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.

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 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-
mented. The most well known are probably SQL for textual and QBE for visual. The combination of relational model and natural language refers to a special NL interface where users must specify concepts and operations of the relational model. This is different from the commonly understood NL database interfaces. For ER model, many textual and visual languages have been proposed and implemented, although few have been commercialized. Although some of these languages are implemented through translations to SQL, the user sees interfaces essentially different from SQL. In other words, the underlying implementations are not important for the definition of the interfaces. The combination of ER and NL refers to the special case where users must specify ER concepts and operations using a natural language.

At the real world level, there is a combination of real world and natural language. This represents an ideal which designers of many NL interfaces are trying to achieve. A typical NL interface with a relational DBMS falls short of this ideal. With such lesser systems, a user and the system may have to go through stages of trying to understand the other. For example, whenever the database system cannot understand a user query, it will ask for clarifications (or worse, it may simply ignore the unknown terms or refuse to continue). “Learning” systems may store these clarifications to provide more links from the system’s world of relations to the real world concepts. Even after extensive use and learning, these systems are still far from the ideal.

**Framework-Based Review of Experiments**

Many experiments have been carried out which compared between some of the combinations of database models and languages discussed above. Typically, subjects were pretrained for the respective combinations being tested, and the studies involved various types of database tasks, including retrieval, update, and modeling. User performance was measured in terms of time taken to complete tasks, confidence, and accuracy. The arrows in Figure 1 represent the experimental comparisons that have been made to-date. It is important to examine the past experiments to determine the progress in the field as well as get indications for what future experiments may be needed.

**Effects of Different Semantic Levels**

The effects of different semantic levels on data retrieval, update and modeling performance of users have been investigated in many experiments, such as (Batra et al., 1990; Batra and Kirs, 1993; Chan et al., 1993; Chan et al., 1994a; Jarvenpaa and Machesky, 1986; Jih et al., 1989). Jarvenpaa and Machesky (1986) compared the modeling performance for the relational model and the ER model for end users. Better performance with the ER model was found. Batra et al. (1990) compared the relational and EER models, with the modeling task further classified based on the various facets of the EER model, such as unary, binary and ternary relationships. The results showed, in general, better user performance for the EER model.

Jih et al. (1989) compared the relational and entity relationship model on end users’ query performance. The experiment required both relational and ER model users to use the same query language SQL. The results did not show much significant difference between the two groups of users. Subsequently, Chan et al. (1993) conducted a similar experiment, but with a customized ER language for the ER model users. The relational model users again used SQL. The results showed significant performance advantages for the ER model users. The difference in this experiment is attributed to the use of a language designed for the ER model, rather than forcing ER users to use a relational language. This conclusion was collaboratively published in Leitheiser (1988). Leitheiser reported that representations using the semantic concepts were faster to learn, compared to representation using storage concepts. The semantic concept users had higher performance for tasks without languages. However, the advantages of the semantic concepts were lost when SQL was learned and used.

Besides studies on the ER model, the newer object-oriented models have also been examined. Wu et al. (1994) reported an experiment on object-oriented model plus object-oriented language against relational model plus SQL. The results showed better query performance for the object-oriented model users, compared to the relational model users. In an experiment comparing the object-oriented model and the ER model, subjects were tested on the modeling task (Hardgrave and Dalal, 1995). Subjects in the object-oriented model group used Object Modeling Technique (OMT), while subjects in the ER model group used Extended ER model. The results indicated that the only difference between the two models was in the time taken to understand. Degree of understanding and perceived ease of use did not show any difference. Another study by Bock and Ryan (1993) comparing the Extended ER model with Kroenke’s (1991) version of the object-oriented model did find some significant differences. The EER model had better modeling performance for attribute / property identifier, unary one-to-one relationship and binary many-to-many relationship. On a different task of comprehending models, the study by Shoval and Frumermann (1994) showed mixed performances for object-oriented and EER models. Subjects using different models were better at different facets of the model. Incidentally, the unclear advantages between the ER and the object-oriented model provided empirical support for placing the ER model and the object-oriented model at the same semantic level in figure 1.

