December 1999

Empirical validation of knowledge-based systems for conceptual database design

Solomon Antony
TexasTech University

Dinesh Batra
Florida International University

Radhika Santhanam
University of Cinncinnati

Empirical validation of knowledge-based systems for conceptual database design

Solomon Antony, Texas Tech University, santony@ttu.edu
Dinesh Batra, Florida International University, batra@fiu.edu
and
Radhika Santhanam, University of Cincinnati, santhar@email.uc.edu

Abstract

Conceptual database design is a complex and difficult task for non-expert designers. However, many of the commonly committed errors can be prevented with a knowledge-based (KB) design support system. The interface to such a system can be programmed using one of two strategies: (1) restrictive strategy in which the user is forced to follow a specific problem solving path or (2) the guidance strategy in which the user is advised on possible next steps in the problem solving process. This study involves the development of two versions of a KB system – one with a restrictive interface and other with a guidance interface – and a control system that offers no KB help. In a lab experiment non-expert designers solved a difficult data modeling task using one of the three systems. Analysis of their performance indicates that the KB systems improve the users' model accuracy. However, there was no significant difference in performance between the two KB system implementations. Subjects in the restrictive interface group rated their system as easier to use than the guidance interface group users.

Introduction

Conceptual database design is a difficult problem and it cannot be fully automated. The designer has full responsibility of ensuring its accuracy (Batini et al. 1992). Although a number of knowledge-based (KB) systems have been proposed (Storey and Goldstein 1993), there have not been an attempt at systematically evaluating any of those tools. The major objective of any of the KB systems is to assist non-experts. Non-experts commit many errors and the KB systems must be able prevent some of those errors. The KB system we developed for this study differ from the earlier ones, because it has features that are specifically aimed at preventing some of the errors a non-expert designer would commit. In this study we empirically compared two implementations of the KB system with a system that does not have knowledge base.

Background

Conceptual database design is one of the most critical phases in database design, because it is the basis for logical and subsequently physical design of the database. It involves identifying objects and their inter-relationships in the application using a data model such as the Entity-Relationship (ER) model (Chen, 1976). While developing the ER model, non-experts commit a number of errors (Batra and Antony 1994). Norman (1983) suggests that analysis of the errors be used in designing better systems. Batra and Antony (1994) have analyzed the errors and recommend ways by which some of those errors can be prevented. Their recommendations, along with the rules for the construction of entities and relationships (e.g. Teorey 1990) and heuristics and rules that were analytically shown to prevent errors (Batra and Zanakis 1994) were used in building the knowledge base of our system. The system does not possess knowledge about the specific task domains (e.g. loan management system for a library), but it has knowledge about the ER modeling methodology. Using a GUI development tool for the Windows environment, we developed two implementations of the KB systems. It differed only at the interface level as explained next.

Silver (1990) proposed system restrictiveness and decisional guidance as two strategies for developing decision support systems. Those two strategies were used in developing the interface to our KB systems too. A restrictive interface offers limited number of choices at each step of the problem solving process, thus forcing the user to follow a specific sequence of steps. In our system, it is implemented by disabling some menu options, windows, command buttons and other controls in the GUI environment. We would expect such interfaces to be very helpful to novices. The guidance interface does not restrict, but it (a) advises the user on possible next step, (b) warns about potential errors, and (c) helps with routine design decision making situations. We expect this interface to prompt the user to pause and think before making the next step. Since these two are distinct interface strategies, it is desirable to know which of the two (i) leads to better performance, (ii) is easier to use and (iii) provides more satisfaction to the user. The system with the guidance interface is hereafter called Guidance system and the one with the restrictive interface is called Restrictive system.

Research questions

We expect that a system that provides external assistance in the ER modeling process would lead to better performance than the a system that does not provide any assistance. So, our first hypothesis is that "The accuracy of data models developed by the KB system users will be higher than the accuracy of data models developed by the control system users". If a system enables a user to exercise better control over his/her cognition – like our Guidance system - it can be expected that the user would perform better (Balzer et al. 1992). The Restrictive system, not being flexible, may stifle user's free exercise of problem solving skills. Hence, our second hypothesis is that "The accuracy of data models developed by the Guidance system users will be higher than accuracy of the data models developed by the Restrictive system users". According to Silver (1991), a restrictive interface would be favored over a non-restrictive interface by non-experts and it leads us to hypothesize that "The perceived ease-of-use of the Restrictive system would be higher than that of the Guidance system". Gill (1996) found that users prefer systems that give them more discretion than systems that give less discretion. So, out next hypothesis is "The satisfaction measure would be higher for the Guidance system users than that of the Restrictive system users".

Research method

Lab experimentation was the chosen method. A pre-pilot and pilot tests were conducted prior to actual data collection. Eighty nine students from an undergraduate Introduction to information systems class volunteered to participate in the study. The participation involved attending two class sessions and a lab session.

