

Investigating the Anomalous States of Knowledge Hypothesis in a Real-Life Problem Situation: A Study of History and Psychology Undergraduates Seeking Information for a Course Essay

Charles Cole, John Leide, Jamshid Beheshti, and Andrew Large

Graduate School of Library and Information Studies, McGill University, 3459 McTavish Street, Montreal, Quebec, Canada H3A 1Y1. E-mail: {charles.cole, john.leide, jamshid.beheshti, andrew.large}@mcgill.ca

Martin Brooks

Institute for Information Technology, National Research Council of Canada, Montreal Road, Building M-50, Ottawa, Ontario, Canada K1A 0R6. E-mail: martin.brooks@nrc.ca

The authors present a study of the real-life information needs of 59 McGill University undergraduates researching essay topics for either a history or psychology course, interviewed just after they had selected their essay topic. The interview's purpose was to transform the undergraduate's query from general topic terms, based on vague conceptions of their essay topic, to an information need-based query. To chart the transformation, the authors investigate N. J. Belkin, R. N. Oddy, and H. M. Brooks' Anomalous States of Knowledge (ASK) hypothesis (1982a, 1982b), which links the user's ASK to a relevant document set via a common code based on structural facets. In the present study an interoperable structural code based on eight essay styles is created, then notions of structural facets compatible with a high-impact essay structure are presented. The important findings of the study are: (a) the undergraduates' topic statements and terms derived from it do not constitute an effective information need statement because for most of the subjects in the study the topic terms conformed to a low-impact essay style; (b) essay style is an effective interoperable structural code for charting the evolution of the undergraduate's knowledge state from ASK to partial resolution of the ASK in an information need statement.

Introduction

Information Need: Start State, Goal State

When a person accesses an information retrieval (IR) system, the person with a known item information need knows

Received May 21, 2004; revised September 17, 2004; accepted October 14, 2004

© 2005 Wiley Periodicals, Inc. • Published online 30 September 2005 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/asi.20248

what information is needed, making the IR system a simple document locating or accessing device. However, in cases where the user first begins solving a problem or performing a task, the user does not know what information he or she needs from the system, indicating an unknown item information need.

Since Cutter (1876/1904), known item and unknown item information needs have been recognized as fundamentally different needs (cf. Cottrell & Eisenberg, 2001; Svenonius, 2000). Known item information need is an effective user start state because it corresponds to an immediate goal of the user, enabling the user to immediately effectuate task performance when the information is found. Unknown item need, on the other hand, is not an effective start state; it places the user in the position of having to predict what information is needed based on vague conceptions of his or her information need.

For unknown item information need, Taylor's (1968) classic article on information need describes the user's actual information need as being unconscious (Level 1). A modified form of the need is vaguely conscious but nevertheless unspecifiable (Level 2), but even the specifiable form of the need (Level 3) is not the one used by the person when accessing information (i.e., asking a librarian or IR system a question). The user accesses information from a librarian or system (Level 4) employing compromised language or word tokens as the query to the IR system because the user thinks the information need will not be processed by the IR system unless it has been "appropriately" coded.

Following the initial studies of Oddy (1975, 1977) in the mid-1970s, Belkin, Oddy, and Brooks examined Taylor's Level 1 or Level 2 conception of information need and began to incorporate it into designing an information need-based IR system (Belkin, 1980, p. 137). Even in cases where

information need is “a statement of what the user does not know” (Belkin, Oddy, & Brooks, 1982a, p. 64), nonetheless the need’s nonspecificity could be incorporated into IR system theory and design (Belkin, 1980, p. 139). If such an IR system could be designed, they believed, the system would facilitate unknown item information need satisfaction without requiring the user to either know or specify that need to the system. Belkin, Oddy, and Brooks’ (1982a, 1982b) two articles on the user’s incomplete state of knowledge producing an information need—anomalous states of knowledge or ASK—are the most highly cited articles ever published in the *Journal of Documentation* (Inside ASIS&T, 2004).

Here we investigate Belkin, Oddy, and Brooks’ ASK hypothesis for IR system design and apply this approach to the study of the real-life problem situation of 59 undergraduate students with an unknown item information need. The students were enrolled in history and psychology courses at McGill University in fall 2003.

The article is structured as follows. First, we present Belkin, Oddy, and Brooks’ ASK hypothesis. Then we integrate important theoretical premises from the ASK hypothesis with the specific problem situation of the study group being examined (the 59 undergraduates researching a course essay). Next, we create an interoperable structural code based on the problem situation of the 59 undergraduates. After which, we report the results of the study of 59 undergraduates and discuss the results of the study in terms of the ASK hypothesis. We conclude by specifying the two important findings of the study, as well as future research projects based on the notion of an interoperable structural code.

Belkin, Oddy, and Brooks’ ASK Hypothesis

In place of information need, Belkin, Oddy, and Brooks (1982a, 1982b) believed it would be possible to derive, analyze, and then represent the user’s anomalous state of knowledge (ASK) to the IR system. The user’s ASK—an anomaly in the user’s knowledge state about the topic or situation for which information is being sought—comes before the user’s information need, causing the latter to form. Therefore, though the user’s ASK is not the same thing as his or her information need, it is both fundamental to the information need’s formation and fundamental to the form the information need will take. Belkin, Oddy, and Brooks believed that by dealing with the user’s ASK, which was identifiable and representable, instead of the user’s information need, which was not identifiable or representable, a properly designed IR system could serve the user seeking information with an unknown information need in a more “natural manner” than current systems (Daniels, Brooks, & Belkin, 1985/1997).

