2008

Information Search Process for a Well-Structured IS Problem: The Role of IS and Application Domain Knowledge

Vijay Khatri
Indiana University - Bloomington, vkhatri@indiana.edu

Iris Vessey
The University of Queensland, i.vessey@business.uq.edu.au

Follow this and additional works at: http://aisel.aisnet.org/icis2008

Recommended Citation
http://aisel.aisnet.org/icis2008/206

This material is brought to you by the International Conference on Information Systems (ICIS) at AIS Electronic Library (AISeL). It has been accepted for inclusion in ICIS 2008 Proceedings by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact elibrary@aisnet.org.
INFORMATION SEARCH PROCESS FOR A WELL-STRUCTURED IS PROBLEM: THE ROLE OF IS AND APPLICATION DOMAIN KNOWLEDGE

Processus de Recherche d'Information pour un Problème de Système d'Information Bien Structuré:
Le Rôle de la Connaissance du SI et du Domaine d'Application

Completed Research Paper

Vijay Khatri
Indiana University
Bloomington, IN 47405, USA
vkhatri@indiana.edu

Iris Vessey
The University of Queensland
Brisbane, QLD 4072, Australia
i.vessey@business.uq.edu.au

Abstract

Prior research has shown that the effect of information systems (IS) domain knowledge and application domain knowledge on problem solving is contingent on task type (Khatri et al. 2006). We build on this study by engaging in an in-depth analysis of how both these types of knowledge influence one type of task referred to as “schema-based problem solving” task. Our theoretical analysis is based on the fact that conceptual schema understanding is a well-structured problem area and that, in such a setting, participants engage in depth-first rather than the breadth-first search that is evident in the more-frequently studied ill-structured problem areas. We used protocol analysis to explore the search process in the context of varying levels of both IS and application domain knowledge. We found that knowledge of the IS and application domains result in similar effects on the search process: both high IS knowledge and familiarity with the application domain result in deeper (more focused) search.

Keywords: Conceptual modeling, conceptual schema understanding, schema-based problem-solving tasks, cognitive fit, well-structured problems, protocol analysis

Résumé

Les recherches antérieures ont montré que l'effet de la connaissance du domaine des SI et celle du domaine d'application sur la résolution de problèmes dépend du type de tâche (Khatri et al. 2006). Nous nous basons sur cette étude en menant une analyse en profondeur sur la façon dont ces deux types de connaissance influencent un type de tâche désigné comme une tâche de « résolution de problèmes basée sur les schémas ». Nous avons constaté que la connaissance du domaine des SI et la connaissance du domaine d'application produisent des effets similaires sur le processus de recherche : une bonne connaissance du SI et la familiarité avec le domaine d'application produisent un processus de recherche plus profond (plus ciblé).

Introduction

In information systems (IS), the term “domain knowledge” has dual significance. IS domain knowledge (ISDK) provides representations, methods, techniques, and tools that form the basis for the development of application systems. Application domain knowledge (ADK), which is required to develop those systems, is used to organize and structure information related to the real-world problems in that given area of business, or application domain. IS problem solving therefore applies theoretical IS knowledge to the application domain under investigation. Hence, knowledge of the IS and the application domains have a symbiotic relationship in solving IS problems.
Study of the role of ADK in conjunction with the use of ISDK is not well supported, for example, in the traditional literature on domain knowledge, which is the realm of researchers in education. In a long history of inquiry, domain knowledge has been studied in such diverse areas as physics and economics, on the one hand, and history and reading, on the other hand. Heavily based in cognitive psychology, domain knowledge in this arena is defined as knowledge of the area to which a set of theoretical concepts is applied (Alexander 1992). As is evident in this definition, a domain is viewed as a single “area” of inquiry; that is, domains are viewed as encompassing both content and principles, while in IS problem solving the principles are found in the IS domain and the content in the application domain.

The role of the application domain in IS development has long been acknowledged, for example, in conceptual papers that argue that application domain knowledge impacts IS problem-solving effectiveness (see, for example, Blum 1989; Glass and Vessey 1992). Evaluative research in the field of IS development has, however, investigated the role of the IS domain almost to the total exclusion of the application domain. Just a few studies have addressed the role of application domain knowledge either directly or indirectly (see, for example, Burton-Jones and Weber 1999; Khatri et al. 2006; Parsons 2002). Note, however, that a recent study in the area of systems development formally examined IS problem solving performance using a dual-task model (Shaft and Vessey 2006). This model, which views problem solving as addressing tasks in both domains together with a third that results in the solution, uses the theory of cognitive fit as the theoretical foundation for each of the three models representing the interlinked tasks.

The role of the application domain in conceptual modeling has received considerably less attention than in systems development. A conceptual model is a formalism that is used to develop an abstract representation of the structure of data (Hoffer et al. 2005), typically known as a conceptual schema. In a recent study, Khatri et al. (2006) developed and tested theory to explain the role of ADK, as well as ISDK, in conceptual schema understanding: the role of ADK is contingent upon the type of understanding task under investigation. Specifically, ADK aids in solving schema-based problem-solving tasks, but not in solving tasks that require only syntactic and semantic knowledge (syntactic and semantic comprehension tasks). It is now important to deepen our understanding of the role played by both ISDK and ADK in those situations in which ADK affects problem solving, that is, on schema-based problem-solving tasks.

