Database Interfaces: A Conceptual Framework and a Meta-Analysis on Natural Language Studies

Hock Chuan Chan  
National University of Singapore

Lai Huat Lim  
National University of Singapore

Data utilization is an important aspect of information systems. Over the last two decades, numerous experimental studies have been conducted on user performance involving database-related tasks with certain database models and/or languages. In this paper, we propose a two-dimensional conceptual framework aimed at classifying and systematically analyzing these studies, in order to provide a bigger picture facilitating systematic understanding of this body of research. The classification exercise shows that studies involving natural language interfaces did not have very consistent findings; correspondingly, we applied the meta-analytic technique to attempt to gain insight into these differences.

Data utilization has been cited as one of the top MIS issues in many surveys (Brancheau and Wetherbe, 1987; McCormick, 1991; Niederman et al., 1991; Grover and Goslar, 1993). In the late 1970’s and early 1980’s, database research focused on relational, network and hierarchical models. In particular, experimental studies on user performance largely compared relational model (and languages) with the network and hierarchical models (and languages), and revealed relative advantages of relational languages like SQL and QBE. In the field of database research, the emphasis has been increasingly shifting from the commercially prevailing relational model and SQL standard to higher level models and languages, which include the entity relationship (ER) models and languages, and more recently the object-oriented models and languages.

It is therefore timely to summarize the experimental studies conducted so far, and to propose how further studies may be designed to further enhance our understanding of these newer models. Toward this end, we classify the various systems tested based on their semantic and syntactic levels. The resulting two-dimensional framework helps to organize findings of the individual experiments. Where necessary and feasible, the findings of the primary studies are quantitatively summarized by performing meta-analysis.

The following section presents a conceptual framework, which is used for mapping past studies. The natural language studies are then compared and integrated with a meta-analysis. This is followed by the conclusion of the paper.

A Conceptual Framework

In human-computer interaction (HCI) research, it is common to classify the interaction knowledge into syntax and semantics; one example is the syntactic/semantic model of user knowledge proposed by Shneiderman (1987). Syntactic knowledge includes the grammar of a language as well as the particular functions of different keys on the keyboard, the use of buttons on the mouse, and the functions of various icons on the screen. Semantics concerns the meanings of the interaction to the user, i.e., the concepts and actions that the user is trying to understand or perform. The distinction between semantics and syntax is also evident in the three-stage process.
model proposed by Ogden (1985) for database retrieval. At the first stage, users decide the real world information needed. At the second stage, users transform the real world requirement to data model requirement. At the final stage, users arrange the requirements into the format dictated by the language. The second stage therefore has to do with semantics, while the third deals with syntax.

We advocate that, with respect to the database area, the relevant surrogate pertaining to semantics is the level of abstraction, which is commonly classified into physical, logical, and conceptual. Correspondingly, along the semantic axis, we have various database models starting from the hierarchical model to the real world model, in increasing closeness to the real world (see Figure 1). A model consists of the concepts as well as the operations that can be performed with these concepts. Progressing down the semantic axis generally means addition of computer information – at the expense of real world information. For example, from “real world” to “ER model”, much information about the real world is lost (i.e., omitted or cannot be modeled in the latter). Going from the ER model to the relational model, the relationship information is lost (i.e., not explicitly stated in the latter; in place of that, there are foreign keys). Notwithstanding the mentioned “loss”, by doing certain joins of keys and foreign keys, relationship information can be reconstructed. In the original relational model without additional constraint mechanisms, there is no trace of the relationship cardinality information (e.g., mandatory and optional participation), which exists in the ER model. Going from the relational model to the hierarchical model, much information about physical pointers is added. With the hierarchical model, users also need to keep track of the order of records during retrievals.

When it comes to syntax, what is of significant (and practical) importance is the related ease (perceived or otherwise) with which users are accorded. In this regard, the common classifications for database interfaces (or languages) are textual (also known as linear keyword or computer language), visual (also known as graphical), and natural language (NL), in increasing order of syntactic ease. Textual language refers to a computer language in words, such as SQL. Visual language refers to a two dimensional interactive direct manipulation system, such as QBE or Visual KQL (Siau et al., 1992). Natural language refers to the use of a language “natural” to the user; it means the language that the user normally uses (e.g., English, French, German, Chinese, etc.).

Based on these simple syntactic and semantic classifications, a two-dimensional framework is produced and shown in Figure 1. The framework is applicable for all the three major database tasks - data modeling, data retrieval and data update.

The combinations of model (semantics) and language (syntax) show the various possibilities. Referring to Figure 1, for hierarchical and network models, the main language implemented is textual. On the other hand, for relational model, many textual and visual languages have been imple-

![Figure 1. Conceptual Framework of Database Models and Languages](image-url)
Related Content

Relational, Object-Oriented and Object-Relational Data Models
www.igi-global.com/chapter/relational-object-oriented-object-relational/11200?camid=4v1a

RDF Keyword Search by Query Computation
Zongmin Ma, Xiaoqing Lin, Li Yan and Zhen Zhao (2018). Journal of Database Management (pp. 1-27).
www.igi-global.com/article/rdf-keyword-search-by-query-computation/227035?camid=4v1a

Adopting Open Source Development Tools in a Commercial Production Environment: Are We Locked In?
Advanced Topics in Database Research, Volume 5 (pp. 27-40).
www.igi-global.com/chapter/adopting-open-source-development-tools/4384?camid=4v1a

A Practical Perspective on Data Quality Issues
www.igi-global.com/article/practical-perspective-data-quality-issues/51192?camid=4v1a