Besides studies using textual keyword languages, a study using visual languages was made by Siu et al. (1995), again comparing the relational and ER models. The results showed better performance for the ER model users.

The results have been fairly consistent across all tasks. In general, the use of a model of a higher semantic level leads to
better performance.

**Effects of Different Syntactic Levels**

Studies on the syntactic level, particularly with respect to the natural language interfaces, have been much less consistent. Shneiderman (1980) counted the valid and invalid queries made by SQL and NL users. While there was no difference in the number of valid queries, NL users made significantly more invalid queries. This was attributed to overall high expectations of the NL users. Vassiliou et al. (1983) studied subjects using SQL and a natural language interface, USL. The correctness difference, in favor of SQL users, was not statistically significant. A similar study by Small and Weldon (1983) also did not find significant accuracy difference between SQL and NL subjects. Jarke et al. (1985) compared SQL with a natural language interface for eight subjects. Accuracy and time were measured. The results showed better performance for SQL subjects. A follow-up study on USL by Turner et al. (1984) found no significant accuracy difference between USL and SQL users. Jenkins and Suh (1992) compared SQL with a natural language interface (Clout) for the relational model, for querying task. Subjects’ queries were entered into the actual systems. The results showed that NL users performed better than SQL users. This is different from the earlier studies. The authors attributed the difference to the shorter semantic distance faced by NL users.

The seemingly contradicting results of NL query studies make the studies a suitable area for meta-analysis, which helps to integrate the results. To maximize consistency in comparison as well as cater to realistically reasonable sample size for the analysis, only those experiments comparing natural language against textual keyword languages are selected.

Other studies have compared languages based on procedural and non-procedural classification. For example, the experiments by Welty and Stemple (1981) compared SQL (procedural) with TABLET (non-procedural). TABLET may be classified as an early form of visual language. Another visual language, QBE, had also been compared with SQL (Greenblatt and Waxman, 1978; Yen and Scamell, 1993). The results tend to favor the visual languages.

**An Exploratory, Integrative Analysis Using the Meta-Analytic Technique**

Meta-analysis is a useful addition to the usual qualitative reviews, as the qualitative reviews “may suffer a very considerable loss of power relative to meta-analytic methods” (Rosenthal, 1984, p18).

**Information Sources**

Several methods were adopted to find the required studies. Extensive search was performed using on-line databases and the Internet. Direct correspondence with some researchers was also attempted. The major types of documents reviewed include books, journals, conference proceedings and dissertations. The sample mainly comprises studies published between the late 70’s and mid 90’s. Generally, three exclusion criteria were employed. First, non-experimental studies were omitted. Second, studies were excluded if they did not address the independent and dependent variables of interest to this research. Third, studies that lack sufficient information (in the form of some numerical data) necessary for calculating effect sizes were also omitted.

**The Meta-Analytic Technique**

Meta-analysis refers to a set of procedures for the quantitative accumulation and analysis of descriptive statistics across primary studies, whose findings are treated as the unit of analysis. Different approaches toward meta-analysis exist, e.g., Glass et al., 1981; Rosenthal, 1984; Hunter et al., 1982. The methods of meta-analysis adopted here are those presented by Rosenthal (1984), Rosenthal and Rosnow (1991) and Hunter and Schmidt (1990). A brief summary of the statistics is presented below for analyzing the two-group (natural language versus textual keyword language) experiments.

Effect size measures the degree of impact of the relationship of interest (between an independent variable and a dependent variable). Effect size can be calculated as d or r. The first choice d is calculated from the means and standard deviations of the groups in an experiment: \( d = (m_1 - m_2) / s \), where \( m_1 \) and \( m_2 \) are the dependent variable means of two experimental groups being compared, and s is the within-group standard deviation. Pearson’s correlation coefficient r can be calculated from d as follows: \( r = d / \sqrt{d^2 + 4} \). A common method to compare or combine effect sizes is to first convert the r or d to Fisher’s \( z_r \) which is equal to \( 0.5 \ln \left( \frac{1+r}{1-r} \right) \).