For such a system, the process of deriving and resolving the user’s ASK can be divided into five stages:

1. The user’s knowledge of the topic for which information is being sought is represented, called the *knowledge representation technique*. First, the user’s knowledge of

the topic is obtained via an analysis of the user’s own description in a lengthy problem statement, and then represented in an *association map*. The association map is created by calculating lexical distances between words in all the word pairs used in the user’s problem statement. Lexical distances are given *r* ratings: $r = 1$, if A & B are adjacent words in the same sentence; $r = 2$, if A & B are words in the same sentence but not adjacent; $r = 3$, if A & B are words in adjacent sentences within the same paragraph, etc. (Belkin, Oddy, & Brooks, 1982b, p. 148).

2. The ASK is then derived from the association map via an ASK representation technique: The structural characteristics of the user’s ASK are analyzed by breaking down the association map into groups (clusters, stars, and lines); the relations among the groups are determined (cluster–cluster, cluster–star, cluster–line, star–star, star–line); and finally, the overall connectivity of the structure is determined (Belkin & Kwasnik, 1986, p. 22).

To determine overall connectivity, first, ASKs are globally categorized as either attached or detached. Belkin, Oddy, and Brooks codified this essential feature of the ASK as a facet, called the *Attachment Facet*. The Attachment Facet is followed by the determination of two other facets called the *Overall Structure Facet* and the *Structure Characteristics Facet*.

- Attachment Facet: If two or more clusters of concepts nodes are not connected with any of the others, the ASK is deemed detached.
- Overall Structure Facet: The Overall Structure is determined by the ratio of number of links in the association graph/map to the maximum number of links possible for the number of nodes present in the graph/map.
- Structure Characteristics Facet: The Structure Characteristics are concerned with local structural and lexical features of the groupings of clusters, etc. in the ASK (Belkin & Kwasnik, 1986, p. 18).

3. The ASKs are classified according to a global structure classification scheme (note that local structure characteristics of the ASK are also used to determine the ASK class; Yuan, Belkin, & Kim, 2002). The ASK’s global structure is measured using a connectivity score derived by using the connectivity formula: $C = N_a - N_{\min}$. As the number of lines in the Association Map of all ASKs stays constant at 40, the minimum number of nodes possible is $N_{\min} = 10$. For the example of a user’s ASK given in Belkin, Oddy, and Brooks (1982b), the number of nodes in the ASK network is 25, so the connectivity formula given above determines that the connectivity score for this particular ASK is 15, meaning this particular ASK is assigned to Class C in the global structure classification scheme which follows (with their range of connectivity scores in parenthesis; Belkin, Oddy, & Brooks, 1982b, p. 153):

- Class A: (0–5)
- Class B: (6–10)
- Class C: (11–15)
- Class D: (16–20)
- Class E: (21–25)

It is important to note that the classes in the ASK classification scheme are meant to be algorithmically generated (Belkin, Oddy, & Brooks, 1982b, p. 152), leading directly to system-invoked information retrieval strategies to resolve the ASK (we return to this point in the Discussion section below).

4. The ASK's class automatically invokes particular information retrieval strategies to resolve the user's ASK, depending on the connectivity score as referenced immediately above. The retrieval strategies are grouped into five basic retrieval strategies:

MATCH: Term identification procedure—Specifies an ASK structure, or area of an ASK structure, from which a list of terms is to be used for straightforward quorum searching.

TRIAD: Structural analysis—Operates on clusters in the ASK structure, specifying triplets of terms whose relationships and position in the ASK structure will be used to rank the texts.

STAR: Structural analysis—Identifies terms for matching and ranking from stars in the ASK structure.

PATH: Structural analysis—Identifies groups of terms for matching and ranking which are attached to clusters in the ASK structure, but are not parts of clusters. Group relations are retained for ranking purposes.

LEXICAL: Special closed vocabularies identified (e.g., words such as “research,” “want,” “find,” “problem,” “literature,” “search”). Closed vocabularies are pointers or noncontent words in the ASK structure, which are eliminated from searching consideration but are useful to identify specific parts of the structure to be operated upon (Belkin & Kwasnik, 1986, p. 16).

5. Once the correct retrieval strategy is invoked and the relevant document set is presented to the user, the user must read and understand the relevant documents to resolve the ASK. However, when this is done, the user's unknown and nonspecifiable information need is also satisfied, even though it is never represented in the transaction between user and IR system.

Integration of ASK Hypothesis With the 59 Undergraduates' Problem Situation

In this section, we integrate or apply the three main premises of the ASK hypothesis to the particular problem situation of the 59 history and psychology undergraduate students in our study.

Premise 1

Belkin, Oddy, and Brooks distinguish their ASK hypothesis from contemporary, early 1980s IR system design principles, principles that apply to current systems as well. For users of IR systems with an unknown item information need who are exploring a topic area, an important ASK premise is that there is “a disconnect” between the structure of the user's expression of his or her need and the structure of the relevant document set. Oddy (personal communication, Sep-

tember 7, 2004) states that “the user's initial statement of need is not an accurate description of relevant documents, because it is a different category of object.”

Belkin, Oddy, and Brooks further theorize this disconnect, referring to the different views of Wersig and Neveling (1975) and Boulding (1956) about the information spectrum involved in a user seeking information in a specific environment. Wersig and Neveling emphasized the structures of the *real-world* environment, while Boulding emphasized the structure of the *image* the user has of the real-world environment. Belkin, Oddy, and Brooks' ASK hypothesis sides itself with the user's image of information structures, while IR systems largely concern themselves with presenting real-world structures to the user (i.e., via subject cataloging, classification, or subject directory schemes; Belkin & Robertson, 1976, p. 198; cf. Belkin, Oddy, & Brooks, 1982a, p. 66; Oddy et al., 1992, p. 143).