While Khatri et al. (2006) examined the average effects of ISDK and ADK on schema-based problem-solving tasks, the objective of the current study is to determine how ISDK and ADK influence problem solving by examining individual problem-solving processes. Specifically, we examined the characteristics of participants’ search processes. In characterizing conceptual schema understanding as a well-structured problem area and characterizing search in well-structured problems as depth-first rather than breadth-first, as is typical on ill-structured problems, our research makes a substantive theoretical contribution to the literature in cognitive psychology, as well as to that in IS. We conduct an empirical study on verbal protocol data produced while solving a well-structured problem to address our overall research question: “How do ISDK and ADK influence the process of solving schema-based problem-solving tasks?”

Next, we present a theoretical conceptualization of conceptual schema understanding tasks and present theory on the nature of problem solving involved in conceptual schema understanding. In the following section, we present the theory underlying our examination of schema-based problem solving. We then present the experimental methodology we used to conduct protocol analysis, followed by our findings. Finally, we present the implications of our findings.

### Theoretical development

We first present the theoretical foundations for conceptual schema understanding, followed by those for our examination of schema-based problem solving.

#### Theory on Conceptual Schema Understanding Tasks

Most of the more recent research on conceptual schema understanding is based on read-to-do tasks (Wand and Weber 2002), which require problem solvers to respond to questions that are based on conceptual schemas. There are two basic types of such tasks, comprehension tasks and problem-solving tasks, each of which can be divided into two sub-types (Khatri et al. 2006). Comprehension tasks may be either syntactic (see, for example, Kim and March
Surface-level understanding can be explained theoretically using the theory of cognitive fit (Khatri et al. 2006; Vessey 1991). Because the knowledge required for task solution is represented directly in the schema, cognitive fit exists and the problem solver can extract the required material directly from the schema. Hence the theory of cognitive fit suggests that ADK plays no role in the solution of either syntactic or semantic comprehension tasks.

Problem-solving tasks, which have been used more recently in conceptual schema understanding research, require deeper understanding than comprehension tasks. Such tasks may be either schema-based or inferential in nature. Schema-based problem-solving tasks can also be solved using knowledge represented in the schema (Khatri et al. 2006; Shanks et al. 2003). Inferential problem-solving tasks, on the other hand, require participants to incorporate information in a conceptual schema into an existing mental model (Bodart et al. 2001; Burton-Jones and Weber 1999; Gemino 1999; Gemino and Wand 2003; Shanks et al. 2003). These types of tasks therefore evaluate participants’ broader knowledge of the domain of application (Gemino and Wand 2003; Parsons and Cole 2005) and the conceptual schema itself has little influence on performance, as shown in a number of studies (Bodart et al. 2001; Gemino 1999; Parsons and Cole 2005). We limit our examination to schema-based problem-solving tasks, which depends on ADK, as well as ISDK.

In schema-based problem-solving tasks, although all the information required to solve the problem is available in the schema, it is not available directly, and therefore cognitive fit does not exist. Conceptual modelers can solve this type of problem by transforming the knowledge in the schema into a form suitable for task solution. In this instance, however, the lack of cognitive fit, and therefore the required transformations, render the solution of such tasks more demanding. A recent experimental study confirmed that the effect of ADK on conceptual schema understanding is, indeed, contingent upon the type of task under investigation, viz., it has no influence on syntactic and semantic schema comprehension tasks, yet it does aid in the solution of schema-based problem-solving tasks (Khatri et al. 2006).

Our current research builds on this prior research. Specifically, we examine how both IS and application domain knowledge influence the solution of schema-based problem-solving tasks.

Theory on Problem Solving in Conceptual Schema Understanding

In examining how ISDK and ADK influence the solution of schema-based problem-solving tasks, we draw on the cognitive psychology literature on problem-solving. This research is based in the paradigm of humans as information processing systems (HIPS) in which information (the stimulus) is viewed as entering the mind, being processed in a series of ordered stages in short-term memory, with further information being retrieved from long-term memory as appropriate, and with the results of cognitive processing being output or stored in long-term memory (Newell and Simon 1972). The primary foci of this literature are the knowledge structures and the cognitive processes and/or strategies that act on them, and their role in expertise.

In this section, we present the theoretical basis for using the nature of search during IS problem solving to examine the influence of both types of knowledge in schema-based problem solving. We first motivate the role of search as a key factor in problem solving. We then characterize schema-based conceptual schema understanding as a well-structured problem rather than an ill-structured problem. Finally, we present propositions for solving our well-structured schema-based problem-solving tasks that reflect the way in which search characteristics differ in well- and ill-structured problem areas.

