useful and related structural constraints like identification (where the cardinality of the attribute domain exactly matches the cardinality of its associated entity set) and composition (attributes with degree > 1 or component attributes) are also represented. In Figure 1, which uses the notation syntax adopted by a popular database text book[1] (Elmasri & Navathe, 2006), we see EmpID is an identifier, Name is a composite attribute, and Phone is a multi-valued attribute for the EMPLOYEES class.

There are a number of other data constraints besides cardinality. For instance, when discussing attributes, one could include constraints on the range of values an attribute can take, including restrictions determined by membership in relationships or subclasses. Often, a simple annotation to the schema or data dictionary is made. For example, the Semantic Database Model (SDM) (Hammer & McLeod, 1981) uses value classes and derivations which are specified in the schema data-dictionary. Aiming to survey and classify all possible rules is a huge task, and would go well beyond the scope of a single chapter.

In this work, we focus on cardinality rules. This is a subset of the possible data integrity rule types, and we refer the reader to work by Thalheim (Thalheim, 1996) that discusses the various constraint categories. Cardinality is an interesting type of rule for a number of reasons, including the variety of constraint sub-types, the ability to formalize rule semantics via first-order-logic and consequently reason about the rules and potential conflicts. Further, a lack of understanding of the distinction among cardinality types can lead to miscommunication (for those following a different scheme) about the data semantics and consequent translation, or a persistent misconception that cardinality is a difficult concept in conceptual modeling.

Figure 1. An example of employees working on projects



Related Content