The 59 undergraduates in our study were deliberately interviewed when they were in an exploration stage of researching their essay topic, just after they had selected their topic. According to Kuhlthau (1991, 1993), this user is now in Stage 3, the exploration stage of a 6-stage information search process (ISP). In this stage, the user must explore the topic to gain a personal understanding of the topic area they have just selected (cf. Cole & Leide, 2003; Cole, Mandelblatt, & Stevenson, 2002). Popper (1975) describes this straining for understanding as World 2, a different world from the objective product, the artifacts of the human mind in World 3. Everything is different between the two worlds. For the undergraduate grasping at understanding of the topic area, in a conceptual sense information is a process (Buckland, 1991), a process of transforming the undergraduate's knowledge structure as it slowly absorbs a sense of the topic's structure (Hjorland, 1997); information is not conceptually a “thing” as topic structure is represented in accessing tools located in books, articles, indexes, subject catalogues, classification schemes, and thesauri.

Premise 2

Users in this exploration stage query an IR system by typing in topic elements to represent their information need, but these topic elements are frequently not equivalent to, nor do they adequately represent the user's information need (Belkin, Oddy, and Brooks, 1982b, p. 160; cf. Saracevic, Kantor, Chamis, & Trivison, 1988).

For undergraduates researching a history or psychology essay, findings reported in Cole, Leide, Large, Beheshti, and Brooks (2005) tangentially corroborate Belkin, Oddy, and Brooks' premise that a query based on topic tokens is not an effective start state for information retrieval. In this study, the same 59 social science undergraduates at McGill University were interviewed, using the same interview schedule shown in the Appendix. We required they do the interview just after they had selected their essay topic, putting them in Stage 3 of the Kuhlthau's (1991, 1993) ISP model. The purpose of the interview was to take the students through a topic

TABLE 1. Subjects' preferred approaches among original topic listing (OTL), original visualization (OV), circle line visualization (CLV), four-question visualization (QV), and last term listing (LTL).

Term list or visualization type	History: 200-level	History: 400-level	Psychology: 400-level	Total
OTL	3 (16%)	1 (5.5%)	1 (5%)	5 (8%)
OV	2 (10.5%)	4 (22.2%)	3 (15%)	9 (16%)
CLV	3 (16%)	4 (22.2%)	8 (40%)	15 (26.5%)
QV	6 (31.5%)	3 (17%)	2 (10%)	11 (19%)
LTL	5 (26%)	6 (33%)	6 (30%)	17 (30%)

focusing process consisting of various visualizations, rank ordering and narrowing cognitive activities, starting from their Topic Statement and ending up with what we labeled their *Information Need Statement*.

For both the Topic Statement and the Information Need Statement, the students were required to list out the terms in those statements—terms we said they would use to access information sources. These terms thus constitute two sets of query terms, one at the beginning of the interview and one at the end. Table 1 shows that only 8% of the 59 students in the study chose their Topic Term listing—terms derived from their topic statement at the beginning of the interview—as their preferred choice compared to 30% for their Information Need Statement term listing. Moreover, all other possible choices were preferred over the Topic Term listing. This result tends to support the premise that query terms derived from topic statements are, according to the users themselves, inadequate representations of the user's real information need.

Premise 3

Belkin, Oddy, and Brooks modeled the user's ASK and the domain knowledge state using compatible structures (Belkin & Robertson, 1976, p. 203), thus creating a single, interoperable structural code which links (a) the user's ASK, (b), the set of documents in the IR system's database, and (c) the strategy for resolving the user's ASK. According to Belkin, this is the major innovation of the ASK theory: "... the base statement of the basic phenomena is nothing particularly new. But defining and relating these phenomena in terms of structure does lead to something new" (Belkin & Robertson, 1976, p. 202). The ASK interoperable structural code is made up of facets which describe the ASK and document structures, as well as rules governing the employment of the information retrieval facets to resolve the user's ASK. As shown in Figure 1, the ASK facets decide connectivity issues for the specific user by first examining the ASK structure then automatically invoking the most appropriate set of retrieval strategy facets for resolving the ASK and satisfying the information need.

We show the interoperable structural code of the ASK hypothesis in operation in the following example, for Subject 14 (s. 14) from Belkin and Kwasnik (1986). The example demonstrates the interconnectedness of the ASK structure,

operationalized into facets, and the information retrieval strategy, also operationalized into facets:

ASK facets:

- If ASK is of category x, then do y, where y is either a retrieval strategy or invoking another rule.
- For Subject 14 (s. 14; a real example from Belkin & Kwasnik, 1986):
- The first facet invoked is *attachment*. The basic rule in attachment says
- If attached,
- Then do *overall structure*.
- Since this is not a detached structure, we proceed to the facet *overall structure*. In this facet, s.14 responds to the rule:
- If two or more clusters linked at PLO,
- Then do *structural characteristics* as a lexical characteristics rule, which
- Invokes the **Information Retrieval Strategy facets**:
 - *Stage 1*:
 - LEXICAL: (finding one closed vocabulary term)
 - TRIAD: operating on type 1 and 2 clusters
 - STAR: operating on stars
 - MATCH: using terms from TRIAD and STAR
 - *Stage 2*:
 - MATCH: must have most involved node
 - TRIAD: rank in order of structure duplication and node strength
 - STAR: modify rank by inclusion of star nodes (Belkin & Kwasnik, 1986, pp. 17–19).

Figure 1 diagrams the example of the interoperable nature of the ASK and Information Retrieval Strategy facets just given. For the specific example, the three ASK facets are gone through in order, in an "if . . . then" format. When the procedure comes to the third ASK facet, the retrieval strategy facets are then invoked.