In the context of conceptual schema understanding, we refer to “expertise” as the level of knowledge in the IS domain because it is the IS domain that contains the formalized knowledge and therefore the principles related to schema-based conceptual schema understanding. We refer to the level of knowledge in the application domain in terms of being familiar or unfamiliar: in familiar domains, ADK is high; in unfamiliar domains, ADK is low.

The Role of Search in Problem Solving

The cognitive psychology community has examined long and in depth problem-solving processes, an examination that is often framed in terms of expertise. Studies of experts and novices in a variety of domains (for example, chess, medical diagnosis, musical performance, programming, and software domains), have revealed a remarkably
similar set of problem-solving skills. Below are some of the well-recognized characteristics of expert problem solving.

First, experts can solve routine problems quickly and efficiently. They use more efficient thinking strategies (Mayer 1997) and more structured problem-solving processes (Chi et al. 1981), and have a range of problem-solving heuristics to draw upon. Experts have also been characterized as using strong problem-solving strategies (that is, specifically adapted to the task at hand), while novices use weak strategies (that is, generally useful and therefore likely to be less effective and efficient than specific strategies; for example means-ends analysis) (Mayer 1992).

Second, experts tend to engage in breadth-first rather than depth-first problem-solving strategies to ensure they do not close constraints on their problem solving before establishing that they can reach a viable solution (see, for example, Adelson and Soloway 1985; Greeno 1978; Jeffries et al. 1981; Rist 1989). When they do encounter an aspect of the problem that they do not know how to solve, they investigate it using a depth-first approach, and then return to their breadth-first approach once satisfied.

Third, experts engage in top-down decomposition, that is, they use a divide and conquer strategy, to simplify cognitive processing. Experts produce good decompositions, while novices frequently make inadequate efforts to compose the problem area or need to start over because their decompositions often do not result in viable solutions (Jeffries et al. 1981).

Fourth, experts recognize when a solution approach is not working and are willing to discard it and seek out a new approach (Feltovich et al. 1997). Hence expert problem solving is characterized by flexibility in choice of problem-solving strategy.

Fifth, experts actively search for evidence to help them understand the underlying structure of the problem, whereas novices simply focus on whatever information is most readily available (Biggs et al. 1988; Bouwman 1984).

Closer examination of the above characteristics of expert problem solving reveals that they are rooted largely in the differences in “search processes” used by experts and novices. And, in fact, researchers in cognitive science typically conceive of the processing strategy used to solve a problem in terms of search. Further, it is generally acknowledged within the HIPS paradigm that the process by which intelligence arises is heuristic search (Newell and Simon 1976).

Problem solvers use search heuristics to solve problems due to their limited cognitive resources. Use of search heuristics results in more effective use of attention and memory (Frederiksen and Breuleux 1990) than would an exhaustive approach to solving the problem. Simon (1955) refers to this phenomenon as satisficing; that is, while heuristics, or “rules of thumb” are generally recognized as resulting in solutions that are sufficiently good, they do not necessarily lead to the best solutions (Leake 2002).

An examination of search strategies in schema-based problem solving is particularly appropriate because there are few or no cues to navigation in graphical models. Search strategies therefore provide valuable insights into the process of visual processing (Hungerford et al. 2004). Hence search is an effective way to examine the effects of solving problems using diagrammatic conceptual schemas.

**Problem Solving in Well-Structured Problem Areas**

Here we differentiate between well- and ill-structured problems, and characterize schema-based conceptual schema understanding as a well-structured problem area. We then present propositions related to the search processes used in schema-based problem-solving tasks in the context of different levels of IS and application domain knowledge.

A key distinguishing factor in problem solving is the nature of the problem, viz., whether the problem under investigation is well- or ill-structured in nature (see Reitman 1964).

Well-structured problems are those that have a well-defined initial state, a clearly-defined goal state, a well-defined, constrained set of transformation functions to guide the solution process, well-defined evaluation processes, and a single optimal solution path (Greeno 1978; Sinnott 1989; Voss and Post 1988). Further, the information needed to solve the problem is contained in the problem statement. On the other hand, ill-structured problems are those for which the initial and goal states are vaguely defined or unclear (Voss and Post 1988), and for which there are multiple solutions and solution paths, or no solution at all (Kitchner 1983). Further, the problem statement does not contain all of the information needed for their solution; hence it is not clear what actions are required to solve them.

Conceptual schemas represent the structure of data. Significant efforts over the past four decades have resulted in extensive formalization (see, among others, Chen 1976; Elmasri and Navathe 2006). As a result, and as stated above, all of the information required to solve schema-based conceptual schema understanding tasks is available in the schema itself. In solving these types of tasks, therefore, there is a clearly-defined initial state, a well-defined goal state, a formal set of transformation and evaluation paths, as well as a well-defined solution path. Hence schema-based conceptual schema understanding can be characterized as a well-structured problem area. Further, because ISDK alone is essential to solving such well-structured problems, any effect of ADK occurs in addition to that of ISDK; that is, there is no interaction between the two types of knowledge. Hence we can consider the distinct effects of each.

