

Typology of User Uncertainty in the selection of Web Search Terms: Insight into the Information Seeking Context of Scholarly Researchers in the Field of Science*

웹 검색어 선택과정에서의 이용자 불 확신성의 유형: 자연과학연구자들의 정보탐색환경에 대한 고찰

김양우 (Yang-woo Kim)**

ABSTRACT

While numerous studies have suggested the significance of uncertainty during the process of information-seeking, less research has investigated user uncertainty in the actual search process using a real system. This study investigated user perceptions of uncertainty in the process of *the selection of Web search terms* in the real information-seeking process. The subjects at the doctoral or post-doctoral level were limited to the discipline of science in order to understand user perceptions in this field. The findings revealed various dimensions, types, and incidents of uncertainty. The typology of uncertainty facilitated an understanding of the subjects' information-seeking context by identifying various aspects of the context that constituted the subjects' uncertainty. The identification of two principal origins of uncertainty based on the different types of uncertainty generated implications to improve information systems and services.

초록

다수의 연구에서 정보추구 과정상 불 확신성(Uncertainty)의 중요성이 지적되었지만, 실제 정보검색시스템을 이용한 탐색과정에서 이용자들의 불 확신성에 대한 연구는 많지 않았다. 본 연구는 실제로 정보를 추구하는 이용자들의 웹 검색어 선정과정에서의 불 확신성 인식을 조사하여, 정보탐색 과정에서의 다양한 불 확신성 유형을 식별하였다. 불 확신성 유형에 입각하여 발견된 불 확신성의 주요 원인(Origins)은 정보검색시스템 및 서비스 발전을 위한 시사점을 제시하여준다.

키워드: 이용자 불 확신성, 정보탐색행태, 정보탐색전략, 이용자연구, 정보검색, 정보서비스
User Uncertainty, Information Search Behavior, User Study, Information Retrieval

* This paper is based on dissertation research conducted in Rutgers University.

** 한성대학교 전임강사 (ykim@hansung.ac.kr)

1. Introduction

While numerous studies have suggested the significance of uncertainty during the process of information-seeking, less research has investigated user uncertainty in the actual search process and related the findings to specific characteristics of their information seeking context. This study investigated user perceptions of uncertainty in the process of *the selection of search terms* in the real information-seeking process.

Various Web search engines have become vital tools for accessing both academic and non-academic information likewise, they serve as important supplementary tools to traditional bibliographic databases in the academic research environment. Reflecting this reality, the present study investigated the search term selection stage of scholarly researchers, specifically at the doctoral or post-doctoral level, who use Web search engines for their academic study. The researchers were limited to the discipline of science (pure and

applied) in order to understand user perceptions in this field.

Based on the analysis of the users' uncertainty, this study found insights into the information seeking context of the users concerning their information need, system and its resource used.

2. Background

2.1 Uncertainty Study

A substantial number of studies in the field of Library and Information Science (LIS) discuss the notion of uncertainty in various contexts associated with the process of information-seeking. Variations in the use of terminology have also been found to represent the concept of uncertainty. These studies can be classified into three broad categories, as indicated below.

The first category discusses a relationship between information and uncertainty in the general flow of information. Relevant literature includes: (1) *information as a means of reducing uncertainty* (Artandi, 1973; Shannon & Weaver, 1949), and (2) *exceptions in*

uncertainty reduction with information provided (Whittemore & Yovits, 1973). These studies deal mainly with the role of information in uncertainty reduction.

The second category discusses uncertainty as it relates to various aspects of the information-seeking process in a focused domain. Relevant literature in additional subcategories includes: (1) cognitive aspects of the information need — difficulty in representing (Taylor, 1968) and articulating an information need (Belkin, 1980) a need for information-seeking with a knowledge-gap (Dervin, 1983) and poly-representative nature of the cognitive space (Ingwersen, 1992, 1996) (2) connecting the internal and expressed needs of users — the necessity of dealing with the discrepancies between the two needs, further developing Taylor's (1968) approach (Gerhan, 1999; Ingwersen, 1982; Michell & Dewdney, 1998; Stevens, 1988); (3) uncertainty for system design — uncertainty as a system design principle (Bates, 1986);

poly-representative nature of the information space and uncertainly, unpredictability and contextualization in IR (Ingwersen, 1992, 1996) types of uncertainty related to the process of indexing and matching (van Rijsbergen, 1996); and the need to accommodate uncertainty in systems design (Kuhlthau, 1999); and (4) relevance judgment — difficulty in determining the usefulness of documents (Rodriguez, 2000), particularly as a concrete binary judgment (Spink, 1997; Spink & Greisdorf, 1997; Spink, Greisdorf, & Bateman, 1998).

The third and last category discusses uncertainty as it relates to various aspects of the information-seeking process in an extensive domain. Relevant literature includes: (1) affective states due to cognitive uncertainty in a six-stage information search process (Kuhlthau, 1993a, b); (2) the role of information-seeking in uncertainty resolution during the problem-solving process (Wilson, Ellis, & Ford, 2000; Wilson, Ellis, Ford, & Foster, 1999); and (3)

correlations between Kuhlthau's feelings dimensions and uncertainty at different problem stages (Wilson, Ford, Ellis, Foster, & Spink 2002).

While the above studies in these three categories present both theoretical and practical discussions related to uncertainty, less attention has been paid to perceptions of uncertainty by end-users who are engaged in an actual search process with a real information system. This is particularly true in examining uncertainty for the characteristics of the search terms selected concerning users' information needs, information systems and resources.

2.2 Web Search Behavior Study

Much research has examined user behaviors in Web-based information-seeking within different contexts, but without specifically focusing on user uncertainty. That is, user behaviors have been examined more generally or with a specified focus other than uncertainty.

