Available online at www.sciencedirect.com
SCIENCE DIRECT®
Data & Knowledge Engineering 55 (2005) 301–326
DATA & KNOWLEDGE ENGINEERING
www.elsevier.com/locate/datak

Complexity and clarity in conceptual modeling: Comparison of mandatory and optional properties

Andrew Gemino [a,*], Yair Wand [b,1]

[a] *Faculty of Business Administration, Simon Fraser University, 8888 University Drive, Burnaby, BC, Canada V5A 1S6*
[b] *Sauder School of Business, University of British Columbia, Vancouver, BC, Canada V6T 1Z2*

Received 14 December 2004; accepted 14 December 2004
Available online 10 March 2005

Abstract

Two versions of the entity-relationship model (ERM) are compared in this empirical study. One model grammar uses optional properties and the other employs mandatory properties and subtypes. The optional grammar produces apparently less complex models than the mandatory with subtypes. An ontological analysis indicates that mandatory properties may be superior to optional properties in providing clearer representations. The Cognitive Theory of Multimedia Learning is used to hypothesize superior local information provided by mandatory properties can lead to improved viewer understanding of a model. An experiment comparing the two ERM grammars is described and results confirm the use of mandatory relationships leads to improved understanding even though the model is apparently more complex. These results suggest clarity within the model may be more important than the apparent complexity of the model when a model is used for developing domain understanding.

Keywords: Conceptual modeling; Systems analysis; Ontology; Cognitive theory; The entity-relationship model

* Corresponding author. Tel.: +604 291 4991; fax: +604 291 4920.
E-mail addresses: gemino@sfu.ca (A. Gemino), yair.wand@ubc.ca (Y. Wand).
[1] Tel.: +604 822 8395; fax: +604 822 0045.

0169-023X/$ - see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.datak.2004.12.009

1. Introduction

Analysts often strike a balance between simplicity and complexity when communicating information system requirements. Too much complexity may overwhelm the audience with detail; too simple a description may hide important issues. The choice between simplicity and complexity is especially important for communication in the early phases of systems analysis. Communication tools used in these early phases are often termed *conceptual models*. These conceptual models describe "*aspects of the physical and social world around us for the purposes of understanding and communication*" Mylopulos [29, p. 2]. Kung and Solvberg [20] suggest conceptual models support *communication* between stakeholders and Rolland and Cauvet [33] indicate conceptual models serve to mediate between users' requirements and systems design.[2] Thus, it is important that conceptual models be understandable to stakeholders, analysts and designers.

Our objective in this paper is to demonstrate how the issue of simplicity versus complexity in conceptual modeling can be studied by combining theoretical considerations and empirical methods (a general rationale for this approach is presented elsewhere [15]). This approach enables us to address conceptual model quality from both the semantic and pragmatic aspects as defined in Lindland et al. [23]. To accomplish this, we have chosen to consider an intra-grammar comparison [43] of two versions of the entity-relationship model (ERM). In one version—optional properties are allowed. In the other version, all properties are mandatory, and whether a property exists or not is captured by the use of subtyping. An empirical comparison of optional and mandatory ER grammars has been reported elsewhere [6], however, not in the general context of complexity and clarity.

To predict why there might be differences, we apply ontological analysis. We then use cognitive considerations to suggest why these theoretical differences might matter when people interact with actual models. The dependent variable is the level of domain understanding developed by people viewing conceptual models. This level is measured by the ability of a participant to reason about domain problems not directly answerable from the presented material. To explain why differences in performance might exist, we use the Cognitive Theory of Multimedia Learning [25]. Our hypothesis is that individuals do not read entire conceptual models at once, as a single chunk of information. Instead, an individual focuses on localized areas in the model and builds his or her understanding by integrating local components. Based on the ontological analysis, we claim the mandatory grammar provides clearer "local" information. Hence, more cognitive resources can be devoted to connecting model components when using the mandatory grammar. This leads to the development of a more sophisticated cognitive model of the domain being represented, and a higher level of domain understanding.

[2] Specifically, Kung and Solvberg [20] identify the following uses of conceptual models: for an analyst to reason about a domain, for communication between analysts and users, for communication between analysts and designers, and for documenting system requirements for future reference. It can be argued that in practice not all these uses take place. However, it is generally accepted that conceptual models are used for reasoning and communicating about system requirements.