The description of expertise presented earlier begs the question of the types of problems that were examined in building that knowledge. Perusal of our references to expertise reveals that past research has focused on ill-structured problems. Hence, in presenting propositions related to schema understanding, we need to identify differences in solving problems in well- and ill-structured problem domains.

A major difference is that search in well-structured problem solving is characterized by depth-first rather than the breadth-first search characteristic of ill-structured problem areas. Problem solvers on ill-structured problems need to engage in breadth-first search to determine how they can solve the complex problem with which they are confronted. For well-structured problems, problem solution relies on the structure inherent in the associated representations, in this case, the conceptual schema. Hence expertise will be reflected in the use of depth-first problem-solving processes that are quite focused in nature. We refer to such search as “deeper,” “more focused.” In general, then, expertise in solving well-structured problems is characterized by focused usage of well-structured search.

Having established that problem solving in well-structured problem areas is characterized by depth-first search, we now address the influence of different levels of ISDK (high and low) and ADK (familiar and unfamiliar) on search in our schema-based problem-solving tasks.

**Influence of IS Domain Knowledge.** We first address the effect of an individual’s ISDK in solving schema-based problem-solving tasks. Typically, experts tend to use their knowledge of the problem domain (that is, the IS domain) to guide their search for data to understand the structure of the task at hand, while novices are heavily influenced by the surface features of the task, simply focusing on whatever information is most readily available (Biggs et al. 1988; Bouwman 1984), and probably therefore engaging in less focused problem solving (Chi et al. 1981).

With regard to our well-structured schema-based problem-solving tasks, we therefore expect that our high ISDK participants (H-ISDK) will engage in deeper, more focused search than low ISDK participants (L-ISDK). Hence we state the following proposition.

**Proposition 1:** Problem solvers with better ISDK engage in deeper, more focused search than those with lesser ISDK, irrespective of their level of ADK.

**Influence of Application Domain Knowledge.** We now address the effect of familiarity with the application domain in IS problem solving. Because problem solvers perform better on schema-based problem-solving tasks in familiar domains (Khatri et al. 2006), we expect that they will exhibit more focused search in familiar than unfamiliar application domains. Hence we state the following proposition.

**Proposition 2:** Problem solvers in familiar application domains engage in deeper, more focused search than those in unfamiliar domains, irrespective of their level of ISDK.

**Research Methodology**

We used the exploratory technique of protocol analysis to examine schema-based problem solving in familiar and unfamiliar application domains.
**Task Setting**

We investigated sales and hydrology as our two application domains. Our expectations were that participants drawn from a business school (see section entitled “Experimental Design”) would be more familiar with a sales application than with a hydrology application.

We investigated the solution of schema-based problem-solving tasks on the conceptual models most commonly used in practice (Davies et al. 2006): the ER and EER models (see, Chen 1976, and Elmasri and Navathe 2006, respectively).

**Participants**

Study participants were 12 undergraduate students, proficient in conceptual modeling, who were drawn from two sections of a data management course offered in the business school of a large university in the U.S. mid-west. Participation in the study was voluntary and the subjects were given $30 to complete the conceptual schema understanding experiment. All of the participants were between 20 and 25 years old, and had a high-school diploma, some work experience, and little database-related work experience. We conceived of our student participants as novice conceptual modelers.

**Experimental Design**

We used a 2 x 2 mixed design with knowledge of the IS domain as the between-subjects factor and familiarity with the application domain as the within-subjects factor. Participants’ “expertise” in conceptual modeling (ISDK) was assessed based on their response to syntactic and semantic comprehension questions. Participants demonstrating high and low ISDK each completed four schema-based problem-solving tasks, two tasks in each of the familiar and unfamiliar application domains, which we refer to as Task 1 and Task 2. Participants were randomly assigned to two groups (ER and EER). Because the schema-based problem-solving tasks investigated in this research involved only entity types/relationships and attributes (ERA), that is, concepts that are common across ER and EER models, we did not investigate further the effects of the formalism. Further, the presentation sequence of the two schemas (familiar and unfamiliar) was counterbalanced, thereby effectively controlling for any order effects.

**Investigating the Influence of IS Domain Knowledge**

To investigate the influence of ISDK on the solution of schema-based problem-solving tasks, we formed groups of participants with high and low expertise in the IS domain. To do so, we examined participants’ scores on syntactic and semantic comprehension tasks. As noted earlier, while scores on the syntactic task assessed the extent to which the participants understood the syntax of the conceptual model, scores on the semantic task assessed the extent to which the participants understood the meaning of the associated constructs on the conceptual schema. Performance on syntactic and semantic comprehension questions is an appropriate measure of IS expertise because it is well established in the cognitive psychology literature that knowledge of surface features, or declarative knowledge, forms the foundation for developing higher forms of knowledge such as procedural knowledge (Anderson 1996). We used this performance measure to form groups of participants with varying levels of knowledge of ERA. We then selected the six highest performers to form the H-ISDK group (participants referred to as H-x), and the six lowest performers to form the L-ISDK group (L-x).