Nevertheless, several relevant findings can be classified into five categories based on different aspects of information-seeking: (1) search term selection — *difficulties* in formulating Web searching queries (Pollock & Hockley, 1996); user *mistakes* on Web query inputs (Jansen, Spink, & Saracevic, 2000); and the *limited capacity* of inexperienced searchers for Web search term selection (Lucas & Topi, 2002) (2) search term selection and system features — *differences* in features among search engines for query formulation (Gordon & Pathak, 1999); *problems* in submitting queries and *difficulties* in using Boolean operations for an academic Web site (Wang, Berry, & Yang, 2003); and the association of Boolean searching on the Web with user *anxiety* (Ford, Miller, & Moss, 2003); (3) system features — *confusion* in using Web search engine features (Su, 2003); (4) results evaluation — *difficulty* and *confusion* in comprehending retrieved resources of library sites (Fitzgerald & Galloway, 2001); and (5) the extensive process of

searching —user *problems* and system *problems* in working with online library systems (Rousseau, et al. 1998);sources of user *frustration* with the Web search and users' suggestions for interface improvements (Fidel, Davies, Douglass, Holder, Hopkins, & Kushner, 1999) and types of *problems* and *anxieties* during the Web search (Wang, Hawk, & Tenopir, 2000). Although these studies present meaningful points related to users' uncertainty on various aspects, they focused primarily on the extensive domain of general Web searching behaviors or specified areas of such behaviors other than uncertainty.

Overall, then, these studies provide only limited discussion on uncertainty in user selection of search terms. Thus, more focused yet extensive investigation is needed in this area, using a data collection framework specifically designed for this topic, user perceptions of uncertainty in the selection of search terms for Web-based information-seeking.

3. Research Method

3.1 Research Scope

A few considerations were necessary to determine the scope of the present study. The first consideration was to choose the disciplinary areas and academic status of the information searchers investigated. The researcher decided to focus on academic researchers in the field of science who search the Web using a Web search engine.

The second consideration was the scope of search engine use. This study focused on investigating the use of the Web search engine as the primary search tool to locate various World Wide Web sites. The use of a search engine only to locate another Web-based search system (i.e., online journal databases or OPACs) while using the located search tool as a main search system was beyond the scope of this investigation. This limitation facilitated the examination of various characteristics of Web search engines as information retrieval systems.

3.2 Collection and Analysis of Data

This study investigated the users' information search process within a natural workplace setting. Fifteen subjects at Rutgers University participated in this study. All participants were engaged in an actual research process with a real information need. They sought to use a commercial Web search engine to search the World Wide Web based on their own need, not because they were participating in this study. Thus, these conditions constituted the natural setting of the data collection process.

Following a qualitative approach for data collection and analysis, the researcher asked each subject to fill in a questionnaire and participate in a subsequent interview. The subjects' responses were coded and grouped, characterizing various aspects of uncertainty. As a result, different dimensions and types of uncertainty were identified. A type of uncertainty is a category of incidents of uncertainty with a common

characteristic.

A dimension of uncertainty specifies a broader category than type it is a category of the *types* of uncertainty. Each dimension represents a category of similar types of uncertainty.

4. Results

Data analysis led to the identification of two dimensions of uncertainty and nine different types of uncertainty out of a total of 40 incidents. The types of uncertainty assigned to the dimensions are shown in Table 1, with the number of incidents of uncertainty in each type and an example of the incidents presented for each type. Each example of incident is accompanied by the subject's corresponding description. In addition, the definition of each type and a list of all incidents of uncertainty are shown in Tables 2 and 3, respectively.

The two dimensions of uncertainty identified in the selection of search terms were: *expressive uncertainty* and *system uncertainty – tool and/or resource*. The first dimension, *expressive*

uncertainty, concerned the subjects' uncertainty in the process of representing their internal need by selecting search terms. For example, an uncertainty in this dimension related to a subject's perception of whether an expressed need or query corresponded to an internal need (see left portion of Figure 1).

This dimension relates to Taylor's (1968) levels of question formation (visceral, conscious, formalized, and compromised need), which were discussed in section 2.1. According to the subjects' responses, their perceptions of prospective system response did not influence their uncertainty here.

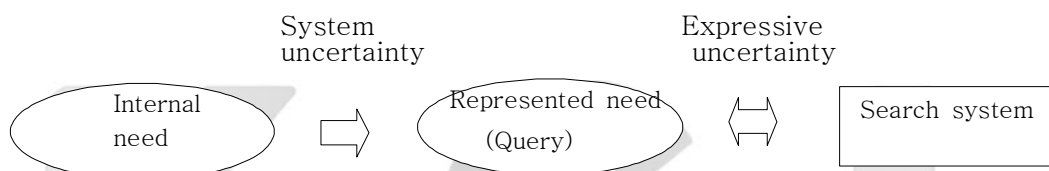


Figure 1. Expressive Uncertainty and System Uncertainty

The type of uncertainty *representation of internal needs* was identified within this dimension (see Tables 1, 2, and 3). The type of uncertainty *representation of internal needs* concerned the selection of search terms that accurately describe a user's perceived information need (see Table 2). Two incidents of uncertainty categorized into this type were:

- *selection of topical terms – lack*

of domain knowledge (S005) and

- *consistency between internal need and expressed need* (S011) (see Table 3).

For example, Subject 005 indicated uncertainty in *selection of topical terms – insufficient knowledge of a topic* by the following response:

"Choice of search terms (antioxidants – glutathione, atherosclerosis, LDL, alpha-tocopherol).... This is

<Table 1> Types of Uncertainty in the Search Strategy – Selection of Search Terms