To compare if the p levels of individual experiments are significantly different, the following steps are needed. The first step is to find the standard normal deviate Z for each p level. Different signs for Z show effects in opposite directions. The second step calculates the value \( \Sigma \left( Z_i \right) \), where \( Z_i \) is the Z for individual experiments. This value is distributed as \( \chi^2 \) with \( K-1 \) df, where \( K \) is the number of experiments being compared. The corresponding \( \Sigma p \) value can be found. A value of \(<0.05\) indicates that the \( p \) levels in the individual experiments are significantly different (Rosenthal, 1984, p77).

To compare the effect sizes for significant differences, the value \( \Sigma (N_i - 3)(z_i - \text{mean } z_i)^2 \) has a \( \chi^2 \) distribution with \( K-1 \) df, where \( K \) is the number of experiments being compared. The \( p \) level corresponding to this value can be calculated. A value of \(<0.05\) indicates that the effect sizes are significantly different (Rosenthal, 1984, p84). The mean \( z_i \) is a weighted mean, equal to \( \frac{\Sigma (N_i - 3)z_i}{\Sigma (N_i - 3)} \). \( N_i \) is the number of sampling units in the experiment.

Heterogeneity of the effect sizes may indicate that moderating variables exist. However, homogeneity tests are used as guidelines rather than strict rules since a non-significant
overall heterogeneity test does not mean that no significant contrast can be extracted from the effect sizes (Gastil, 1994). The heterogeneity may be caused by differences in the subjects, methods used and other experimental conditions adopted by different researchers.

If moderating variables are suspected to exist, the potential moderating variable is focused tested to confirm its existence (Hunter et al., 1982). In the focused test, theoretically derived weights are assigned to values from individual experiments. For p level significance testing, $\Sigma(\lambda_i \cdot Z_i)/\sqrt{\Sigma(\lambda_i^2)}$ has a Z distribution. The weights $\lambda_i$ are chosen such that the sum $\Sigma(\lambda_i)$ is 0. For effect size focused testing, the value $\Sigma(\lambda_i Z_i)/\sqrt{\Sigma(\lambda_i^2/(N-3))}$ is distributed as Z. Corresponding p levels of <0.05 indicate significance for the theoretically derived weight and the suspected moderating variable.

Besides comparison, meta-analysis can also be used to combine the results of individual experiments into overall p and effect size values. The combined overall p value is calculated from $\Sigma(w_i Z_i)/\sqrt{\Sigma(w_i^2)}$ which is distributed as Z, and $w_i$ are weights assigned to studies. These weights may be theoretically derived, or based on the number of sampling units. The combined overall effect size is calculated from mean Fisher’s $z_i = \Sigma(w_i Z_i)/\Sigma w_i$.

**Results**

As mentioned in Section 3, the meta-analysis performed involves the comparison between natural language (NL) interfaces and textual keyword languages (KL). The columns in Table 1 and 2 show the source of the study, the number of subjects (N) in the study, the significance of the result (p) and the effect size (r), for timing and accuracy respectively. These values are either reported in the original studies, or derived from other reported values.

If p-value is all that is given in the study, the table of normal distribution is used to find the Z associated with the reported p-value. When results are reported as significant or non-significant with no actual statistics, the conservative p-values of 0.05 and 0.50 (Z=0.00) are used respectively (Rosenthal, 1984). Whenever information regarding the derivation of the effect size is not provided in a primary study, the effect size was estimated through a variety of approximation-conversion techniques (Glass et al.; 1981; Rosenthal, 1984; Strube, 1985). The approximation formulas used include these: $t=0.5 \cdot d \cdot \sqrt{df}$, and $r=\sqrt{\frac{F(1,-)/(F(1,-)+df \text{ error})}{N}}$ (Rosenthal, 1984, p21 and 25). For more details on the approximations, readers are referred to the cited publications.