In the first 2 hour 40 minute class session, they were taught relational database concepts. In the next class session, the same instructor taught them concepts of ER modeling and he demonstrated the ER modeling methodology. The subjects also practiced ER modeling by solving a problem (4 entities 3 binary relationships). Next, in the lab session the subjects solved another ER modeling problem (4 entities, 1 binary and 1 ternary relationships). Their performance in this problem was used as the pre-test score (PTS). Then, they were randomly assigned to one of the three groups – restrictive, guidance and control – and were given the appropriate software on diskettes. After they got familiarized with the software, they solved another problem – the experimental task - using the software. The task has 5 entities, 1 ternary and 2 binary relationships. Their performance in the experimental task is the dependent variable. System type is the independent variable and pre-test score is the co-variate. There are unique solutions to both the pre-test and experimental tasks. The grading scheme assigned 25 points to entities and 75 points to relationships. In other words, relationships are three times as important as entities.

Results

The accuracy of the ER models were assessed by two independent graders (Cohen's kappa = 0.89). Since modeling relationships is more important, only those subjects who scored at least some points for modeling relationships in the pre-test task were considered in the analysis. The results of the statistical analysis are summarized in Table below.

	Average of experimental task score	N	Average of perceived Ease-of-use measure	Average of User Information Satisfaction
Control	40.25	19		
Restrictive	55.85	17	6.19	4.16
Guidance	60.54	25	5.50	4.41
Scale	0 – 100		1 – 7	1 – 5
Statistical test	Analysis of covariance		One tailed *t*-test	One tailed *t*-test
Result	$F = 3.91$, $p < 0.05$		$T = 1.748$, $p < 0.05$	$t = 1.410$, $p < 0.10$

The results indicate significant difference between the three groups. The pair-wise comparison of Control Vs. KB systems ($t = 2.798$, $p < 0.05$), Control Vs. Guidance ($t = 2.914$, $p < 0.05$) and Control Vs. Restrictive ($t = 1.901$, $p < 0.10$) indicate statistically significant differences. However, the pair-wise comparison between the Restrictive and Guidance systems did not indicate a significant difference ($t = 0.574$, $p = 0.570$). The statistical tests on perceived ease of use and user satisfaction tests reveal significant differences too.

Discussion

The finding that the KB system users outperformed the control system users validates the knowledge embedded in the systems. An analysis of errors committed indicate that subjects in KB systems group committed fewer errors than the control group subjects. We could not find a definite answer to which type of interface – the restrictive or the guidance - leads to better performance. The finding that non-expert users find more restrictive interface easier to use confirms the expectations of other researchers (e.g. Benbasat, Dexter and Marulis 1981). We also found that the subjects who scored poorly in the pre-test task performed poorly in the experimental task also - immaterial of which system they used. This finding leads us to believe that absolute novices cannot be helped with knowledge-based systems and that they need to possess at least some problem solving knowledge before the systems can effectively help them. It has been demonstrated that concepts of system restrictiveness and decisional guidance can be implemented in design support tools. This demonstration contributes to the Information Systems field towards the important goal of attaining disciplinary status.

References

Batini, C., Ceri, S., and Navathe, S.B. *Conceptual database design: An Entity-Relationship approach.* Benjamin Cummings. Redwood City, CA. 1992.

Balzer, W.K., Sulsky, L.M., Hammer, L.B, and Sumner, K.E.. "Task information, cognitive information, or functional validity information: Which components of cognitive feedback affects performance?," O*rganizational behavior and human decision processes*, (53), 1992, 35-54.

Benbasat, I., Dexter, A.S. and Marulis, P.S. "An experimental study of the human computer interface,". *Communications of the ACM*, (24), 1981,. 752.

Batra, D. and Antony, S.R.. "Novice errors in conceptual database design," *European Journal of Information Systems*, (3), 1994, 57-69.

Batra, D. and Zanakis, S.H. "A conceptual database design approach based on rules and heuristics," *European Journal of Information Systems*, (3), 1994, 228-239.

Chen, P.P. "The Entity-Relationship model - Toward unified view of data," *ACM Transactions on Database Systems*, (1), 1976, 9-36.

Cohen, J. "A coefficient of agreement for nominal scales," *Educational and Psychological Measurement*, (1), 1960, 37-46.

Norman, D.A. Design rules based on analysis of human error, *Communications of the ACM*, (26), 1983, 254-258.

Silver, M.S. "Decision support systems: Directed and non-directed change," *Information Systems Research*, (1), 1990, 47-70.

Silver, M.S. *Systems that support decision makers: Description and analysis*. John Wiley & Sons, Chichester, England, 1991.

Storey, V.C. and Goldstein, R.C. "Knowledge-based approaches to database design," *MIS Quarterly*, (17), 1993, 25-46.

Teorey, T.J. *Database modeling and design, Morgan Kaufmann Publishers*, San Mateo, California. 1990.