Dimension of uncertainty	Types of Uncertainty	No. of incidents	Example of Incident of Uncertainty	Subjects' Descriptions (Subject # – Instrument)
Expressive uncertainty	Representation of internal needs	2	Consistency between internal need and expressed need	"Search term(s) <u>may not exactly point out my need.</u> " (S011- Q)
System uncertainty – tool and/or resource	Search output	4	Avoiding excessive number of returns	"Before I try it, I am <u>wondering [about the] effectiveness of my keywords since the Web searching commonly showed me a ton of hits.</u> " (S001 - Q)
	Search mechanism – polysemy/homonymy	9	Relevance of items retrieved – terms in different contexts	"Nickel. Gives results unrelated to catalysts." (S013 - Q); " <u>...which are pretty common words, you know, nickel, or catalysis, and upon typing those words individually into Google, you get results most of which are not related to what you're interested in.</u> " (S013 - I)
	Search mechanism – use of search feature	4	Use of Boolean operators	"There's <u>no set method of [for]the formatting of the operators on the different databases so once again you have to look everything up. And also on some of the search engines, they don't seem to work the same.</u> " (S010 - I)
	Search mechanism – abbreviation, singular/plural & truncation of term	3	Deciding the use of singular/plural for search term	"Technical reports" OR "Technical report" (S003 - Q); " <u>I don't know actually whether or not Google performs stemming or whether they truncate this 'reports' term to 'report.'</u> " (S003 - I)
	Search mechanism – resource type	2	Specifying the type of resource – 'lecture note' instead of 'review'	"Which term should I use to specify a particular resource?" (S015 - Q) "Search term <i>Review</i> worked for another need but I am not sure how it will work for this particular need. I may want to be more specific for resource type using <i>Lecture Note</i> as a search term." (S015 - I)
	Search skill	3	Formulating query – search terms as a phrase	"Not sure how to present search terms as a phrase." (S007 - Q); "What I want is to find out some site that included every search term...together...If I put like 3-D cell migration...there are three words. <u>Three-D cell migration. I wanna search every term together... not separately. (asked about using a quotation mark or the exact phrase search available in advanced search) No... I did not know.</u> "(I)
	Specificity of search term	6	Extent to narrow down search terms	"When I use <u>narrow search terms, I could get a good result but might miss other good results.</u> " (S001 - I)

	Inter-disciplinary & inter-author/organization discrepancy	8	Inter-disciplinary variation in the use of terminology – selecting right nomenclature in the field	"I am a biochemical engineer, but I am trying to find information from [the] biology and chemistry field[s]...there might be some difference between engineering terms and science terms." (S001-Q) "I do have [an] engineering background, but I'm working on plant science research field. I have not studied plant biology well. Maybe I can miss some important terminology." (S001-I)
Total		40		

Note. Q = Questionnaire; I = Interview

related to how well the term you choose to put in your search engine describes what you need to find and whether there would be some other terms that would describe your need better. And this is related to the knowledge also on your knowledge of the subject... So if you don't have sufficient knowledge of your topic yet and you are at the initial stage, you may not choose the best, the most appropriate terms for the search you need. So you won't get that relevant information to your topic.... It's not relevant to the search engine, it's an uncertainty of the terms you choose to put in your search engine."

In another example, Subject 011 indicated uncertainty in *consistency between internal need and expressed need* by the following response:

"Search terms may not exactly point out my need...What matters is that maybe my choice of the words or my combination of the words there do not express exactly what I am looking for. So it can take me somewhere else, some different things that I don't want, so that's the kind of uncertainty" (see Table 1).

Both of these incidents of uncertainty have no relation to the subjects' perceptions of system responses on query terms. Accordingly, these incidents belong to the first dimension.

<Table 2>Uncertainty in the Search Strategy – Selection of Search Terms:
Definition of Types

Dimension I. Expressive uncertainty

Types of Uncertainty	Definitions
Representation of internal need	Selection of search terms that accurately describe the user's perceived information need

Dimension II. System uncertainty – tool & resource

Types of Uncertainty	Definitions
Search output	Yielding an ideal quantity of results or relevant results
Search mechanism – polysemy/homonymy	Retrieval mechanisms of search engines concerning a search term (phrase) with the different related senses of a single meaning (polysemy) or more than one literally different meaning (homonymy)
Search mechanism – use of search feature	Retrieval mechanisms of search engine with respect to using its search features
Search mechanism – abbreviation, singular/plural & truncation of term	Retrieval mechanisms of search engine with respect to the following: using abbreviation as (a) search term(s), using a singular or plural search term or determining the extent to truncate (a) search term(s)
Search mechanism – resource type	Retrieval mechanism of search engine, which locates a specific type of resource necessary to the user
Search skill	Subjects' self-perceived skill at presenting effective terms to search engine
Specificity of search term	Degree of broadness and narrowness of a term with respect to its meaning
Inter-disciplinary & inter-author/organization discrepancy	Differences between authors, disciplines, and organizations with respect to the use of terminology in various aspects (i.e., selection of different terms in representing an identical concept/entity; inconsistent use of acronym and abbreviation)

The second dimension, *system uncertainty – tool and/or resource*, related to the subjects' perceptions of how the system would respond to selected search terms (see right portion of Figure 1). Subjects' perceptions of different aspects of an information system regarding a search engine as a search *tool* and/or WWW sites as the *resource*

of the system influenced uncertainty in this dimension.

Uncertainties in this dimension were not further categorized according to the two components of the information system: *search tool* and *resource*. This was because most uncertainties or types of identified uncertainty can

<Table 3> Uncertainty in the Search Strategy – Selection of Search Terms: Incidents for Each Type

Dimension I. Expressive Uncertainty

<p>Type – Representation of internal need (2) Selection of topical terms – lack of domain knowledge (S005) Consistency between internal need and expressed need (S011)</p>