**Investigating the Influence of Application Domain Knowledge**

Our experimental design called for the use of two domains with which our participants would not be equally familiar. We refer to these application domains as familiar and unfamiliar. As a manipulation check on ADK prior to the experiment proper, each subject was asked to describe five hydrology terms (seep, playa, bore hole, lithology and pump) followed by five sales terms (product line, sales person, warehouse, area headquarter and manufacturer), terms that mapped to concepts on the conceptual schemas with which they later interacted. Hence this exercise highlighted for the participants what they knew about aspects of each domain. The participants were then asked to rate their familiarity, on a 7-point scale, with sales and hydrology applications (where 7 = high and 1 = low
familiarity). The self-reported familiarity of all the subjects was far greater in the sales than in the hydrology domain.

**Experimental Materials**

**Conceptual Schemas**

Two schemas (along with the corresponding data dictionaries) were presented to each participant, one in the familiar domain (sales) and the other in the unfamiliar domain (hydrology); the data dictionary included application-oriented descriptions of each entity type/relationship and attribute on the schema. The schemas were syntactically equivalent; only the labels used for entity types, relationships, and attributes differed. The schema and the data dictionary were adapted from Khatri et al. (2006).

The sales schema was a typical order-processing application that included concepts such as SALES AREA, SALES TERRITORY, PRODUCT, PRODUCT LINE, and MANAGER. The hydrology schema was adapted from a schema for a groundwater application at the U.S. Geological Survey. This application included hydrological concepts such as SEEP, PLAYA, BORE HOLE, CASING, and ACCESS TUBE.

**Schema-Based Problem-Solving Tasks**

Our participants responded to two schema-based problem-solving tasks in each of the sales and hydrology domains. As noted above, both tasks focused on ERA only. The tasks were structurally equivalent in each domain; that is, structurally-corresponding entity types and attributes were needed to respond to the corresponding task in each application domain.

The two tasks are as follows.

**Task 1** (Sales): Managers in the finance and marketing divisions need to decide which products to keep in their product portfolio. These decisions are based on measures of advertising budget, miscellaneous expenditure, and target audience for a given product line. Based on the material provided, can you find an answer to the above problem? If yes, describe how you would find the answer. Please be specific.

**Task 1** (Hydrology): Geologists and hydrologists need to decide which bore holes to include in their groundwater study. These decisions are based on measures of leakance, horizontal conductivity, and vertical conductivity at a bore hole site. Based on the material provided, can you find an answer to the above problem? If yes, describe how you would find the answer. Please be specific.

**Task 2** (Sales): A group of customer service managers needs to understand the defects associated with recently delivered products. We need to find the address and CEO of a manufacturer that produced products in the last five years. Based on the material provided, can you find an answer to the above problem? If yes, describe how you would find the answer. Please be specific.

**Task 2** (Hydrology): A group of earth scientists needs to understand the rock formations associated with recently constructed bore holes. We need to find the age and formation name of a lithology that is related to bore holes constructed in the last five years. Based on the material provided, can you find an answer to the above problem? If yes, describe how you would find the answer. Please be specific.

**Task Analysis**

We review our tasks to provide a basis for understanding the types of analyses we conducted and the comparisons we made.

**Task 1**

In the familiar application domain, the solution of Task 1 requires making a decision on retaining products in a PRODUCT LINE according to a set of criteria. The criteria are presented as attributes of PRODUCT LINE on the
schema. Any analysis of further entity types other than PRODUCT LINE is unnecessary. In the unfamiliar application domain, a similar situation arises, this time for BOREHOLE SITE.

Task 2

In the familiar application domain, the solution of Task 2 requires finding the CEO and address of a MANUFACTURER that produced PRODUCTS in the last five years. A response to this task requires participants to determine the history of the PRODUCTS that are associated with MANUFACTURER. However, the schema does not present the date on which the PRODUCTS were produced. It is therefore not possible to answer this question. Any analysis of further entity types other than PRODUCT and MANUFACTURER is unnecessary. In the unfamiliar application domain, a similar situation arises, this time for LITHOLOGY associated with BOREHOLE history.

Pilot Study and Experimental Procedure

We conducted a pilot study with graduate students who had conceptual modeling experience. The pilot study helped us to eliminate ambiguity in question wording, test the experimental procedures, and determine the length of time that the experiment would take to complete. The pilot data were used to develop the coding scheme for the experiment proper.

In the study itself, the researchers used a common script to conduct 12 individual sessions. Following an introduction to the study, participants completed a background questionnaire that sought demographic information and measured their a priori familiarity with the sales and hydrology application domains. Participants were then given an information sheet that described the syntax of the assigned model and viewed a PowerPoint video, developed using Camtasia Studio, that recapped key conceptual modeling concepts. Next, participants were given instructions on the verbal protocol (think-aloud) technique used to collect experimental data and completed a practice exercise to become familiar with thinking aloud while problem solving. Finally, participants completed two sets of experimental exercises, one in each of the familiar and unfamiliar application domains. Participants responded in sequence to syntactic and semantic comprehension tasks, and the schema-based problem solving task. Responding first to syntactic and semantic tasks allowed participants to become conversant with the given schema. Further, they responded to Task 1 prior to Task 2. All responses were audio-taped and later transcribed.

