An Ontology for Database Design Automation

Veda C. Storey
Department of Computer Information Systems
College of Business Administration
Georgia State University
P.O. Box 4015 Atlanta, Georgia 30302-4015
E-mail: vstorey@gsu.edu
Telephone: (404)-651-3894, Fax: (404)-651-3842

Harald Ullrich
William E. Simon Graduate School of Business Administration
University of Rochester
Rochester, NY 14627

Shankar Sundaresan
316 Beam Building
The Smeal College of Business Administration
Pennsylvania State University
University Park, PA 16802
E-mail: shankar-s@psu.edu

Abstract
Although it is possible to encode a great deal of process knowledge about database design into a system, experience has shown that the contribution of a human designer extends beyond his or her knowledge of database design techniques. The next step in the evolution of automated database design tools is to incorporate knowledge and reasoning capabilities to support this higher level of participation. Doing so, requires some understanding of what different terms mean. This paper presents an ontology that can be used as a surrogate for the meaning of words in a database design system to simulate the contributions that a designer would make to a design session with a user based on the designer's general knowledge. The ontology classifies a term into one or more categories such as person, abstract good, or tradable document. It is comprised of a semantic network, a knowledge base containing information on the meaning of terms that have been classified, an expert system knowledge acquisition component, and a distance measure for assessing the distance between the meanings of terms. The ontology was tested by different types of users on a variety of problems and was shown to be quite effective.

1 This research was supported by the National Science Foundation under Research Initiation Award IRI-9209252 to the first author, a research grant from the College of Business Administration, Georgia State University, and the William E. Simon Graduate School of Business Administration, University of Rochester. The authors wish to thank Robert C. Goldstein, Roger Chiang and Deb Dey for useful discussions on this research and Shigetaka Yamakawa for his assistance in analyzing the data.
1 Introduction

Database design tools attempt to capture and represent the data items that are important to the application domain for which the database is being developed. Many of these systems have quite sophisticated rules and heuristics for developing a good design for an application. However, they typically rely on the user for all information about the application [Storey and Goldstein, 1993]. A system guides the user through the process of specifying entities, relationships, and attributes. It then applies its expertise to detect missing or conflicting data, expresses the requirements as a well-formed conceptual model, and translates this into a form suitable for implementation. Database design systems, however, cannot compare, for example, a term that a user provides with one from a system's stock of knowledge about that application domain to ascertain if they are the same. This would require an understanding, or some surrogate, for the meaning of the terms. The objective of this research, therefore, is to: develop an ontology that can be used as a surrogate for the meaning of words. The ontology can then be incorporated into other expert systems for database design as an additional module to automatically make comparisons between terms based upon some understanding of their semantics.

2 Database Design Ontology

Any knowledge-based system is based on some abstract, simplified view of the world that is called a conceptualization. An ontology is an explicit specification of such a conceptualization [Gruber, 1993]. If two (or more) parties — humans and/or computer programs — can agree on a common ontology as a specification of their shared domain of interest, the ontology can be used to support communication among them, even though they may use entirely different internal knowledge representation mechanisms. The use of ontologies is found in many different areas: the CYC project [Lenat, 1995]; semantic interoperability [Goh et al., 1994, Sciore, 1994]; conceptual database design [Bergamaschi and Sartori, 1992]; and context interchange [Madnick, 1995]. Research on Naïve Semantics which is "commonsense knowledge associated with words" [Dahlgren, 1988; Dahlgren et al., 1989], is of particular interest.

This section presents our database design ontology. The ontology classifies a term (entity name) into one or more categories such as person, abstract good or tradable document. The purpose of the ontology is to store and provide information on the meaning of terms, to acquire information regarding the meaning of new terms, and to compare two terms to determine if and how they might be related. The ontology consists of the following components:

- a semantic network describing the different categories into which terms can be classified;
- a knowledge base containing information on the meaning of terms that have already been classified;
- an expert system-based knowledge acquisition component supporting an interactive, dialog-oriented extraction of the meaning of terms from a user; and
- a distance measure for assessing the distance between the meanings of terms.