

User Perceptions of Uncertainty in the Selection of Information Retrieval System: Implications for System and Service Improvement

Yang-woo Kim*

Division of Knowledge & Information, Hansung University,
389 Samsun-dong 3-ga, Sungbuk-gu, Seoul,
Republic of Korea

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 information retrieval system in the real information-seeking process. Considering the role of commercial Web search engines as supplementary tools for traditional bibliographic databases in academic research environments, this study analyzed the selection behavior of scholarly researchers, who use such search tools for their academic study. The researchers 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. Variations appeared in different incidents of uncertainty relating to the unique characteristics of the subjects' information-seeking context. The identification of three principal origins of uncertainty based on the different types of uncertainty generated implications to improve information systems and services.

Keywords: Information Seeking Behavior; Uncertainty; User Study; Information Retrieval System; Information Service

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 using a real system.

This study investigated user perceptions of uncertainty in the process of the selection of information retrieval system 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 system 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 generated implications to improve information retrieval systems and services.

2. BACKGROUND

2.1 Uncertainty Study

A substantial number of studies from the relevant research community 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: *information as a means of reducing uncertainty* [1], [2], and *exceptions in uncertainty reduction with information provided* [3]. 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 [4] and articulating an information need [5]; a need for information-

* Corresponding author. E-mail : ykim@hansung.ac.kr

Manuscript received Apr. 24, 2009 ; accepted Jun. 22, 2009

This research was financially supported by Hansung University in the year of 2008.

seeking with a knowledge-gap [6]; and poly-representative nature of the cognitive space [7], [8];
2) connecting the internal and expressed needs of users — the necessity of dealing with the discrepancies between the two needs, further developing Taylor's [4] approach [9]-[12];
3) uncertainty for system design — uncertainty as a system design principle [13]; poly-representative nature of the information space and uncertainly, unpredictability and contextualization in IR [7, 8]; types of uncertainty related to the process of indexing and matching [14]; and the need to accommodate uncertainty in systems design [15]; and
4) relevance judgment — difficulty in determining the usefulness of documents [16], particularly as a concrete binary judgment [17]-[19].

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 [20], [21];
- 2) the role of information-seeking in uncertainty resolution during the problem-solving process [22], [23]; and
- 3) correlations between Kuhlthau's feelings dimensions and uncertainty at different problem stages [24].

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 retrieval system. This is particularly true in examining uncertainty for the characteristics of the information retrieval system selected, either as an information search tool (i.e., system features such as search or display features), or as an information resource (i.e., quantity and quality of stored information).

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 [25]; user *mistakes* on Web query inputs [26]; and the *limited capacity* of inexperienced searchers for Web search term selection [27];
- 2) search term selection and system features — *differences* in features among search engines for query formulation [28]; *problems* in submitting queries and *difficulties* in using Boolean operations for an academic Web site [29]; and the association of Boolean searching on the Web with user *anxiety* [30];
- 3) system features — *confusion* in using Web search engine features [31];
- 4) results evaluation — *difficulty* and *confusion* in comprehending retrieved resources of library sites [32]; and
- 5) the extensive process of searching — user *problems* and system *problems* in working with online library systems [33]; sources of user *frustration* with the Web search and users'

suggestions for interface improvements [34]; and types of *problems* and *anxieties* during the Web search [35]. 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 do not provide in-depth discussion on uncertainty in user selection of information retrieval systems. 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 Web-based information system.

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 the use of Web search engines as the main search tools to locate various World Wide Web sites. The scope of investigation does not include the selection of search engines only as a means to locate another Web-based search system¹ that the subjects might regularly use while also using the located search system as the main search tool.

3.2 Collection and Analysis of Data

This study investigated the users' information search process within a natural workplace setting. Twenty subjects from a university in the northeastern state of U.S. 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.