The Specific Problem

The specific problem addressed in this article is to investigate and apply Belkin, Oddy, and Brooks' premise of an interoperable structural code for undergraduates accessing information from an IR system who are exploring a topic

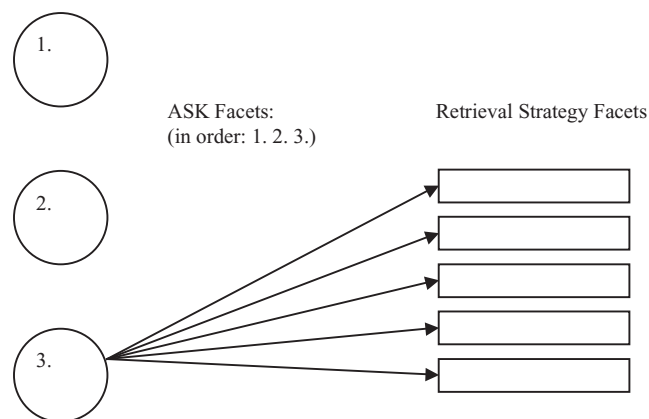


FIG. 1. ASK facets and retrieval strategy facets.

area for a history or psychology course essay and who must put their learning on display to get a high mark from their course instructor. The interoperable structural code must link the undergraduates' ASK, however it can be defined, to the most appropriate information retrieval strategy for resolving the ASK.

An Interoperable Structural Code for Undergraduates Researching an Essay

The intention of Belkin, Oddy, and Brooks was to create a top-down, across-domain IR system (Oddy, Palmquist, & Crawford, 1986, p. 249). The connectivity classes that make up the scheme are meant to be algorithmically generated (Belkin, Oddy, & Brooks, 1982b, p. 152), leading directly to information retrieval strategies to resolve the ASK. However, in Belkin, Oddy, and Brooks' Table 6 (p. 154), they create an interim step, describing the five classes in the ASK classification scheme as five different problem situation types (these problem situation descriptions were taken from Brooks' masters thesis; Brooks, 1978).

The place of the specific problem situation of the user gained greater prominence when Oddy et al. (1992) revisited ASK 10 years later. He observed that the underlying structures of the user's problem statement and the abstract of the documents created by professional indexers did not, in fact, mesh all that well, due to several reasons. One of the reasons was that users in writing a problem statement concentrate too much on topic aspects of the information need, thus ignoring other issues in their problem situation which probably have an effect on the ASK (Oddy et al., 1992, p. 145).

We share these misgivings that the user's problem statement misses something fundamental that a finer instrument tuned to the specific problem situation of the 59 undergraduates in our study would capture. Therefore, in our conceptualization of an interoperable structural code linking the undergraduate's (a) ASK with (b) the set of relevant documents in the IR system's database, and (c) an information retrieval strategy for resolving the ASK, we decided to replace (b) the set of documents, with a task-problem situation structure appropriate for undergraduates researching a course essay.

We have previously defined the undergraduate's task as putting his or her learning on display in the essay so that the course instructor will give the essay a high mark. The essay structure selected by the undergraduate will be a key variable in performing the task effectively. Our strategy is to assist the user in determining the desired essay structure; the retrieval strategy is then derived from the combination of the chosen essay structure with the essay topic.

There are eight commonly used essay styles (Howard University, 2003):

1. Narration essay: Transformation of an experience into a story. The writer seeks to make a point and to provide a detailed, descriptive, and sometimes, personal account of an event or experience.

2. Descriptive essay/narrative essay: Yields a word-portrait of a person, place, or thing. Provides factual information about a topic (e.g., an encyclopedic entry) or provides a subjective depiction of an experience (e.g., how racial prejudice affects the author or an audience).
3. Process/analysis essay: Provides a prescribed "menu." Provides instruction and information (e.g., "how to" essays, procedural writing).
4. Compare/contrast essay: Demonstrates how things, views, or issues are similar and/or different—Juxtaposition of elements.
5. Definition essay: Identifies and explains the special nature of an institution, theory, philosophy, or group of people.
6. Cause and effect essay: Engages in speculation, asks "what if" questions, and hypothesizes about the factors which might bring about an event or the consequences of certain actions.
7. Persuasion and argument essay: Presents a case and attempts to change the reader's mind. Sometimes includes a call to action.
8. Ambiguous Essay: An essay with no discernable structure.

In Figure 2, we deconstruct one of the eight essay styles, the Compare and Contrast Essay style, breaking it down into structural facets similar to the ASK facets on the left-hand side of Figure 2 (showing stars, paths, and triads). The structural elements for a Compare and Contrast Essay style are (a) the essay thesis, (b) the two events or concepts being compared and contrasted by the student (the two circles in the diagram), and (c) the three linking concepts the undergraduate wishes to use to prove or disprove his or her thesis. It is possible that these three components of a Compare and Contrast Essay style can be considered facets. If the undergraduate's ASK can be categorized as an essay style, then reduced to facets, so that a retrieval strategy based on a similar structure—a common interoperable structural coding scheme—can be invoked by the IR system to resolve the undergraduate's ASK, then we have gone part way in designing an operating IR system based on the ASK hypothesis.

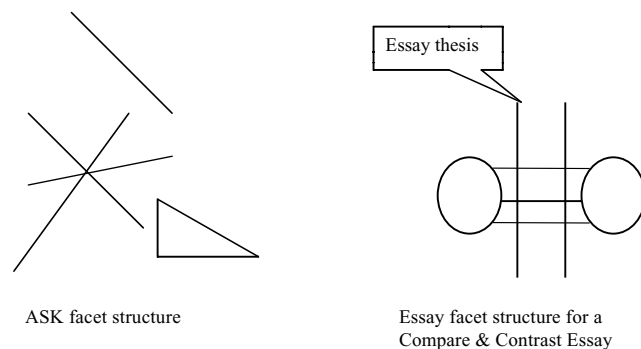


FIG. 2. On left-hand side, the facet structure of Belkin, Oddy, and Brooks' (1982a, 1982b) ASK, made up of (from left to right) stars, paths, and triads, versus the facet structure of a Compare and Contrast Essay style (on right-hand side of figure), which consists of two types: relational facets (the three lines linking the two circles) and conceptual facets (the two circles).

Research Question

The research question posed in the following study is: Does essay style have the potential to be an interoperable structural code capable of linking (a) the undergraduate's ASK, defined as topic terms from their topic statement that we categorize by essay style, to (b) the undergraduate's task–problem situation, which we define as an essay style that fulfills the task of putting the undergraduate's learning on display, via (c) an effective retrieval strategy determined by this linkage?