For example, in Table 1, effect sizes for Vassiliou et al. (1983), Turner et al. (1984), Jarke et al. (1985) and Suh and Jenkins (1992) were calculated from the reported means and standard deviations. Those for Small and Weldon (1983) and Chan et al. (1994) were derived from the reported F value. That for Shneiderman (1980) is estimated at a conservative 0, with estimated p value of 0.5, since that study reported non-significant difference without reporting other values suitable for derivation. For Table 2, effect sizes are derived from reported means and standard deviations for Suh and Jenkins (1992) and from reported F value for the other two studies.

**Query Accuracy**

Accuracy is probably the most important measure for user performance with query languages. The results for accuracy are summarized in Table 1. The combined figures (p<0.01, r=0.18) show that users of NL are more accurate than users of KL. These figures are derived by using weights based on the number of subjects in the individual experiments.

As indicated in Table 1, homogeneity tests for both the p values and the effect sizes of the experiments show that these are significantly different (p<0.01) across experiments. This heterogeneity prompted the need to look into moderating variable effects.

Previous research reviews have posited that several variables could moderate the effects of query language type on user productivity. Unfortunately, many of these hypothesized moderators and their effects were not reported in a number of studies, thus making it impossible to analyze them. Some examples are age, examination grades, gender, and personality of subjects. After careful consideration of all the different environments present in these studies, the most important difference across experiments appears to be the difference in the natural language system. The keyword language is basically SQL, which has not changed much. Other differences do not appear to have any clear pattern. The individual differences of subjects do not have any detectable trend. The different queries used in the experiments also do not show any trend, such as in terms of complexity. Similarly, grading of query accuracies does not have any clear pattern.

Over the years, the natural language system used were

<table>
<thead>
<tr>
<th>Experiments</th>
<th>N</th>
<th>p</th>
<th>r</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shneiderman (1980)</td>
<td>22</td>
<td>0.50</td>
<td>0.00</td>
</tr>
<tr>
<td>Small &amp; Weldon (1983)</td>
<td>20</td>
<td>0.33</td>
<td>-0.23</td>
</tr>
<tr>
<td>Vassiliou et al. (1983)</td>
<td>51</td>
<td>0.30</td>
<td>-0.11</td>
</tr>
<tr>
<td>Turner et al. (1984)</td>
<td>8</td>
<td>0.47</td>
<td>-0.06</td>
</tr>
<tr>
<td>Jarke et al. (1985)</td>
<td>8</td>
<td>0.01</td>
<td>-0.99</td>
</tr>
<tr>
<td>Suh &amp; Jenkins (1992)</td>
<td>60</td>
<td>0.01</td>
<td>0.66</td>
</tr>
<tr>
<td>Chan et al. (1994b)</td>
<td>102</td>
<td>0.01</td>
<td>0.31</td>
</tr>
<tr>
<td>Test for Homogeneity</td>
<td></td>
<td>&lt;.01</td>
<td>&lt;.01</td>
</tr>
<tr>
<td>Weighted Combined Values</td>
<td></td>
<td>.01</td>
<td>.18</td>
</tr>
</tbody>
</table>

Note: N=number of subjects, p=probability of significant difference, r=effect size. A positive r means that KL is less accurate than NL.
getting better and better, as a result of research on natural language implementations. Improvements were made to both functionality and usability. We called this moderating variable **improvement in design**. As later experiments use later NL systems, and as NL systems are improved through the years, the year of publication was used as surrogate for improvement in design. This is a “next” best measure. The best measure of actually accessing the systems and evaluating them is not available.