A number of coding methods has been adopted in analyzing the data including open, axial and selective coding [36], [37]. Overall, a close examination of data identified dimensions, types, and incidents of uncertainty (open coding). In other

¹Online journal databases or OPACs are examples of such tools.

words, open coding pertains specifically to naming and categorizing of phenomena through such examination of data. Through the process, axial and selective coding were used in a supplementary manner: linking categories to their subcategories (axial coding) and at a higher level of abstraction, elaborating the core categories by grouping and integrating developed categories.

3.3 Homogeneity of Sample

This research consists of 15 case studies with in-depth questionnaire and interviewing conducted for each study. As generally known, the sample size is of less importance in a qualitative approach as these studies. An important factor that strengthened the generalizability of the research findings was the homogeneity of the sample. As discussed, these case studies deal with homogenous individuals, who are scholarly researchers in the field of science who use Web search engines for academic research purpose. The homogeneity of a sample provides a greater generalizing power to the research findings.

4. RESULTS

Analysis of the data identified three dimensions related to the information retrieval system and eight types of uncertainty within a total of 31 incidents. The types of uncertainty assigned to different dimensions are shown in Table 1, with the number of incidents of uncertainty in each type and one or two examples of the incidents per type. Table 1 also shows subjects' descriptions of each exemplary incident. In addition, the definition of each type and a list of all incidents of uncertainty are shown in Tables 2 and 3, respectively.

The dimensions of uncertainty identified in selecting an information retrieval system reflected two different aspects of search engines:

- 1) Web search engines as search tools, and
- 2) WWW sites as resources for the search systems.

The three dimensions identified were:

- *search tool* — *Web search engine*,
- *search tool and resource*, and
- *resource* — *World Wide Web sites*.

The first dimension, *search tool – Web search engine*, related to the different functions and performances of Web search engines as information search devices, not to the characteristics of WWW sites as information resources. Incidents of uncertainty within this dimension originated from the subjects not being sure of the best choice of a search engine. That lack of certainty was explained only by the different retrieval mechanisms of search engines, not by the characteristics of resources stored in the World Wide Web.

As a database, Web search engines including Google and Yahoo, which were selected by the subjects, have the WWW as an identical resource. If a difference exists on WWW sites crawled by the spider software of different search engines, that difference is attributed to the crawling mechanism which the search engine itself adopted, while the crawling mechanism is a component that characterizes the search engine as a search tool. Each search engine stores a subset of URLs on the WWW, and only these subsets are likely to differ according to the search engines' different crawling mechanisms. However, users frequently misunderstand this aspect of search engines.

As presented, the type of uncertainty *choice of search engines* was identified within this dimension (see Tables 1, 2, and 3). This uncertainty type involved determining the best search engine, as compared with other engines (Table 2). The three incidents of uncertainty categorized into this type were *choice of the best search engine* (S004, S005, and S015) (Table 3). For example, Subject 015 explicitly indicated uncertainty in *choice of the best search engines* by the following response: "Is Yahoo better than Google in terms of getting more relevant sites?" (see Table 1).

Two related factors could influence users' perceptions within this type:

- 1) appropriateness of a search engine for their information needs, and
- 2) variations on user skills in working with different search engines.

Yet, in either case, the user perceptions related only to the functions and performances of different search engines, not to the characteristics of WWW sites.

The second dimension was *search tool & resource*, which encompasses user uncertainty that can be explained either by an attribute of a search engine as a search tool or by a characteristic of WWW sites as an information resource. Uncertainty in this dimension can be explained by either:

- 1) characteristics of Web search engines as search tools, as compared with those of other search tools, i.e., academic journal databases, or
- 2) characteristics of WWW sites as an information resource, as compared with those of other resources, i.e., resources stored in academic journal databases or printed resources.

For example, Subject 010's incident originated from the system's mistreatment of ambiguous and deceptive labeling presented by a site provider (search engine as a search tool) and an excessive number of commercial sites on the WWW (Web sites as resources of the information system selected). The subject indicated: "I've noticed a lot the pages will bring up either they'll be mislabeled, generally I call those pages traps...trying to bring people in there to sell something."