The Study

In the winter term, January to April, 2003, 59 undergraduate students attending social science courses at McGill University were interviewed using the interview schedule shown in the Appendix. McGill University is a large (26,000 fulltime students), 175-year-old research university offering doctoral programs in a full range of disciplines, including medicine and law. The 59 students were interviewed about their topic for a course essay, an essay that would be evaluated by the course instructor and used as a component of the total mark the student received for the course. The interviews, therefore, were about the students' real problem situation, task, and information need.

The researchers gathered subjects for the study by entering courses and asking for volunteers. The study sample, therefore, was self-selected; it does not necessarily represent the larger population of history or psychology undergraduates at McGill University or elsewhere. The volunteers were contacted by e-mail when they had selected their essay topic and had done some preliminary thinking about it (i.e., when they were in Stage 3 of Kuhlthau's (1991) information search process (ISP) model). The interview schedule required the students to list subject terms or draw visualizations of their essay topic on large sheets of newsprint-type paper. The interviews were video–audio taped and lasted from 30 minutes to 1 hour.

To distinguish domain and experience differences, students came from four different types of social science courses, representing both the humanities or “soft” end of the social science spectrum (history students) and the science or “hard” end of the social science spectrum (psychology students; cf. Whitmire, 2002), as well as both inexperienced first-year students and more experienced third-year students. The number aimed for in each cell of the 2×2 structure was 20 students.

History 200: Twenty students from one large (over 150 students) 200-level (first year) history course.

History 400: Nine and ten students, respectively, from two small (10–15 students) 400-level (third–fourth year) history courses.

Psychology 400: Twenty students from one large (over 100 students) 400-level (third–fourth year) psychology course.

We will complete a 2×2 cell structure study design by adding a Psychology 400-level course with an essay requirement in the fall term, 2003 (such a course was not given at McGill during the term when the study reported here was conducted). The purpose of the data analysis reported here is to investigate the ASK-related research question as it is defined above, not to explore subject domain differences, which we leave to a later article.

Data Analysis

The interview schedule contains 14 cognitive activities (listing, visualizing, ranking, narrowing, selecting) designed to narrow the undergraduates' topic statements about their course essay from a topic statement at the beginning of the interview to an information need focus at the end. We assessed the essay style of the undergraduate's evolving ASK at three different points in the interview schedule:

1. Topic statement and terms derived from the topic statement at the beginning of the interview schedule.
2. Four Questions and the terms derived from the question selected as most important by the subject in the middle of the interview schedule.
3. Information Need Statement and the terms derived from the Information Need Statement at the end of the interview schedule.

Results

In Tables 2–4, we list the findings by course: History 200 (Table 2), History 400 (Table 3), and Psychology 400 (Table 4). For each course, we ascertained the essay style at three different stages of the interview schedule (Appendix): the Topic stage at the beginning of the interview, the “Four Questions stage” in the middle of the interview, and the Information Need stage at the end of the interview. Table 5 summarizes the three course tables.

In Table 2, seven subjects in the History 200 course started the interview in the Ambiguous Essay style. The Ambiguous Essay style is the catchall category used when the essay style did not comply with any of the other essay styles. A further eight subjects started the interview with “Descriptive/Narrative Essay” style. Over the course of the interview most of these students switched to other essay styles. In the last stage of the interview, 15 of the 20 students in History 200 were labeled either a *Cause and Effect Essay style* or a *Persuasion and Argument Essay style*.

In Table 3, eight subjects in the History 400 course started the interview in the Descriptive/Narrative Essay style; a further five students started the interview in the Ambiguous Essay style. Over the course of the interview, the subjects shifted out of these two essay styles; almost half of the students in this course (9 of 19 subjects) ended up in the *Cause and Effect Essay style*.

In Table 4, 12 of the 20 subjects in the Psychology 400 course started the interview in the Ambiguous Essay

TABLE 2. Evolving essay styles: History 200.

Essay style	Narrative	Descriptive/ narrative essay	Process/ analysis essay	Compare/ contrast essay	Definition essay	Cause & effect essay	Persuasion & argument essay	Ambiguous essay style
Stage of interview schedule								
1. Topic		8			1		4	7
2. Four questions		3			6	2	8	1
3. Information need		1			3	6	9	1

TABLE 3. Evolving essay styles: History 400.

Essay style	Narrative	Descriptive/ narrative essay	Process/ analysis essay	Compare/ contrast essay	Definition essay	Cause & effect essay	Persuasion & argument essay	Ambiguous essay style
Stage of interview schedule								
1. Topic		8		1	2	3		5
2. Four questions		4			6	7	2	
3. Information need		3		1	4	9	2	

TABLE 4. Evolving essay styles: Psychology 400.

Essay style	Narrative	Descriptive/ narrative essay	Process/ analysis essay	Compare/ contrast essay	Definition essay	Cause & effect essay	Persuasion & argument essay	Ambiguous essay style
Stage of interview schedule								
1. Topic		2		1		5		12
2. Four questions		3				17		
3. Information need		1				19		

TABLE 5. Evolving essay styles at three separate stages in the interview schedule: Total for all subjects.

Essay style	Narrative	Descriptive/ narrative essay	Process/ analysis essay	Compare/ contrast essay	Definition essay	Cause & effect essay	Persuasion & argument essay	Ambiguous essay style
Stage of interview schedule								
1. Topic		18		2	3	8	4	24
2. Four questions		10			12	26	10	1
3. Information need		5		1	7	34	11	1

style. By the end of the interview, a dramatic shift occurred with 19 of 20 students ending up in the Cause and Effect Essay style.