Data Analysis

Verbal protocol data is collected when participants speak aloud while carrying out specific tasks; that is, they verbalize the “inner voice” of problem solving. This verbal data must then be converted into a form suitable for analysis.

We first coded characteristics of the search process based on references to both entities/relationships and attributes, in keeping with the characteristics of ER diagrams. We examined search processes by capturing the transitions participants made among different objects in solving the schema-based problem-solving tasks under investigation, where objects are entity types/attributes/relationships, the problem statement, and the data dictionary.

Representing Search Processes

To gain insight into problem-solving processes, we used transition graphs, which displayed the problem-solving processes selected participants used during problem-solving. Transition graphs illustrate references to different aspects of the experimental material as a sequence of coded utterances over time. Transition graphs facilitate seeing at a glance the extent to which the problem solver exhibited a “focused” problem-solving process.

In these graphs, each indicator represents a participant’s verbalization of the problem statement (✱), reference to the data dictionary (▲), an E/Rs (⊙), or an attribute (●). The numbers in the attribute symbol refer to the attribute referenced by the subject; for example, in Figure 1-1a, the participant referred to an attribute in PRODUCT LINE, denoted by “1,” twice, as shown by the appearance of a black circle containing the figure “1” on two occasions.

Note that the transition graphs also present a pictorial representation of the degree of focus in the search process. The focused nature of the search for E/Rs is evidenced in the number of E/Rs referenced on the vertical axis: the
fewer the references to different E/Rs, more focused the search. A focused search for attributes is reflected in the number of black circles with different numbers on the horizontal axis: more black circles with different numbers illustrates less focused search within the E/Rs referenced.

Findings

In keeping with the tenets of protocol analysis, we examine the problem-solving processes of individual participants. We first use transition graphs to illustrate the search processes of four participants, one each from the H- and L-ISDK groups for each of Tasks 1 and 2. We then conduct a formal analysis of the search processes used by each of our 12 participants.

Recall that transition graphs are visual representations of the flow of a participant’s search processes. The participants whose search processes we selected for display are those centered immediately above and below the median of each of the high and low ISDK groups. We examine the processes of those participants below the median (H-4 and L-4) on Task 1, and those of participants above the median (H-3 and L-3) on Task 2. Figure 1 shows, from the left, the transition graphs for each of the four participants on Tasks 1 and 2, respectively. The number of transitions is reflected in the vertical lines linking objects in a given transition graph. The higher the number of transitions above those needed to solve the problem (one in Task 1, three in Task 2), the greater the cognitive stress the participant experienced.

We examine, first, the search processes for participants H-4 and L-4 on Task 1. Visual scanning of the four transition graphs to the left of Figure 1 shows, first, that the graph of L-4 in the unfamiliar domain appears to be quite different from the other three graphs. While H-4’s search behavior changes little from the familiar to the unfamiliar application domain, L-4 exhibited more focused search behavior in the familiar (F-AD) than the unfamiliar application domain (U-AD); L-4, for example, examined a number of E/Rs unnecessarily. Second, while the search processes of these two participants are similar at the E/R level in the familiar application domain, L-4 shows greater uncertainty by referencing the same attributes on a number of occasions. It appears, therefore, that while the overall problem-solving processes of H-4 and L-4 follow similar patterns across application domains, L-4 does reveal some stress in addressing Task 1 in both domains, most notably in the unfamiliar domain.

Next we turn to the transition graphs for participants H-3 and L-3 on the more complex task, Task 2. On this task, reference needed to be made to two entity types, as opposed to one in Task 1. First, visual scanning of the four transition graphs shows immediately an increased level of activity over that observed for Task 1 for all graphs except that for H-3 in the familiar application domain. Second, L-3 in both application domains and H-3 in the unfamiliar domain referenced many more E/Rs than necessary to solve the problem. Third, H-3’s search process involved 6 transitions in the familiar application domain and 13 in the unfamiliar domain, while the corresponding figures for L-3 were 9 and 21, respectively. Hence, the cognitive stress evidenced by L-3 in both domains, but especially in the unfamiliar application domain, is readily apparent.

Fourth, participant H-3’s search process for ISDK in the familiar application domain is reasonably similar to that of H-4 on Task 1, given that reference needed to be made to two entity types. H-4 did reveal some uncertainty, however, in trying to make intuitive sense of the question by reading it several times, and then employing a deep search for attributes by referencing the same attributes on a number of occasions, as did L-4 in the familiar domain on Task 1. On the other hand, in the unfamiliar domain, H-3 searched the schema extensively, focusing widely on E/Rs, and then on the attributes within the E/Rs, although not particularly deeply. Participant L-3 searched very widely for E/Rs, but also quite deeply for attributes in the familiar application domain.