Table 1. Types of Uncertainty in the Selection of Information Systems

Dimension-Uncertainty	Types of Uncertainty	No. of Incidents	Example of Incident	Subjects' Descriptions (Subject # – Instrument)
Search tool –Web search engine	Choice of search engines	3	Choice of the best search engine	“Is <u>Yahoo better than Google</u> in terms of getting more relevant sites?” (S015 – Q)
Search tool & Resource	Search output	2	Number of (relevant) returns	“I may get <u>too many results or nothing.</u> ” (S001 – Q)
	Search (& Display) mechanism	7	Amount of information searchable – ranking of relevant results	“Not sure how much information would be obtained.” (S008 – Q) “World Wide Web does give sometimes ten thousand, several million results. <u>If the most relevant result comes very late</u> in the, in the sequence, then <u>I don't have a tendency to look until the end.</u> ” (S008 – I)
			Relevance of items retrieved – coping with (commercial) sites with ambiguous/deceptive labeling	“I've noticed a lot the pages will bring up either they'll be <u>mislabeled</u> , generally I call those pages traps... <u>trying to bring people in there to sell something</u> ... if you haven't been able to find things using the long names, you'll start using the abbreviations, like chemical symbols and it sends you into irrelevant sites.” (S010 – I)
	Search skill	3	Sorting out sites effectively – coping with excessive number of returns	“Many sites can be returned; how can I effectively sort out what I need?” (S007 – Q)
	Search engine vs. other resources	3	Using a search engine – compared to print resource	“Using <u>Yahoo vs. books/journal</u> (review) articles.” (S004 – Q)
Resource – World Wide Web sites	Availability of information	9	Availability of specific resources on the Web	“I'm <u>not sure if all works on thermal control have been published on the Web.</u> ” (S006 – Q); “People who are not related to computers normally don't write the document and put it on the Web. <u>They don't normally have a web page for every researcher...</u> There are some <u>copyright restrictions sometimes.</u> ” (S006 – I)
	Characteristics of information available	2	Technicality/specificity of information available	“ <u>Not sure there is very technical information on the Web, specific information.</u> ” (S001 – Q) “I am expecting more general and broad information...” (S001 – I)
	Quality of information source	3	Source credibility	“ <u>Not sure about credibility</u> for the information.” (S001 – Q) “If I use [an] academic database, all data are from already published papers...proved, also accurate...but <u>on the Web any individuals can put information on it.</u> ” (S001 – I)
Total		31		

Table 2. Uncertainty in the Selection of Information Systems: Definition of Types

Dimension I. Search tool –Web search engine

Types of Uncertainty	Definitions
Choice of search engines	Determining the best search engine compared to other engines

Dimension II. Search tool & resource

Types of Uncertainty	Definitions
Search output	Yielding an ideal quantity of results or relevant results

Search (& Display) mechanism	Retrieval (and display) mechanism of search engines that provide search results to users
Search skills	Subjects' self-perceived skills at using search engines and/or search results; related to either use of search systems or domain knowledge on the search topic
Search engine vs. other resources	Using search engine compared to using other resources such as print resource or scientific databases

Dimension III. World Wide Web sites as an information resource

Types of Uncertainty	Definitions
Availability of information	Availability of specific resource on the WWW
Characteristics of information available	Attributes of information available on the WWW
Quality of information source	Degree to which source of information is credible and information is accurate

Note. Number of incidents for each type is in parentheses.

Subject number for each incident is in parentheses.