Table 5 contains a summary of the results for all 59 subjects in the study, charting the evolution of essay style over

the three stages of the interview schedule. Twenty-four of the 59 students started off in the Ambiguous Essay style in the Topic statement section of the interview, then switched out of this style in the Four Questions part of the interview schedule to various other styles of essay, before finally ending

the interview in mainly the Cause and Effect Essay style (34 students) or the Persuasion and Argument Essay style (11 students).

Limitation of Research Method

A limitation of the data analysis is the inherent subjectivity of assigning essay styles. Data analysis and table construction were carried out by the research assistant (RA) for the study project—a PhD candidate at McGill's Graduate School of Library and Information Studies with 15 years previous experience working as a librarian in public libraries and other types of information centers. For the important finding, from Table 5, the two first authors disagreed with the RA for the number of Ambiguous Essay style cases as follows (by stage of interview):

1. Topic Statement, $n = 25$ (first two authors) instead of $n = 24$ (RA).
2. Four Questions, $n = 0$ instead of $n = 1$.
3. Information Need, $n = 2$ instead of $n = 1$.

Discussion

The research intervention occurred when the students were estimated to be in Stage 3 of their ISP, the exploration stage of researching their essay topic, just after they had selected their essay topic. The students were first asked to write out their topic statements and then list the terms, including terms derived from the topic statements, which they would use to look for information for their essay. The subjects mostly started out in the topic statements in either an Ambiguous Essay style ($n = 24$) or a Descriptive/Narrative Essay style ($n = 18$). The Ambiguous Essay style indicates no discernable essay style, while the Descriptive/Narrative Essay style is a simple description or narration of an event that we estimate to be a low impact essay style from the point of view of the undergraduates' task of putting their learning on display for the course instructor marking the essay.

The Four Questions stage of the interview asked the students to list four research questions they wished answered by information sources, to select the most important question for their essay, then to list terms from that question. This part of the interview schedule was designed to narrow the focus of the students' perspective on their essay to something like a thesis statement. In the Four Questions stage, the students shifted dramatically out of the so-called low impact essay styles to mainly the Cause and Effect Essay style but also to the Definition Essay style, and, for the students in the History 200 course, the Persuasion Essay style. These are much higher impact essay styles from the point of view of the students putting their learning on display so that they will get a high mark from the course instructor marking the essay. Both styles demonstrate critical thinking based on a refined essay structure.

The Information Need stage of the interview schedule is designed to encourage the students to think over the whole

interview and to come up with a definitive statement about what their essay is about, and to list terms from that statement. We have labeled this stage Information Need because the promotion of information need identification for the student is the underlying intention of the interview schedule. We are just beginning to supply proof of the contention that this final interview statement of the students in our study contains elements of the students' real information need (Cole, Leide, Large, Beheshti & Brooks, 2005). In this Information Need stage of the interview, the students in the study shifted strongly to the Cause and Effect Essay style ($n = 33$ of a total of 59) or the Persuasion Essay style ($n = 12$). These essay styles are high impact styles from the point of view of the students' problem situation/task of putting learning on display to obtain a high mark from the course instructor who will mark the essay.

The results support Belkin, Oddy, and Brooks' premise that the undergraduate's topic statement is not their information need but closer to their ASK because for most of the subjects in the study the topic statement did not conform to the style of essay that effectively put their learning on display. After the subjects in our study were taken through the cognitive activities in the interview schedule, the vast majority agreed with this deduction, as they indicated one of the later stages of the interview schedule as their preferred approach to accessing information for their essay (Table 1). As the effect of the interview was to push the subjects out of a low-impact Essay Style into a higher impact Essay Style, the interview schedule may be considered an instrument that gets undergraduates to an effective start state for accessing information from the system. If this is proven (after further testing), then the interview instrument has built in an effective retrieval strategy for resolution of the undergraduate's ASK.

The results tentatively indicate that for this particular user group it is possible to conceive of Essay Style as the basis of an interoperable structural code linking the undergraduates' ASK in the topic statement to partially resolving the ASK by facilitating an information need statement, which forms the query and start state for an appropriate information retrieval strategy.

For the Compare and Contrast Essay style, the query and start state for an appropriate information retrieval strategy based on the undergraduate's information need is embedded in the facets and how they are related to each other. As we conceive it in Figure 2, the Compare and Contrast Essay style has three relational facets (the three lines) linking the two conceptual facets the undergraduate wishes to compare and contrast (the two circles). For example, to prove an undergraduate thesis that the conditions leading up to World War I and World War II were similar and that WWII was thus a continuation of WWI, there are social, economic, and political relation issues (which will be specified by the student). Each of the three relation issues (the lines), along with both WWI and WWII concepts (the circles), forms a separate query to the IR system, for a total of three queries. The other essay styles will have different structural codes and different retrieval strategies.

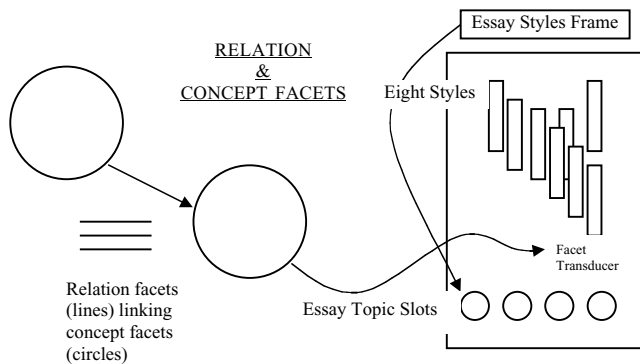


FIG. 3. Shifting essay style facets (left) to universal interoperable code based on Minsky's (1975) frame theory (right-hand side).

