International Journal of Information and Communication Technology Education
Volume 16 • Issue 2 • April-June 2020

Students' Difficulties in Identifying the Use of Ternary Relationships in Data Modeling

Rami Rashkovits, Yezreel Valley College, Jezreel Valley, Israel
Ilana Lavy, Yezreel Valley College, Jezreel Valley, Israel

ABSTRACT

The present study examines the difficulties novice data modelers face when asked to provide a data model addressing a given problem. In order to map these difficulties and their causes, two short data modeling problems were given to 82 students who had completed an introductory course in database modeling. Both problems involve three entity sets with relationships between them, either ternary or binary. The students' solutions were classified according to the types of errors they committed. More than half of the students provided faulty solutions. After an analysis of these results, open interviews were conducted with a selected group of students in order to figure out the reasons underlying the students' erroneous decisions regarding the data model. Among the reasons for their erroneous solutions were insufficient experience, lack of reflection on their solution, and lack of immediate feedback. In addition, the authors suggest instructional modifications derived from the research results.

KEYWORDS

Cardinality of Relationship, Data Model, Novice Data Modelers, Ternary Relationship

INTRODUCTION

The research literature includes an academic discussion on the difficulties novice modelers may encounter when designing a data model addressing given requirements in general (Dey et al., 1999) and regarding ternary relationships in particular (Hitchman, 2003; Batra, 2007). They encounter many difficulties, mostly concerned with cognitive complexity; among them, No Flexibility for errors, lack of immediate feedback, and information overload (Batra, 2007). As a result, data models designed by novice modelers tend to be inaccurate and erroneous, and hence the cause for the faulty behavior of information systems.

During their studies, novice data modelers study how to design a data model addressing given requirements. They study how to identify entities and how to set relationships between them. They also learn how to transform the entities and relationships into tables, fields and keys in order to form a relational schema.

One of the main challenges novice modelers face during the design phase is the identification of relationships between the entities involved. Novice data modelers find the setting of relationships between entities as their main challenge, mostly when non-binary relationships are involved (Batra, 1994).

DOI: 10.4018/IJICTE.2020040104

However, existing research is mainly focused on theoretical rather than empirical aspects of data modeling. That is, there has been little exploration of empirical data gathered from novice data modelers. Such empirical findings might shed light on the causes of the difficulties novice data modelers encounter during the design phase, and help instructors to improve their practice.

The aim of this study is to explore the difficulties novice data modelers encounter as novice data modelers regarding relationships between three entities. For this purpose, students who had completed a database course were asked to fill out a questionnaire including two problems dealing with various requirements, necessitating that their solutions use both binary and ternary relationships.

The research questions derived with the above aim are:

1. What are the types of error relating to the use of ternary relationships?
2. What are the underlying reasons for these errors?

THEORETICAL BACKGROUND

In this section, we present a brief theoretical survey of data modeling complexity, ternary relationships and students' difficulties in data modeling.

Data Modeling Complexity

Building a data model for an information system is a complex task, especially for novice data modelers (Topi & Ramesh, 2002). Novices encounter difficulties mainly in modeling relationships between entities (Batra & Antony, 1994). The main source of errors in data modeling by novices is attributed to cognitive complexity (Batra, 2007). Cognitive complexity, in the context of data modeling, refers to all the factors that make it difficult for one to grasp and understand all aspects of the problem at hand. These difficulties depend on the problem's structure as well as on the knowledge and previous experience of the designer. Four major sources of cognitive complexity were identified: problem solving principles, design principles, information overload, and systems theory (Batra, 2007). Though these papers refer to extensive factors, we focus only on those affecting data modeling, with special focus on ternary relationships. As to problem solving principles, the ones that are relevant to our research are connectivity and time delay. The first refers to a high degree of interrelatedness between the entities involved. The latter refers to the time gap between the design of the data model and its use. Time delay results in a lack of immediate feedback, therefore only upon use can one understand the quality of the design. As to design principles, the ones that are relevant to our research are the gap between the problem space and the solution space, no flexibility for errors, and a lack of knowledge of normalization rules. The gap between the problem space and the solution space refers to the ability to prune to a minimum the relationships described in the problem by the verbal constraints. No flexibility for errors refers to the fact that only one 'right' solution exists according to the normalization rules. A lack of knowledge of normalization rules refers to the lack of experience of novice data modelers as regards to normalization rules and their consequences. As to information overload, the only factor that is relevant to our research is noise, which refers to the presence of irrelevant information (Batra, 2007). In the process of analysis and evaluation of the results obtained in the present study, we use the factors listed above.

Ternary Relationships

The entity-relationship model (ER) (Chen, 1976), is commonly used to design databases (Lenzerini & Nobili, 1990). An ER model consists of entity-sets and relationships between them, representing real-world objects and their connections. Entity-sets include attributes of the objects, and specify an identifying key. Relationships connect two or more objects. The connections stand for a relationship representing a real-world association between the objects. However, not all real-world associations