Dimension II. System uncertainty – tool & resource

<p>Type – Search output (4) Avoiding an excessive number of returns (S001) Selecting the right set to avoid an excessive number of irrelevant returns (S006) Combining terms to reduce the number of returns (S006) Avoiding an excessive number of irrelevant returns (S008)</p>	<p>Type – Search mechanism – polysemy/homonymy (9) Relevance of items retrieved – terms in different contexts (S005) Relevance of items retrieved – use of acronyms (S005) Relevance of items retrieved – use of numbers (S009) Relevance of items retrieved – terms in different contexts (S013) Relevance of items retrieved – terms in different contexts (S013) Relevance of items retrieved – terms in different contexts (S014) Relevance of items retrieved – terms in different contexts of disciplines (S005) Relevance of items retrieved – terms in different contexts of disciplines (S012) Relevance of items retrieved – terms in different contexts of disciplines (S015)</p>
<p>Type – Search mechanism – use of search feature (4) Use of search feature – quotation mark for exact phrase search (S003) Use of search feature – Boolean operators and quotation mark (S003) Use of search feature – Boolean operators (S004) Use of search feature – Boolean operators (S010)</p>	<p>Type – Search mechanism – abbreviation, singular/plural & truncation of terms (3) Deciding the use of singular/plural of search terms (S003) System reaction to abbreviation (S007) Determining the extent to truncate terms (S010)</p>
<p>Type – Search mechanism – resource type (2) Specifying the type of resource – article instead of commercial sites (S002) Specifying the type of resource – lecture note instead of review (S015)</p>	<p>Type – Search skill (3) Appropriateness of search terms (S001) Formulating query – search terms as a phrase (S007) Matching with index term (S011)</p>
<p>Type – Specificity of search term (4) Extent to narrow down search terms (S001) Determining specificity of search terms (S002) Determining specificity of search terms (S006) Specificity of terms selected (S008)</p>	<p>Type – Inter-disciplinary & inter-author/organization discrepancy (8) Inter-disciplinary variation in the use of terminology – selecting right nomenclature in the field (S001) Variation in word form – singular/plural (S004) Variation in word form – acronym (S004)</p>

	Inter-author variation in the use of terminology (S008) Variation in word form - spelling (S010) Variation in word form - abbreviation(S010) Variation in word form - acronym (S013) Inter-disciplinary & inter-author variation in the use of terminology - selecting right nomenclature in the field (S010)
--	---

Note. Number of incidents for each type is in parentheses.
 Subject number for each incident is in parentheses.

possibly relate to both components of the information system. For example, Subject 003 expressed uncertainty in determining whether to use the singular or plural form of a search term, indicating as follows:

"I don't know actually whether or not Google performs stemming or whether they truncate this 'reports' term to 'report'."

This uncertainty explicitly related to the function of an information system as *a search tool*. Yet, the search result can be influenced by whether the search term "Technical reports" or "Technical report" was presented as singular or plural in the text of the *WWW resource*. This suggests that this particular incident related to a characteristic

of *WWW resource* as well (see Table 1).

Eight different types of uncertainty identified within this second dimension were:

- *search output*
- *search mechanism - polysemy/homonymy*
- *search mechanism - use of search feature*
- *search mechanism - abbreviation, singular/plural & truncation of term*
- *search mechanism - resource type*
- *search skill*
- *specificity of search term and*
- *inter-disciplinary & inter-author/organization discrepancy* (see Tables 1, 2, and 3).

The above types were assigned to this dimension because they

related to the users' perceptions of different aspects of an information system. Different aspects (i.e., functions of interface features) influenced the identification of these types. The following discusses the types of uncertainty in the second dimension and presents the incidents of uncertainty assigned to each type.

The type of uncertainty *search output* concerned yielding an ideal quantity of results or relevant results (see Table 2). The four incidents of uncertainty categorized into this type included:

- *avoiding an excessive number of returns* (S001);
- *selecting the right set to avoid an excessive number of irrelevant returns* (S006);
- *combining terms to reduce the number of returns* (S006); and
- *avoiding an excessive number of irrelevant returns* (S008) (see Table 3).

For example, Subject 001 indicated uncertainty in *avoiding an excessive number of returns* by following response:

"Before I try it, I am

wondering [about the] effectiveness of my keywords since the Web searching commonly showed me a ton of hits" (see Table 1).

The type of uncertainty *search mechanism – polysemy/homonymy* concerned retrieval mechanisms of search engines concerning a search term/phrase with the different related senses of a single meaning (polysemy) or more than one literally different meaning (homonymy). The definitions of polysemy and homonymy reflect those provided in relevant literature on the notion of *ambiguity*. The researchers include Cottrell (1989) and Small, Cottrell, and Tanenhaus (1988). (see Table 2). The nine incidents of uncertainty categorized into this type included:

- *relevance of items retrieved – terms in different contexts* (S005);
- *relevance of items retrieved – use of acronym* (S005);
- *relevance of items retrieved – use of numbers* (S009);
- *relevance of items retrieved –*

terms in different contexts (S013, S013, S014); and

- *relevance of items retrieved – terms in different contexts of disciplines* (S005, S012, S015) (see Table 3).

For example, Subject 013 indicated uncertainty in the *relevance of items retrieved – terms in different contexts* by the following response:

"'Nickel.' Gives results unrelated to 'catalysts'...which are pretty common words, you know, 'nickel,' or 'catalysis,' and upon typing those words individually into Google, you get results, most of which are not related to what you're interested in" (see Table 1).

In this incident, the subject questioned the system response to an ambiguous word, "nickel," which can be either a coin or a kind of metal.

The type of uncertainty *search mechanism – use of search feature* concerned the retrieval mechanisms of search engines regarding their search features (see Table 2). The four incidents

of uncertainty categorized into this type included:

- *use of search feature – quotation mark for exact phrase search* (S003);
- *use of search feature – Boolean operators and quotation mark* (S003); and
- *use of search feature – Boolean operators* (S004, S010) (see Table 3).

For example, Subject 010 indicated uncertainty in the *use of Boolean operators* by the following response:

"There's no set method of [for] the formatting of the operators in the different databases, so once again you have to look everything up. And also on some of the search engines, they don't seem to work the same"(see Table 1).

The type of uncertainty *search mechanism – abbreviation, singular/plural & truncation of term* concerned the retrieval mechanisms of a search engine with respect to using an abbreviation as a search term,

using a singular or plural search term or determining the extent to truncate a search term (see Table 2). The three incidents of uncertainty categorized into this type included:

- *deciding the use of singular/plural of a search term* (S003);
- *system reaction to abbreviation* (S007); and
- *determining the extent to truncate terms* (S010) (see Table 3).