The contrast weight chosen for the focused test starts from 0 in 1986. Studies before 1986 has negative weights, e.g., Jarke et al. (1985) has a weight of -1, and Shneiderman (1980) has a weight of -6. Studies after 1986 has a positive weight, e.g., Suh and Jenkins (1992) has a weight of 6.5 and Chan et al. (1994) has a weight of 8.5. These were increased slightly to keep the sum of weights to zero. The focused test reveals an increasing trend of the effect size ($Z=3.57$, $p<.01$). The test shows that the $p$ values of individual experiments tend to get more significant (in favor of NL) as the year of publication increases. This finding supports the hypothesis that improvement in design is a moderating variable. This is confirmed by a focused test on effect sizes, using the same weights, showing the same results.

The difference in performance over the years is attributed to the different levels of system sophistication. While the keyword languages have been almost invariant, the natural language interfaces vary from one experiment to the next; natural language implementations tend to improve with time. In spite of such increasing improvement, it ought to be noted that implementations of the model are but approximations of the real world.

In many systems, natural language simply means some synonyms for the relational tables and columns, plus some accepted terms for predefined joins. Indeed, the gap between the implemented model and the real world has been cited in early studies as the cause of much difficulty experienced by users of natural language; users thought that the systems “knows” the real world, and had overly high expectations from the systems. First, NL (e.g., English) frequently has multiple or ambiguous meanings; examples include fuzzy words like “similar” and “almost”. It also permits partial specification, with the listener filling in the missing information based on his or her understanding of the context of the statement (Malhotra and Wladawsky, 1975). Furthermore, NL tends to encourage users to make requests beyond the language, data, or knowledge boundaries of the system (Shneiderman 1980). Users, because of imperfect knowledge about the scope of the system, tend to make references beyond the system domain; this results in the return of either no answer or an incorrect response. As the gap is narrowed by better designed NL systems, users experience less frustration. This further supports the hypothesis that improvement in design is the moderating variable.

We offer two possible reasons based on the conceptual framework in Figure 1 to account for the relative advantage of NL. A first explanation has to do with the familiarity factor; natural language, being more familiar to users, does not require extra effort on syntax formulation. With the exception of Chan et al. (1994b), which compared NL with KL at the same semantic level, the experiments have used NL systems with implicit models semantically closer to the real world, compared with KL at the relational model level. Therefore, the observed advantage of NL interfaces can possibly be attributed to the associated shorter semantic distance, which requires less cognitive effort.

### Query Formulation Time

A less important measure of user performance of query language is time taken to write the queries. Fewer experiments measured this. Consequently, Table 2 showing timing results has fewer rows compared to table 1. Overall, the combined effect size ($r=0.23$) shows that **NL query can be formulated faster than KL query**.

Homogeneity tests for both the p-values and the effect sizes show that results of the individual experimental studies are statistically heterogeneous. This is similar to the findings for accuracy. The same moderating variable is applied. As the number of experiments is different, a different set of weights based on the year of publication is applied. The weights are -7, 2.5 and 4.5, adding to a sum of zero. The focused test ($Z=3.47$, $p<0.01$) supports the hypothesis that improvement in systems is a moderating variable, as for the case of accuracy.

### Implications and Limitations

The meta-analysis only addresses the syntactic dimension of the framework, and involves one type of comparison—that between natural language interfaces and textual keyword languages. The sample size is admittedly small. However, the meta-analyses provided some consistent interpretation of the otherwise apparently contradictory results of individual experiments. The synthesis is supported with statistically analyses instead of simple hypothesizing.

The overall results show that users can write NL queries

<table>
<thead>
<tr>
<th>Study</th>
<th>N</th>
<th>p</th>
<th>$r$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Small &amp; Weldon (1983)</td>
<td>20</td>
<td>.02</td>
<td>-.49</td>
</tr>
<tr>
<td>Suh &amp; Jenkins (1992)</td>
<td>60</td>
<td>.01</td>
<td>-.44</td>
</tr>
<tr>
<td>Chan et al. (1994b)</td>
<td>102</td>
<td>.01</td>
<td>.24</td>
</tr>
<tr>
<td>Test for Homogeneity</td>
<td>&lt;.01</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Weighted Combined Values</td>
<td>&lt;.01</td>
<td></td>
<td>.23</td>
</tr>
</tbody>
</table>

Note: N - number of subjects, $p$ - probability of significant difference, $r$ - effect size.