Fifth, in the unfamiliar application domain both the low and high ISDK participants referenced the problem statement a number of times throughout their solution attempt. Given that participants did not feel such a need in the familiar application domain, this evidence of uncertainty in their problem-solving processes probably reflects lack of understanding of the application domain. L-3 also referred to the data dictionary a number of times in a search for ADK when addressing the task in the unfamiliar domain. It appears, therefore, that only H-3 was comfortable addressing Task 2, as evidenced in the use of a search process that was as focused as that of H-4 on Task 1. The search process used by L-3 in the familiar application domain was both broader and deeper. Both participants in the unfamiliar application domain used search processes that were both broad and shallow.
In summary, then, Figure 1 shows that H-ISDK participants (H-3 and H-4) exhibited much more focused problem-solving processes (that is, they made fewer transitions) than L-ISDK participants (L-3 and L-4), in both the familiar and the unfamiliar domains on both Tasks 1 and 2.

We now examine more formally the search processes of each of the 12 participants. Table 1 presents an analysis of the transitions undertaken by all participants; for example, participant H-4, an H-ISDK subject, undertook 3 transitions in F-AD in Task 1. From the viewpoint of AD familiarity (columns “F-AD” and “U-AD”), we observe that out of 24 comparisons between “F-AD” and “U-AD” (for Tasks 1 and 2), 20 comparisons suggest that participants engaged in fewer transitions in the familiar application domain compared to the unfamiliar application domain suggesting that participants engage in more focused search when the AD is familiar. From the viewpoint of ISDK, Table 1 shows that the median number of transitions by L-ISDK participants is more than that of H-ISDK participants, suggesting that knowledge of the IS domain results in more focused search on a conceptual schema. Additionally, the summary analysis of Table 1, shown in Tables 2a and b, shows that median ADK and ISDK ratios (median U-AD/median F-AD and median L-ISDK/median H-ISDK, respectively) are greater than 1, i.e., median participant transitions are more focused: 1) in F-AD than in U-AD; and 2) for H-ISDK than for L-ISDK participants. Therefore, the transition analysis illustrated in Figure 1 reflects the more in-depth analyses presented in Tables 1, 2a, and 2b.

<table>
<thead>
<tr>
<th>Subject</th>
<th>F-AD</th>
<th>U-AD</th>
<th>F-AD</th>
<th>U-AD</th>
</tr>
</thead>
<tbody>
<tr>
<td>H-1</td>
<td>1</td>
<td>3</td>
<td>7</td>
<td>11</td>
</tr>
<tr>
<td>H-2</td>
<td>4</td>
<td>7</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>H-3</td>
<td>1</td>
<td>1</td>
<td>6</td>
<td>13</td>
</tr>
<tr>
<td>H-4</td>
<td>3</td>
<td>5</td>
<td>1</td>
<td>6</td>
</tr>
<tr>
<td>H-5</td>
<td>9</td>
<td>3</td>
<td>4</td>
<td>7</td>
</tr>
<tr>
<td>H-6</td>
<td>1</td>
<td>3</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td><strong>Median</strong></td>
<td><strong>2</strong></td>
<td><strong>3</strong></td>
<td><strong>4</strong></td>
<td><strong>6.5</strong></td>
</tr>
</tbody>
</table>

L-1 3 6 9 11
L-2 1 4 6 9
L-3 3 4 9 21
L-4 5 7 5 4
L-5 1 4 8 9
L-6 7 1 4 5
**Median** 3 4 7 9

### Discussion

This research focuses on the role of both IS and application domain knowledge (ISDK and ADK, respectively) in conceptual schema understanding. More recent research has shown that schema-based problem-solving tasks are facilitated by both ISDK and ADK. In this exploratory research, we sought to build on that prior research by determining how both ISDK and ADK influence the solution of schema-based problem-solving tasks. Our overall research question is: “How do ISDK and ADK influence the process of solving schema-based problem-solving tasks?”

We characterized conceptual schema understanding as a well-structured problem area, and examined differences in problem solving on these types of tasks compared with the ill-structured problems that have typically been studied. To provide insight into the potential effects, we used protocol analysis to examine the search behavior on schema-based problem-solving tasks of problem solvers possessing both high and low knowledge of the IS domain in both familiar and unfamiliar application domains. This section discusses our findings and presents the contributions of our research.

Because of the central role that search plays in problem solving, we examined the effects of search in solving well-structured schema-based problem-solving tasks. Specifically, we examined the effects of both ISDK and ADK in IS problem solving based on the inherent structure of conceptual schemas; that is, we examined search for E/Rs and attributes.

Our findings support our hypotheses that both IS and application domain knowledge have a positive influence on problem-solving behavior. Specifically, search during problem solving is more focused, and therefore deeper, when
both IS and application domain knowledge are high. Further, search is less focused as the complexity of the task increases.

Our research makes a number of contributions to the literature. First, well-structured problem solving appears to have been understudied in cognitive psychology research. Characterizing conceptual schema understanding as a well-structured problem area opens the way for the examination and characterization of problem solving on well-structured problems, in general.