For example, Subject 003 indicated uncertainty in deciding on the use of the singular/plural of a search term by the following response:

"'Technical reports' OR 'Technical report.'... I don't know actually whether or not Google performs stemming or whether they truncate this 'reports' term to 'report.'" (see Table 1).

The type of uncertainty *search mechanism – resource type* concerned the retrieval mechanism of a search engine, which located a specific type of resource

necessary for the user (see Table 2). The two incidents of uncertainty categorized into this type were:

- *specifying the type of resource – article instead of commercial sites* (S002) and
- *specifying the type of resource – "lecture note" instead of "review"*(S015) (see Table 3).

For example, Subject 015 indicated uncertainty in *specifying the type of resource – "lecture note" instead of "review"* by the following response:

"Which term should I use to specify a particular resource?... Search term 'Review' worked for another need, but I am not sure how it will work for this particular need. I may want to be more specific for a resource type using 'Lecture Note' as search term" (see Table 1).

The type of uncertainty *search skill* concerned subjects' self-perceived skills at presenting effective terms to a search engine (see Table 2). The three incidents of uncertainty categorized

into this type included:

- *appropriateness of search terms* (S001);
- *formulating query – search terms as a phrase* (S007); and
- *matching with index term* (S011) (see Table 3).

For example, Subject 007 indicated uncertainty in *formulating query – search terms as a phrase* by the following response:

"Not sure how to present search terms as a phrase... What I want is to find out some site that included every search term...together...if I put like 3-D cell migration.... There are three words. Three-D cell migration. I want to search every term together...not separately. [asked about using a quotation mark or the exact phrase search available in advanced search] No...I did not know" (see Table 1).

In this incident, the subject did not indicate difficulty in expressing his need. Instead, he experienced uncertainty in presenting his need appropriately to the information system.

The type of uncertainty *specificity of search term* concerned the degree of broadness and narrowness of a term with respect to its meaning (see Table 2). The four incidents of uncertainty categorized into this type included:

- *extent to narrow down search terms* (S001);
- *determining specificity of search terms* (S002, S006); and
- *specificity of terms selected* (S008) (see Table 3).

For example, Subject 001 indicated uncertainty in the *extent to narrow down search terms* by the following response:

"When I use narrow search terms, I could get a good result, but might miss other good results" (see Table 1).

This incident suggests that the user was aware of the relationship between recall and precision.

The type of uncertainty *inter-disciplinary & inter-author/organization discrepancy* concerned the differences between authors, disciplines, and organizations with respect to the use of terminology

in various aspects (i.e., selection of different terms in representing an identical concept and entity inconsistent use of acronym and abbreviation) (see Table 2). The eight incidents of uncertainty categorized into this type included:

- *inter-disciplinary variation in the use of terminology – selecting right nomenclature in the field*(S001);
- *variation in word form – singular/plural* (S004);
- *variation in word form – acronym* (S004);
- *inter-author variation in the use of terminology* (S008);
- *variation in word form – spelling* (S010);
- *variation in word form – abbreviation* (S010);
- *variation in word form – acronym* (S013);
- *inter-disciplinary & inter-author variation in the use of terminology – selecting right nomenclature in the field* (S010) (see Table 3).

For example, Subject 001 indicated uncertainty in

inter-disciplinary variation in the use of terminology by the following response:

"I am a biochemical engineer, but I am trying to find information from [the] biology and chemistry fields...there might be some difference between engineering terms and science terms... I do have [an] engineering background, but I'm working in the plant science research field. I have not studied plant biology well. Maybe I can miss some important terminology"(see Table 1).

In this incident, the subject was aware of a problem in selecting search terms, but showed the lack of ability to solve it.

5. Discussion

This section has the following three subsections. The first subsection, 5.1, discusses exemplary incidents of uncertainty in relation to the information

seeking context of the subjects. The second subsection, 5.2, addresses the usefulness of the typology of uncertainty for understanding the subjects' uncertainty and their information seeking context. The final subsection, 5.3, explains primary origins of uncertainty and its implications to improve information systems and services.

5.1 Uncertainty and Understanding the Information Seeking Context

The analysis of uncertainty in the information-seeking process offers insights into the understanding of individuals' information-seeking context. As indicated, the incidents of uncertainty identified in the search term selection phase reflected the subjects' perceptions of their information needs, the systems used, and the resources of the systems. Accordingly, the incidents of uncertainty identified in this phase facilitated an understanding of the following components with respect to

selecting search terms: scholarly researchers' information needs in the area of science the Web search engines selected by the researchers and WWW sites, the resources provided by the search engines.

Some incidents of uncertainty identified in this phase of the information search did not directly pertain only to the characteristics of Web search engines. Rather, those incidents may occur in many other retrieval systems. For example, Subject 013 showed uncertainty in the search mechanism of a Web search engine with respect to handling an *ambiguous* term (*nickel*) with more than one meaning (i.e., nickel for metal or for coin). Yet, in this particular incident — a chemistry researcher using a search term, *nickel* — the user would likely have a greater uncertainty in selecting this search term (*nickel*) when using a commercial Web search engine compared to a specialized database in the chemistry field. The term would have less variation in its meanings in the texts of a chemistry

database, compared to those of WWW sites which have no limitations in topical domain.

The notion of *ambiguity* has gained significant attention in the IR community. A considerable body of literature discussed the ambiguity of (query) terms in IR contexts, pointing out the shortcomings of retrieval systems in handling ambiguous terms. Most recent works include a study by Kim (2003), who extensively reviewed relevant studies and presented a typology of various ambiguities in user representations of information problems. By contrast, the present study focused on various information seeking phases and identified the ambiguous representation as one of many reasons for user uncertainty. Kim (2003) analyzed user representations of information needs in his ambiguity study. By doing so, he discussed the prospective improvement of retrieval systems and services in

handling such representations. In Kim' study, the following three aspects of retrieval systems and services were emphasized as an attempt to enhance user-system interaction and user-information intermediary interaction: (1) increasing user-inputs; (2) reducing search space by disambiguating queries and (3) clustering search results according to the characteristics of located resources.