We have thought about how to maintain the Belkin, Oddy, and Brooks' principle of creating a top-down, cross-domain, cross-problem situation—a "one-size-fits-all" interoperable structural code—rather than the problem situation-based interoperable code we have provided in this article. In Figure 3, we illustrate the transition from problem situation (essay style) to a universal scheme based on Minsky's (1975) theory of frames. On the left-hand side of Figure 3 is the Compare and Contrast essay style, taken from Figure 2, but in a deconstructed form, with separated-out concept and relational facets represented by circles and lines respectively. The Compare and Contrast facets, which are problem specific in their configuration, are represented in Figure 3 as shifting to the universal scheme-orientation of Minsky's frame theory on the right-hand side of the figure.

In Figure 3, we imagine we have accessed an Essay Styles Frame from a complete library of frames. The upper level vertical boxes represent the eight possible types of essay style. These styles are always true. The slots (round circles) at the bottom of frame "must be filled by specific instances or data" (Minsky, 1975, p. 212), which in the case of a history or psychology essay is the data constituting the subject topic of the user's essay. There is a Facet Transducer in between as interoperable structural code. These Essay Topic slots, in turn, constitute separate frames that can be accessed from a library of topic frames. Each topic frame shows the structure of that topic (derived from a thesaurus tree or some other subject organization tool).

Conclusion

Popular search engines interfaces, as well as other information accessing interfaces, require the user to either type in key words or to specify in some other way their query to the system, thus creating a word-representation of their information need. Belkin, Oddy, and Brooks' seminal articles in the *Journal of Documentation* are important, we believe, because they firmly face the issue of unknown item information need and the nonspecificity of this sort of need. Their ASK hypothesis offers an alternative based on this premise, creating a framework for resolving the user's information need without forcing the user to specify it.

The second important premise of the ASK hypothesis is the notion that there must be a sort of transducer built into the IR system that can link the user's ASK to the relevant document set contained within the system's database. We have referred to this transducer aspect of the IR system design, based on facets, as an interoperable structural code.

Here, we have applied these two premises but changed from a top-down, cross-domain, cross-problem situation interoperable code to one based on the specific problem situation of the group being studied. We argue that many undergraduates researching a course essay approach an IR system when they are in Stage 3 of their ISP, with an unknown item information need—i.e., they do not know and are not able to specify their information need in a query to the system. To upgrade performance and user satisfaction, new interactive IR systems must go beyond their present reliance on this type of user typing in topic concepts as a representation of their information need. For undergraduates researching a course essay, we see essay styles as an effective anchor for both representing the undergraduate's ASK to the IR system and resolving the ASK via concrete information retrieval strategies based on a facet analysis of essay styles.

The two important findings from the study are:

1. The topic statements of the undergraduates in the study did not effectively represent their information need, defined as the task-problem situation of the undergraduate, with 24 of the 59 subjects indicating a low-impact Ambiguous Essay style, and 18 of 59 indicating a low-impact Descriptive/Narrative Essay style in their topic statements. The topic components that now constitute the query to an IR system for this type of user is not the undergraduate's information need but rather something closer to the undergraduate's ASK.
2. The study shows, for undergraduates researching a social science essay, that an interoperable structural code based on essay styles has potential for:
 - Linking the undergraduate's ASK to the undergraduate's task-problem situation-information need
 - Facilitating the resolution of the undergraduate's ASK and satisfying the undergraduate's information need as evidenced by the shift from Ambiguous and Descriptive/Narrative Essay styles to higher impact essay styles over the course of the interview

Undergraduates researching a social science essay is a user group that is relatively well known. Also, the task for which they seek information—the undergraduate essay—can be specified both in its structure and from a sociological perspective when the task of essay researching is considered in the wider context of their problem situation. Other user groups are more complicated to analyze. However, we can see the type of information need analysis shown here being applicable to health care information seeking, with differential diagnosis for health information seekers forming an interoperable structural coding system akin to essay styles for undergraduate students.

Our immediate future research goal is to determine which cognitive activity (see Appendix) has the most impact in the process we describe here to simplify that process. We are also analyzing the data from the point of view of domain (history vs. psychology) and experience level of the undergraduates in the study (first-year students vs. third-year students). These two subjects constitute the subject of our next article.

Acknowledgments

The first four authors wish to acknowledge the financial support of the Social Science and Humanities Research Council of Canada (SSHRC), File No. 410–2002–1394. The authors would also like to thank our research assistants for assisting us in the data analysis and the creation of the results tables.

References

Inside ASIS&T: Annual Meeting Coverage. (2004). *Bulletin of the American Society for Information Science and Technology*, 30(3), 4–11.

Belkin, M.J. (1980). Anomalous states of knowledge as a basis for information retrieval. *Canadian Journal of Information Science*, 5, 133–143.

Belkin, N.J., Oddy, R.N., & Brooks, H.M. (1982a). ASK for information retrieval: Part I. Background and theory. *Journal of Documentation*, 38(2), 61–71.

Belkin, N.J., Oddy, R.N., & Brooks, H.M. (1982b). ASK for information retrieval: Part II. Results of a design study. *Journal of Documentation*, 38(3), 145–164.

Belkin, N.J., & Kwasnik, B.H. (1986). Using structural representations of anomalous states of knowledge for choosing document retrieval strategies. In F. Rabitti (Ed.), *Proceedings of the 9th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval* (pp. 11–22). New York: ACM.

Belkin, N.J., & Robertson, S.E. (1976). Information science and the phenomenon of information. *Journal of the American Society for Information Science*, 27(4), 197–204.

Boulding, K.E. (1956). *The image*. Ann Arbor, MI: University of Michigan Press.

Brooks, H.M. (1978). *The knowledge structure underlying information needs*. Unpublished master's thesis, Centre for information Science, The City University, London.

Buckland, M.K. (1991). Information as thing. *Journal of the American Society for Information Science*, 42, 351–360.

Cole, C., & Leide, J.E. (2003). Using the user's mental model to guide the integration of information space into information need. *Journal of the American Society for Information Science and Technology*, 54(1), 39–46.