Table 3. Uncertainty in the Selection of Information Systems: Incidents for Each Type

Dimension I. Search tool – Web search engine

<p>Type – Choice of search engines (3) Choice of the best search engine (S004) Choice of the best search engine (S005) Choice of the best search engine (S015)</p>
--

Dimension II. Search tool & resource

<p>Type – Search output (2) Number of (relevant) results (S001) Number of (relevant) results (S002)</p>	<p>Type – Search (& Display) mechanism (7) Relevance of items retrieved – resource type (S002) Relevance of items retrieved – topicality (S002) Amount of information searchable – ranking of relevant results (S008) Relevance of items retrieved – avoiding commercial sites with mislabeling (S010) Relevance of items retrieved – copying with ambiguous/deceptive labeling (S010) Accessibility to searched links (S010) Matching between user intention and system's interpretation (S012)</p>
<p>Type – Search skills (3) Sorting out sites effectively – coping with excessive number of returns (S007) Sorting out sites with clarity – coping with difficult terminologies (S007) Locating all references – coping with limited search skills for research topic (S011)</p>	<p>Type – Search engine vs. other resources (3) Using a search engine – compared to using print resource (S004) Using a search engine – compared to using print resource (S005) Using a search engine – compared to scientific databases (S004)</p>

Dimension III. World Wide Web sites as an information resource

<p>Type – Availability of information (9) Availability of additional information beyond other database search (S001) Availability of specific resource on the Web – seminal papers (S003) Availability of specific resource on the Web – open problems (S003) Availability of specific resource on the Web – technical report of leading scholars (S003) Amount of information available on research topic (S003) Availability of specific resources on the Web – topicality (S005) Availability of specific resources on the Web – unavailability of Home page from researchers; copyright restriction (S006) Amount of information available (S008) Availability of all necessary references on the Web (S011)</p>	
<p>Type – Characteristics of information available(2) Technicality/specificity of information available (S001) Specificity of information available (S008)</p>	<p>Type – Quality of information source (3) Source credibility (S001) Source credibility (S007)</p>

	Accuracy of information available (S002)
--	--

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

The four different types of uncertainty identified within the second dimension were:

- *search output*;
- *search (& display) mechanism*;
- *search skill*; and
- *search engine vs. other resources* (see Tables 1, 2, and 3).

As indicated, the above types were assigned to the second dimension because the subjects' responses within these types related to certain characteristics of search engines and of WWW sites.

The third dimension, *resource – World Wide Web sites*, related only to the characteristics of WWW sites as an information resource of search engines. User uncertainty assigned to this dimension was not explained by specific attributes of search engines as search tools. The three different types of uncertainty identified within this dimension were:

- *availability of information*;
- *characteristics of information available*; and
- *quality of information source* (see Tables 1, 2, and 3).

The above types were assigned to this dimension because they were concerned with only certain attributes of Web resources. Different aspects related to the resources (i.e., availability and quality of information) constituted the identification of these types.

For example, Subject 006 indicated uncertainty about the *availability of specific resources on the Web* by the following response:

"I'm not sure if all works on thermal control have been published on the Web.... People who are not related to computers normally don't write the document and put it on the Web, they don't normally have a web page for every researcher.... There are some copyright restrictions sometimes." (see Table 1)

This subject did not question specific functions of the search engines to describe this incident. The subject's response for the above incident was only concerned with the availability of necessary resources on the Web, not the process of locating such resources using a search system.

5. DISCUSSION

5.1 Uncertainty and Information-Seeking Context

The analysis of uncertainty in the process of information-seeking provides insights into individual's information-seeking contexts. The incidents of uncertainty identified in the system selection phase reflected the subjects' perceptions as mostly related to the information retrieval systems used and their resources. Accordingly, the incidents of uncertainty identified

in this phase facilitated an understanding of the selected Web search engines and WWW sites as information resources.