Some incidents of uncertainty in this information seeking phase more explicitly explained the users' information seeking context. The following different aspects of the information seeking process influenced these incidents: user need concerning its disciplinary area, and information system with regard to inter-text variation in representing an acronym. Ultimately, three incidents of uncertainty were identified, as shown in Table 4.

<Table 4> Uncertainty and the Information Seeking Context – Search Term Selection

Uncertainty		Information Seeking Context (ISC)	
Exemplary Incidents	Subjects' Descriptions (Subject # - Instrument)	Components of the ISC	Related Characteristics of the Components
Inter-disciplinary variation in the use of terminology – selecting right nomenclature in the field	"I am a <u>biochemical engineer</u> , but I am trying to find information from [the] biology and chemistry fields... <u>there might be some difference between engineering terms and science terms.</u> "(S001-Q)	User need – disciplinary area (biochemical engineering) Resource	Multidisciplinary aspect of research/information need Field of pure science
Inter-author variation in the use of terminology –selecting right nomenclature in the field	"Also, the problem is that <u>even within chemistry, people will use different terms.</u> If terms are, if things were discovered and synthesized before IUPAC* came out with a law for them, if you're dealing with older professors in older groups, they will use the other terms. <u>This is very frequent, this very frequently happens in the organic field...</u> For a common chemical we use now, in the United States, there are twelve names commonly being used for it." (S010- I)	Resource – disciplinary area (chemistry)	A high degree of inter-author discrepancy in the use of terminology
Relevance of items retrieved – use of acronym	"Use of acronyms (e.g., LDL Low-Density Lipoproteins)" (S005 - Q) " <u>This acronym is usually recognized by an academic database if you put it as a search term, while if you put it in a commercial search engine, this could cause a problem because there may be other substances or companies or anything else that has [have] these three letters as an acronym.</u> " (S005 - I)	Information system & resource – Web search engine and WWW sites	A high degree of variation in the representation by an acronym (WWW sites)

Note. IUPAC = *International Union of Pure and Applied Chemistry*. (The subject meant glossary books published by IUPAC); Q = Questionnaire; I = Interview

The first two incidents in the table relate to the users' disciplinary areas of biochemical engineering and chemistry (see Table 4). In the first incident, Subject 001 showed uncertainty

concerning *inter-disciplinary variation in the use of terminology*.

This particular uncertainty relates to the multidisciplinary aspect of the subject's information need as a biochemical engineering

researcher and the disciplinary area of necessary resources. The second incident also pertained to the user's disciplinary field and a characteristic of resources in the field. The chemistry researcher (Subject 010) emphasized a high degree of inter-author discrepancy in naming chemicals in her field in the following response:

"For a common chemical we use now in the United States, there are twelve names commonly being used for it."

This caused Subject 010 uncertainty in selecting the right nomenclature in the field.

In the third incident, Subject 005 attributed her uncertainty to the use of a Web search engine instead of an academic database. Concerning the use of the acronym LDL as a search term, the subject expressed uncertainty in the relevance of the items retrieved because of a possible variation in the use of the acronym in WWW sites. The subject pointed out a difference between using a search engine or a database in the following response:

"This acronym is usually recognized by an academic database if you put it as a search term, while if you put it in a commercial search engine, this could cause a problem because there may be other substances or companies or anything else that has [have] these, these three letters as an acronym."

The incidents of uncertainty identified in the search term selection phase offered insights into the users' information seeking context concerning their information need and the information system and resource they selected. The findings revealed variations in different incidents of uncertainty with respect to their relatedness to the unique characteristics of the information seeking context. As indicated, some incidents more explicitly differentiated the users' information seeking context from other such contexts.

5.2 Types of Uncertainty and Facilitation of Understanding the Information Seeking Context

Identifying the dimensions and types of uncertainty in this study facilitated an understanding of the users' uncertainty, thereby providing insights into their information seeking context. Both dimensions and types characterized specific aspects of the user, the information system, and/or its resources. For example, in the search term selection phase, the dimensions and types facilitated an understanding of the users' needs, the information system, and selected resources. The typology in this phase was useful for identifying various *aspects* of the uncertainty related to the scholarly researchers' information needs, Web search engines, and World Wide Web sites.

As discussed, the analysis of uncertainty in selecting the search terms identified two dimensions of uncertainty: *expressive uncertainty* and *system uncertainty – tool & resource*. This indicated that the users' uncertainty in this phase

related to the following aspects of the information seeking process: *correspondence between an expressed need and an internal need or the response of a system to search terms presented* for the first and second dimensions, respectively.

The types of uncertainty identified within each dimension indicated that the subjects perceived uncertainty in various aspects related to the information seeking process. For example, the type of uncertainty within the first dimension, *expressive uncertainty*, was *representation of internal needs*, which concerned the accurate representation of the users' perceived information needs (see Tables 1, 2, and 3).

The types of uncertainty identified within the second dimension, *system uncertainty – tool and/or resource*, revealed more diverse aspects of the information-seeking process. These different aspects constituting the users' uncertainty in each type were: quantity of search results, as revealed in the type of uncertainty *search output*;

a search term with the different related senses of a single meaning (polysemy) or multiple meaning (homonymy), as revealed in the type of uncertainty, *search mechanism – polysemy/homonymy*; search features and the retrieval mechanism of search engines, as revealed in the type of uncertainty *search mechanism – use of search feature*; abbreviation, singular and plural, and truncation of terms, as revealed in the type of uncertainty *search mechanism – abbreviation, singular/plural & truncation of term*; a resource type necessary for the user, as revealed in the type of uncertainty *search mechanism – resource type*; subjects' self-perceived skills, as revealed in the type of uncertainty *search skill*; the broadness and narrowness of the meaning of search terms, as revealed in the type of uncertainty *specificity of search term*; and differences between authors, disciplines, and organizations with respect to the use of terminology, as revealed in the type of uncertainty *inter-disciplinary &*

inter-author/organization discrepancy (see Tables 1, 2, and 3).