A positive $r$ means that KL required more time than NL in formulating queries.

Table 2: Effect on Query Formulation Time Due to NL vs. KL
faster and more accurately than KL queries. The moderating variable that can help explained individual differences is improvement in system design. The surrogate measure for this is the year of publication. The results put to rest earlier fear that NL systems might not be that effective. The practical implication is that developers and users should target for more NL interfaces, to replace the KL interfaces that currently dominate database management systems.

A common criticism of meta-analysis is that apples and oranges are compared. Glass (1978) suggested that these are good things to mix if we are generalizing to fruits. Indeed, the experiments had many differences, such as NL systems, actual queries written and human subjects. On the other hand, these differences are variations of the same general classes, just as apples and oranges are different fruits. What is important is to try to ascertain if the differences are systematic enough to show clear trends in the results. That had been done for the factor of improvement in NL system design.

Concluding Remarks

There have been more than two decades of experimental studies on user-data base interfaces. Technologically, database models and languages have also undergone much changes. It is timely for a review, both in the conceptual as well as the empirical aspects, to be done. This paper has made a first attempt in this regard.

Two dimensions, semantic and syntactic, have been utilized to develop a conceptual framework to classify experimental works on database interfaces. Along the semantic dimension, experiments generally show that the use of a model of a higher semantic level leads to better performance for users. Along the syntactic dimension, our analysis suggests that natural language, on the whole, appears to fare better than textual keyword languages, both in terms of query formulation speed and accuracy. Notwithstanding the above, moderating effect involving “improvement in design” exists.

The conceptual framework provides a meaningful classification system for experimental works on database interfaces. This study has demonstrated that meta-analysis methods can be useful for integrating experiments in this research area of database interfaces. It is hoped that meta-analysis will be applied to other subsets of this rich source of experimental data, such as comparisons of visual and textual languages. Such efforts will provide researchers with deeper insights at an integrative level.

References


Malhotra A. and Wladwsky I. (1975). The Utility of Natural


Dr. H. C. Chan is a senior lecturer at the Department of Information Systems and Computer Science, National University of Singapore. He has B.A. (Hon) and M.A. degrees from the University of Cambridge, U.K., and a Ph.D. in management information systems from the University of British Columbia, Canada. He has taught courses in programming, database and information systems, and has published articles in various conferences and journals, including ICIS, HICSS, MISQ, and IJHCS.

Dr. Lai-Huat Lim is a Lecturer in the Department of Information Systems and Computer Science, National University of Singapore. He is also Head of the Information Systems Research Center. Dr. Lim received his bachelor’s degree in electrical engineering and his master’s degree in MIS, both from the National University of Singapore, and his Ph.D. from the University of British Columbia. His current research interests include negotiation support systems, group support systems, Internet commerce, as well as managerial issues pertaining to collaborative technology.
Related Content

Hierarchical Fuzzy Sets to Query Possibilistic Databases
www.irma-international.org/chapter/hierarchical-fuzzy-sets-query-possibilistic/20358/

Framework for a Geographic Districting DSS using an Intelligent Object-oriented Model
www.irma-international.org/article/framework-geographic-districting-dss-using/51180/

Mathematics of Generic Specifications for Model Management, II
www.irma-international.org/chapter/mathematics-generic-specifications-model-management/11173/

Research Note: Ontology-Based Structuring of Conceptual Data Modeling Patterns
Wim Laurier and Geert Poels (2012). *Journal of Database Management* (pp. 50-64).
www.irma-international.org/article/research-note-ontology-based-structuring/74703/

Flexible Querying of Imperfect Temporal Metadata in Spatial Data Infrastructures
www.irma-international.org/chapter/flexible-querying-imperfect-temporal-metadata/52300/