Cole, C., Leide, J.E., Large, A., Beheshti, J., & Brooks, M. (2005). Putting it together online: Information need identification for the domain novice user. *Journal of the American Society for Information Science and Technology*, 56(7), 684–695.

Cole, C., Mandelblatt, B., & Stevenson, J. (2002). Visualizing a high recall search strategy output for undergraduates in an exploration stage of researching a term paper. *Information Processing & Management*, 38, 37–54.

Cottrell, J.R., & Eisenberg, M.B. (2001). Applying an information problem-solving model to academic reference work: findings and implications. *College & Research Libraries*, 62(4), 334–347.

Cutter, C.A. (1904). *Rules for a dictionary catalog* (3rd ed.). Washington, DC: Government Printing Office. (Original work published 1867)

Daniels, P.J., Brooks, H.M., & Belkin, N.J. (1997/1985). Using problem structures for driving human-computer dialogues. In K. Sparck Jones & P. Willett (Eds.), *Readings in information retrieval* (pp. 135–142). San Francisco: Morgan Kaufmann. (Reprinted from RIAO'85, Actes of the

conference: Recherche d'informations assistée par ordinateur [Proceedings: Computer Assisted Information Retrieval] (pp. 645–660). Grenoble: I.M.A.G.

Hjorland, B. (1997). *Information seeking and subject representation*. Westport, CT: Greenwood Press.

Howard University College of Arts and Sciences. (n.d.). *Essay styles*. Retrieved July 8, 2003, from http://www.howard.edu/CollegeArtsSciences/Freshman%20Seminar/essay_styles.htm

Kuhlthau, C.C. (1991). Inside the search process: Information seeking from the user's perspective. *Journal of the American Society for Information Science*, 42(5), 361–371.

Kuhlthau, C.C. (1993). *Seeking meaning: A process approach to library and information services*. Norwood, NJ: Ablex Publishing.

Minsky, M. (1975). A framework for representing knowledge. In P. Winston (Ed.), *The psychology of computer vision* (pp. 211–277). New York: McGraw-Hill.

Oddy, R.N. (1975). *Reference retrieval based on user induced dynamic clustering*. Unpublished doctoral dissertation, University of Newcastle upon Tyne.

Oddy, R.N. (1977). Information retrieval through man-machine dialogue. *Journal of Documentation*, 33(1), 1–14.

Oddy, R.N., Liddy, E.D., Balakrishnan, B., Bishop, A., Elewononi, J., & Martin, E. (1992). Towards the use of situational information in information retrieval. *Journal of Documentation*, 48(2), 123–171.

Oddy, R.N., Palmquist, R.A., & Crawford, M.A. (1986). *Proceedings of the 49th Annual Meeting of the American Society for Information Science* (pp. 248–254). Medford, NJ: Learned Information.

Popper, K. (1975). *Objective knowledge: An evolutionary approach*. Oxford: Clarendon Press.

Saracevic, T., Kantor, P., Chamis, A.Y., & Trivison, D. (1988). A study of information seeking and retrieving. III. Searchers, searches, and overlap. *Journal of the American Society for Information Science*, 39(3), 197–216.

Svenonius, E. (2000). *The intellectual foundation of information organization*. Cambridge, MA: MIT Press.

Taylor, R.S. (1968). Question-negotiation and information seeking in libraries. *College & Research Libraries*, 29(3), 178–194.

Wersig, G., & Neveling, U. (1975). The phenomena of interest to information science. *The Information Scientist*, 9(4), 127–140.

Whitmire, E. (2002). Disciplinary differences and undergraduates' information seeking behavior. *Journal of the American Society for Information Science and Technology*, 53(8), 631–638.

Yuan, X.-J., Belkin, N.J., & Kim, J.-Y. (2002). The relationship between ASK and relevance criteria. In K. Jarvelin, M. Beaulieu, R. Basza-Yates, & S.H. Myaeng (Eds.), *SIGIR'02: Proceedings of the 25th Annual ACM SIGIR International Conference on Research and Development in Information Retrieval* (pp. 359–360). New York: ACM.

Appendix

Interview Schedule January–April 2003

1 st sheet	1.	Could you please describe and write down the topic of your paper?
	2.	What words, search terms, keywords or concepts will you use to look for information to write your paper? Please write out the terms.
	2b.	Include any terms from the topic sentence.
2 nd sheet	3.	Could you please label these terms, using initials or abbreviations? Could you rank these terms according to their importance in your paper?
	4.	Could you in some way visualize, draw, map or chart these words, terms or concepts in your paper showing the relationship(s) between them, if any? *Use the labels you have chosen.

- [When finished]: Please write the search term(s) represented by each label on the diagram. Please verbally explain your diagram.
- 3rd sheet 5. Could you now visualize these same terms, using circles and lines this time? The circles should vary in size according to importance of terms. Their distance apart should be representative of how closely related the concepts are. The lines should show inter-relationships amongst the terms.
*Please use the labels you chose for each term to label the circles rather than write the term in the circle.
[When finished]: Please write the search term words on the diagram and please verbally explain your diagram.
- 4th sheet 6. Could you please now write down 4 questions you would like answered by your information source?
7. Could you please rank these 4 questions by their importance to your paper?
- 5th sheet 8. Concerning the question you just ranked as most important, please write out the terms you would use to look for information for that question. Please label these terms with initials or abbreviations.
Could you map out these terms in a new diagram? Please use these labels on your diagram.
[When finished]: Please write the search term words on the diagram and please verbally explain your diagram.
- 6th sheet 9. Having gone through this process, can you now write down what your paper will be about? Feel free to make changes or integrate new thoughts.
10. What words, search terms, keywords or concepts would you use now?
Please write them down, and rank these terms according to their importance to your paper.
11. Of these approaches [the original list, your initial visualization, the circles and lines diagram, the last diagram, or your final list of terms] which would you find most useful in finding information to write your paper?
Please mark the one you prefer and verbally explain why would you prefer it.