As addressed above, the typology of uncertainty helped to identify different aspects of the information seeking process that constituted the users' uncertainty. The identification of those aspects facilitated an understanding of the users' information seeking context, which concerned varying components in the different phases of information seeking.

5.3 Origin of Uncertainty and Its Implications

This section discusses two principal origins of uncertainty and the implications of such findings on improving information systems and services. The information system perspective mainly focused on discussing the various aspects of Web search engines that caused the users' uncertainty.

The information service perspective may require an explanation about the study's data collection method. Although the dataset for this study was collected without a human

information intermediary engaged in the subjects' search process, the findings can be nonetheless meaningfully extended to the prospective design of information services. The dialogue interaction between researcher and subjects in the actual information-seeking process, as examined in this study, can contribute a good prototype for interactions between

intermediaries and information users.

For example, understanding a patron's uncertainty would be an important segment of a reference librarian's task and would, in turn, facilitate the librarian in providing appropriate service to the user. This supports speculation on how to improve information services, as discussed below in this section.

<Table 5> Origins of Uncertainty and Types of Uncertainty Related

Origins of Uncertainty		Types of Uncertainty Related
User-side	Limited knowledge of the functions and/or features of information systems	Search skill Search mechanism -use of search feature
System-side	Limited retrieval mechanisms for utilizing the characteristics of resources	Search mechanism - polysemy/homonymy Inter-disciplinary & inter-author/organization discrepancy

The findings, as shown in Table 5, revealed the following principal origins of user uncertainty. The user-side involved *limited knowledge of the functions and/or features of the information system*. The system-side included *limited retrieval mechanisms for utilizing the characteristics of resources*.

The identification of the above origins did not aim to exclusively encompass all types of uncertainty or incidents of identified uncertainty. Some types or incidents of uncertainty related to more than one origin. In some incidents, the subjects' descriptions did not provide a sufficient rationale to relate uncertainty to

any of the two origins. To some extent, the above origins were not completely comprehensive. Yet, identifying the two principal origins was important as it led to practical ideas that generated the implications of this study.

The origin of uncertainty on the user-side concerned *limited knowledge of the functions and/or features of the information system*. A type of uncertainty related to this origin was *search mechanism – use of search feature*. The subjects showed limited knowledge of the following area in this type of uncertainty: retrieval mechanisms of search engine with respect to using its search features.

This origin of uncertainty suggested the need to improve information systems and to provide relevant services for users who have limited knowledge of the above areas. The system perspective concerned the enhancement of the Help section interface. HCI research in Help systems indicates that most computer users seldom read the instruction manual or use "Help."

In the information service perspective, a possible intervention for users who have limited knowledge of the above aspect would involve providing appropriate explanations about the system or suggesting the use of other applicable systems.

The origin of uncertainty on the system-side involved *limited retrieval mechanisms for utilizing the characteristics of the resource*. The types of uncertainty related to this origin included: *search mechanism – polysemy/homonymy*, and *inter-disciplinary & inter-author/organization discrepancy*. These types of uncertainty related to the limited retrieval mechanisms of Web search engines for utilizing characteristics of Web resources. In other words, Google and Yahoo, the Web search engines used in this study utilized only restricted aspects of the resource characteristics during the retrieval process. Such limitations influenced the users' uncertainty.

For instance, as an example of *search mechanism – polysemy/homonymy*, if the system

had further utilized the related characteristics of resources, the retrieval of irrelevant sites, relying on the occurrence of a certain search term, would have been avoided. An example of such utilizations comes from the analysis of terminologies that appear in the text to identify topics covered or intended audience. The analysis of a Web site's terminology would indicate its domain or targeted audience. This suggests that users would experience less uncertainty if the system had a function to further utilize the characteristics of prospective resources.

Many of the above limitations of information systems have been widely treated in the IR research community, yet progress has been moderate. This study empirically supports the significance of these aspects.

From an information service perspective, a possible intervention would involve explaining the limitations of the information

system so that users have a greater understanding of the system. Thus, they would know which resource characteristics were utilized by the system during the retrieval process.

6. Conclusion

The analysis of uncertainty in this study was useful for generating implications of the research findings to improve information systems and services. It would be a meaningful attempt for LIS research to apply a similar approach to different information-seeking contexts. Related components other than researchers in science, Web search engines, and World Wide Web sites would constitute different information-seeking contexts. Further development of the research method may enable better facilitation of future research processes.

References

- Artandi, S. 1973. Information concepts and their utility. *Journal of the American Society for Information Science*, 24: 242-245.
- Bates, M. 1986. Subject access to online catalogs: A design model. *Journal of the American Society for Information Science*, 37: 357-376.
- Belkin, N. J. 1980. Anomalous states of knowledge as the basis for information retrieval. *Canadian Journal of Information Science*, 5: 133-143.
- Cottrell, G. W. 1989. *A connectionist approach to word sense disambiguation*. San Mateo, CA: Morgan Kaufmann Publishers, Inc.
- Dervin, B. 1983. *An overview of sense making research: Concepts, methods, and results to date*. Seattle, WA: School of Communication, University of Washington.
- Fidel, R., Davies, R. K., Douglass, M. H., Holder, J. K., Hopkins, C. J., Kushner, E. J., et al. 1999. A visit to the information mall: Web searching behavior of high school students. *Journal of the American Society for Information Science*, 50(1): 24-37.
- Fitzgerald, M. A., & Galloway, C. 2001. Relevance judging, evaluation, and decision making in virtual libraries: A descriptive study. *Journal of the American Society for Information Science and Technology*, 52(12): 989-1010.
- Ford, N., Miller, D., & Moss, N. 2003. Web search strategies and approaches to studying. *Journal of the American Society for Information Science*, 54(6): 473-489.
- Gerhan, D. 1999. When quantitative analysis lies behind a reference question.