Second, a key finding of this study is the fact that expert problem solving in well-structured problem areas is characterized by focused, depth-first search rather than the breadth-first search typical of ill-structured problem areas. Our research therefore makes a contribution to the cognitive psychology literature on solving well-structured problems.

Third, we observed that problem solvers in well-structured problem areas rarely seek knowledge of the application domain explicitly; that is, there were few direct references to the data dictionary. Instead, knowledge of the application domain was manifested through search for ISDK where search was deeper and more focused in the familiar application domain than in the unfamiliar.

Fourth, we observed that certain of our participants exhibited signs of limitations in cognitive capacity (that is, stress) that, on certain occasions, resulted in less focused (broader and shallower) search and a somewhat haphazard problem-solving process. This observation may not be surprising in view of the fact that cognitive limitations are likely to come into play when cognitive fit does not exist, as in these types of tasks.

The existence of stress, which is particularly salient in our visual transition graphs, which depict participant’s search processes, appears to be further exacerbated when ISDK is low, the domain is unfamiliar, and the task is more complex, as in Task 2. This situation is illustrated by the transition graph of the L-ISDK participant on Task 2 in the unfamiliar application domain. While all problem solvers with the exception of the L-ISDK participant in the unfamiliar domain exhibited focused problem-solving processes on Task 1, only the H-ISDK participant in the familiar domain exhibited focused search processes on Task 2.

Fifth, this study presents a way of evaluating search behavior in conceptual modeling. While search behavior has been widely studied in cognitive psychology (see, for example, Adelson and Soloway 1985; Johnson et al. 1981), our work shows how analysis of search processes can be applied to conceptual schema comprehension.

Our study has the following limitations. First, we conducted our investigation using students who were relatively inexperienced in using real world conceptual schemas. We characterize them as novice conceptual modelers. Second, in order to stretch our problem solvers we had them respond to two different problems. Task 2, which could not be solved, was more complex than Task 1. It is possible that using a second task that could have been solved may have led to different findings. Third, rather than assessing performance on schema-based problem-solving tasks directly, we sought a richer characterization through an experimental design that manipulated the levels of both IS and application domain knowledge.

**Implications for Research**

Our findings have a number of implications for researchers. First, this research focused only on ERA in ER and EER models. Further research needs to be conducted in the context of other conceptual models such as the class diagrams of UML. We note, again, however, that the ER model is by far the most popular data model in practice (Davies et al. 2006). Second, prior research has found differences in performance on conceptual schema understanding tasks resulting from semantic ambiguities of ontologically-unsound representations (see for example, Burton-Jones and Meso 2006; Burton-Jones and Weber 1999). Protocol analysis therefore needs to be conducted in the context of ontologically sound and unsound representations.

Third, while this study has extended our knowledge of schema understanding, further protocol analysis studies need to be conducted in the context of schema development. Srinivasan and Te’eni (1995), for example, found that specific strategies for building a conceptual schema affect the quality of the resulting representation. Their research needs to be extended to include the roles of both ISDK and ADK in conceptual schema development.
Implications for Practice

Our research also has several implications for practice. First, our study presents insight into the search process used in solving well-structured problems represented diagrammatically. Hence, our study addresses a key in our information rich society – the ability to reason with diagrams. Our research therefore has implications for managers in fields such as accounting, marketing, decision sciences, and information systems.

Second, this study provides further evidence for organizations providing training for data analysts. Because lack of both types of domain knowledge results in use of information that is not relevant to the task at hand, training should focus not only on tool knowledge and ISDK, but also on ADK.

Third, the growing body of evidence pointing to the importance of ADK in certain types of IS problem solving suggests that model tool builders should investigate ways to incorporate characteristics of the application domain into their tools. The role of a tool is to help reduce time and effort, for example, through the use of domain-specific modeling patterns and templates.

Fourth, this study provides initial insights into the processes underlying application and technical risks in application development (Blum 1989). While the former refers to what is known about the problem, the latter refers to confidence that an implementation can be produced that will satisfy the stated requirements. Analysts who engage in broader, unfocused search might be exposing themselves to application and technical risks. Application risk may be mitigated by prototypes/simulations and incremental development, while technical risks may be managed by better training of designers/analysts in using software tools.

Conclusions

The role of the application domain is an issue that has been largely neglected in research on conceptual modeling. In this study, we addressed how both IS domain knowledge (ISDK) and application domain knowledge (ADK) influenced the solution of schema-based problem-solving tasks by examining search behavior of novice conceptual modelers in a study based theoretically on the fact that conceptual schema understanding is a well-structured problem area. Our research shows that existence of both ISDK and ADK each results in deeper, more focused search. Conversely, unfamiliarity with the application domain (low ADK) and low ISDK result in broader and therefore less focused, and shallower search.

Our study contributes to the growing recognition and study of the role of application domain knowledge in conceptual schema understanding, providing guidance for training students and practitioners, for example, by recommending the development of effective search behavior, building tools to support the conceptual schema understanding process, as well as aiding the management of application and technical risks.

References


Figure 1: Transition Graphs