Chapter I
Mapping Generalizations and Specializations and Categories to Relational Databases

Sikha Bagui
University of West Florida, USA

INTRODUCTION

An Entity Relationship (ER) model that includes all the concepts of the original ER model and the additional concepts of generalizations/specializations and categories is often referred to as the Extended ER (EER) model (Elmasri & Navathe, 2007). With the rising complexity of database applications, and also in light of today's web data applications (Necasky, 2006), the basic concepts of the ER model, as originally developed by Chen(1976), are no longer sufficient. Hence the basic ER model has been extended to include generalizations and specializations (Bagui & Earp, 2003; Elmasri & Navathe, 2007), and the concept of categories (Elmasri, et al., 1985). In this short article we shed some light on these relationship concepts, concepts that database designers often find difficult to directly model (Engels et al., 1992/93). We also discuss the mapping rules for generalizations/specializations and categories. Important contributions in this area are also reported in (Elmasri et al., 1985; Gogolla & Hohenstein, 1991; Markowitz & Shoshani, 1992; Dey, et. al., 1999). Dullea, et. al. (2003) discusses the structural validity of modeling structures with ER models.

Due to the short nature of this paper, we will keep the discussion in this paper focused on implementing generalizations and specializations in relational databases; their parallel implementation in objects will not be covered. Also, the discussion of the concept of inheritance will center around generalizations/specializations and categories in EER diagrams, without getting into an almost equivalent notion in Object-oriented (OO) theory, ORM (Object-Role Modeling) and UML (Unified Modeling Language) class diagrams.

BACKGROUND

A generalization/specialization relationship models a superclass/subclass type of relationship. A generalization is an abstraction mechanism that allows for viewing of entity-sets as a single generic entity-set. The attributes and associations which are common to the entity-sets are associated with the generic (generalized) entity-set. The inverse of generalization is called specialization.

GENERALIZATION / SPECIALIZATION RELATIONSHIPS

If we are modeling a hospital database, for example, and we want to store information about the hospital's nurses, technicians, and physician assistants, we could create separate entities such as NURSE, TECHNICIAN and PHYSICIAN ASSISTANT. But, these three entities would also have a lot of fields in common, for example, name, social security number, address, phone, etc. may be common to all three entities. So, it would be a good idea to have an entity set called EMPLOYEE containing these common fields, and entity subsets, NURSE, TECHNICIAN and PHYSICIAN ASSISTANT, that could inherit this information from the EMPLOYEE entity set. In this case the EMPLOYEE entity set would be called the ***superclass***. This superclass is a ***generalization*** of the entity subsets, NURSE, TECHNICIAN and PHYSICIAN ASSISTANT. The NURSE, TECHNICIAN and PHYSICIAN ASSISTANT would be called the ***subclasses***. The subclasses are ***specializations*** of the superclass, EMPLOYEE, and inherit from the superclass. Several specializations can be defined for the same entity type (or superclass).

The subclass, denoted by a separate entity rectangle in the EER diagram, is considered to be a *part* of the superclass entity set, EMPLOYEE. Although it will have attributes that distinguish it from other subclasses, it is considered only a subset of the EMPLOYEE entity set. That is, all nurses are employees, but the reverse is not true - not all employees are nurses. Likewise, all technicians or physician assistants are employees, but all employees are not technicians or physician assistants.

Figure 1 shows this generalization/specialization example. We use Elmasri and Navathe's (2007) diagrammatic notations for the EER diagrams. The subset symbol, "$\subset$", indicates the direction of the superclass/subclass or parent-child, inheritance relationship. This superclass/subclass relationship is also often referred to as a ***IS-A*** or ***IS-PART-OF*** relationship (Sanders, 1995).

Constraints on Generalization/ Specialization Relationships

Generalizations and specializations can have two types of constraints: (i) the ***disjoint/overlap*** relationship constraint, and, (ii) participation constraints – total or partial. The combinations of these constraints can be: (i) disjoint and total participation; (ii) disjoint and partial participation; (iii) overlap and total participation; (iv) overlap and partial participation. First we will discuss disjoint/overlap relationship constraints and then we will discuss participation constraints, giving examples of combinations of the constraints along the way.

Disjoint/Overlap Relationship Constraints

Generalization/specialization relationships may be ***disjoint*** or they may ***overlap***. A disjoint relationship is shown by a "d" in the circle attaching the superclass to the subclass or subclasses (as shown in Figure 1). A disjoint relationship means that an entity from the superclass can belong to only one of the subclasses (can be of only one specialization). For example, according to figure 1, an EMPLOYEE can be at most a member of only one of the subclasses – PHYSICIAN ASSISTANT,

Related Content

Resource Provisioning and Scheduling of Big Data Processing Jobs
Rajni Aronand Deepak Kumar Aggarwal (2018). *Handbook of Research on Big Data Storage and Visualization Techniques (pp. 382-401).*
www.irma-international.org/chapter/resource-provisioning-and-scheduling-of-big-data-processing-jobs/198771

Dynamic Workflow Restructuring Framework for Long-Running Business Processes
Ling Liu, Calton Puand Duncan Dubugras Ruiz (2005). *Advanced Topics in Database Research, Volume 4 (pp. 1-49).*
www.irma-international.org/chapter/dynamic-workflow-restructuring-framework-long/4367

Validating an Evaluation Framework for Requirements Engineering Tools
Raimundas Matulevicius (2005). *Information Modeling Methods and Methodologies: Advanced Topics in Database Research (pp. 148-174).*
www.irma-international.org/chapter/validating-evaluation-framework-requirements-engineering/23013

Discovering and Analysing Ontological Models From Big RDF Data
Carlos R. Rivero, Inma Hernández, David Ruizand Rafael Cochuelo (2015). *Journal of Database Management (pp. 48-61).*
www.irma-international.org/article/discovering-and-analysing-ontological-models-from-big-rdf-data/142072

Multi-Label Classification: An Overview
Grigorios Tsoumakasand Ioannis Katakis (2009). *Database Technologies: Concepts, Methodologies, Tools, and Applications (pp. 309-319).*
www.irma-international.org/chapter/multi-label-classification/7918