- Reference and User Services Quarterly*, 39(2): 166-185.
- Gordon, M., & Pathak, P. 1999. Finding information on the World Wide Web: The retrieval effectiveness of search engines. *Information Processing and Management*, 35: 141-180.
- Ingwersen, P. 1982. Search procedures in the library -analyzed for the cognitive point of view *Journal of Documentation*, 38(3): 165-191.
- Ingwersen, P. 1992. *Information retrieval interaction*. London: Taylor Graham.
- Ingwersen, P. 1996. Cognitive perspectives of information retrieval interaction: Elements of a cognitive IR theory. *Journal of Documentation*, 52(1): 3-50.
- Jansen, B. J., Spink, A., & Saracevic, T. 2000. Real life, real users, and real needs: A study and analysis of user queries on the web. *Information Processing and Management*, 36(2): 207-227.
- Kim, Y. 2003. Typology of ambiguity on representation of information problem: An exploratory study. *Proceedings of the Annual Conference of the American Society of Information Science and Technology*, 366-380.
- Kim, Y. 2003. User-oriented representation of documents for Interactive IR: Connecting beyond-topical criteria of user-defined relevance with the surrogate features of an IR database. *Online Proceedings of 2003 Joint American Society of Indexers (ASI) -- Indexing and Abstracting Society of Canada (IASC/SCAD) Annual Conference, Vancouver, BC, Canada*.
- Kim, Y. 2004. User-oriented indexing: Accommodating beyond-topical aspects of documents. *Key Words: the Bulletin of the American Society of Indexers*, 12(1): 20-25.
- Kuhlthau, C. C. 1993a. A principle of uncertainty for information seeking. *Journal*

- of Documentation*, 49(4): 339-355.
- Kuhlthau, C. C. 1993b. *Seeking meaning: A process approach to library and information services*. Norwood, NJ: Ablex Publishing Corporation.
- Kuhlthau, C. C. 1999, February/March. Accommodating the user's information search process: Challenges for information retrieval system designers. *Bulletin of the American Society for Information Science*: 12-15.
- Lucas, W., & Topi, H. 2002. Form and function: The impact of query term and operator usage on Web search results. *Journal of the American Society for Information Science and Technology*, 53(2): 95-108
- Michell, G., & Dewdney, P. 1998. Mental models theory: Applications for library and information science. *Journal of Education for Library and Information Science*, 39: 275-281.
- Pollock, A., & Hockley, A. 1996. What's wrong with Internet searching. *D-Lib Magazine*, 3, 1-5 Internet. [cited 2003. 7.7.]
<<http://www.dlib.org/march97/bt/03pollock.html>.>
- Rodriguez, L. 2000. Uncertain relevance judgments: A quantitative and qualitative study. *Great Lakes Conference* [October 28, 2000]
- Rousseau, G. K., Jamieson, B.A., Rogers, W. A., Mead, S. E., & Sit, R. A. 1998. Assessing the usability of on-line library systems. *Behavior and Information Technology* 17: 274-281.
- Shannon, C. E., & Weaver, W. 1949. *The mathematical theory of communication*. Urbana, IL: University of Illinois Press.
- Small, S. I., Cottrell, G. W., & Tanenhaus, M. K. (Eds.). 1988. *Lexical ambiguity resolution: Perspective from psycholinguistics, neuropsychology & artificial intelligence*. San Mateo, CA:

- Morgan Kaufmann Publishers, Inc.
- Spink, A. 1997. Partial relevance judgments during interactive information retrieval: An exploratory study. *Proceedings of the 60th ASIS Annual Meeting, 34*: 111-122.
- Spink, A., & Greisdorf, H. 1997. Users' partial relevance judgments during online searching. *Online and CD-ROM Review, 21*: 271-279.
- Spink, A., Greisdorf, H., & Bateman, J. 1998. From highly relevant to not relevant: Examining different regions of relevance. *Information Processing and Management, 34*: 599-621.
- Stevens, N. D. 1988. The importance of the verb in the reference questions. *The Reference Librarian, 22*: 241-244.
- Su, L. T. 2003b. A comprehensive and systematic model of user evaluation of Web search engines: II. An evaluation by undergraduates. *Journal of the American Society for Information Science and Technology, 54*(13): 1193-1223.
- Taylor, R. S. 1968. Questions-negotiation and information seeking in libraries. *College and Research Libraries, 29*: 178-194.
- Wang, P., Berry, M. W. & Yang, Y. 2003. Mining longitudinal Web queries: Trends and Patterns. *Journal of the American Society for Information Science and Technology, 54*(8): 743-758.
- Wang, P., Hawk, W. B., & Tenopir, C. 2000. Users' interaction with World Wide Web resources: An exploratory study using a holistic approach *Information Processing & Management, 36*(2): 229-251.
- Whittemore, B., & Yovits, M. C. 1973. A generalized conceptual development for the analysis and flow of information. *Journal of the*

*American Society for
Information Science, 24:*
221-231.

Wilson, T. D., Ellis, D., & Ford, N.
2000. *Uncertainty in
information seeking: A
research project in the
Department of Information
Studies, University of
Sheffield.* Sheffield, UK:
Department of Information
Studies, University of
Sheffield. [cited 2003. 7.7.]
<[http://www.shef.ac.uk/~is
/research/imrg/uncerty.html](http://www.shef.ac.uk/~is/research/imrg/uncerty.html)>

Wilson, T. D., Ellis, D., Ford, N.,
& Foster, A. 1999.
*Uncertainty in information
seeking. Final report to the
British Library Research and
Innovation Centre/Library
and Information Commission*

*on a research project carried
out at the Department of
Information Studies,
University of Sheffield.*
(Library and Information
Commission Research
Report 59). Sheffield, UK:
Department of Information
Studies, University of
Sheffield. [cited 2003. 7.7.]
<[http://www.shef.ac.uk/~is/
publications/unis/.](http://www.shef.ac.uk/~is/publications/unis/)>

Wilson, T.D., Ford, N., Ellis, D.,
Foster, A., & Spink, A.
2002. Information seeking
and mediated searching.
Part 2. Uncertainty and its
correlates. *Journal of the
American Society for
Information Science and
Technology, 53(9):*
704-715.