Four different incidents of uncertainty identified in this study more explicitly differentiated the system used from other information retrieval systems (see Table 4). The first two incidents are from the type of uncertainty *search (& display) mechanism*, which concerns the retrieval and display mechanism of search engines. For example, one incident of uncertainty, *relevancy of items retrieved – coping with commercial sites with ambiguous/deceptive labeling* (Subject 010), can be explained by the limitations of the mechanisms of storing and retrieving the World Wide Web resource. The retrieval of items, which relies on the occurrence of a search term in full-text documents, can generate a major drawback to many search systems. It can possibly cause end-users' uncertainty about the systems perform retrieval. The retrieval of irrelevant sites based on the search terms, which are associated with chemical symbols in this particular incident (Subject 010), is an example of such a drawback. Yet, ambiguous or deceptive labeling presented by site providers, particularly in commercial sites, characterizes this uncertainty, thereby differentiating it from similar incidents caused by other retrieval systems, i.e., journal databases which are a major type of resource for scholarly researchers.

The last two incidents, shown in Table 4, are from the type of uncertainty, *quality of information source*. The incident of uncertainty, *source credibility*, related to the characteristics of the WWW as an information resource included:

- insufficient quality control mechanisms for WWW resources (Subject 001), and
- insufficient organization of bibliographic information in the WWW (Subject 007).

Again, such shortcomings would be less significant when researchers use scholarly journal databases, which are major information sources for the user group. Even journal databases would have different levels of reliability. Nevertheless, in general, the search results of journal databases, particularly those used by scholarly researchers, would have a higher level of reliability compared with the search output of a search engine. The latter could include any Web-based resources provided by heterogeneous individuals.

The incidents of uncertainty identified in this study provided insights into the users' information-seeking context of the information retrieval system and its selected resources. The findings revealed variations in how different incidents of uncertainty related to the unique characteristics of the information-seeking context. As indicated, some incidents more explicitly differentiated the stem used in this study from other information retrieval systems.

Table 4. Incidents of Uncertainty and the Information-Seeking Context

Uncertainty		Information seeking context
Exemplary Incidents	Subjects' Descriptions (Subject # – Instrument)	Related Characteristics of Web Search Engines and WWW Resources
Amount of information searchable – ranking of relevant results	“ <u>Not sure how much information would be obtained.</u> ” (S008 – Q) “ <u>World Wide Web does give sometimes ten thousand, several million results. If the most relevant result comes very late in the, in the sequence, then I don't have a tendency to look until the end.</u> ” (S008 – I)	Excessive volume of resources in WWW
Relevance of items retrieved – copying with (commercial) sites with ambiguous/deceptive labeling	“ <u>I've noticed a lot the pages will bring up either they'll be mislabeled, generally I call those pages traps...trying to bring people in there to sell something...</u> if you haven't been able to find things using the long names, you'll start <u>using</u> the abbreviations, like <u>chemical symbols</u> and it sends you into irrelevant sites.” (S010 – I)	Limited mechanism in storing and retrieving WWW resources – ambiguous/deceptive labeling in commercial sites: the system's mistreatment of ambiguous/deceptive labeling an excessive number of commercial sites
Source credibility	“I am <u>not sure about credibility</u> for [of] the <u>information.</u> ” (S001 – Q) “If I use a formal academic database, all data are already published papers... proved, also accurate, published in good journal, <u>but in the Web any individuals can put information on it.</u> So, I do not know whether information is accurate or not in the Web environment.” (S001 – I)	Insufficient quality control mechanism for WWW resources – source credibility
Source credibility	“Not sure how to sort out sites with sufficient degree of clarity.” (S007 – Q) “ <u>Sometimes, I could not identify the sources of the whole site or source of information on a site.</u> ” (S007 – I)	Insufficient organization of bibliographic information on the WWW

5.2 Origin of Uncertainty and Its Implications

This section discusses three principal origins of uncertainty and the implications of such findings on improving information retrieval systems and services. The information retrieval 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.

The findings revealed the following principal origins of user uncertainty. The user-side included *limited knowledge of the information need or topical area* and *limited knowledge of the functions and/or features of the information retrieval 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 three origins. To some extent, the above origins were not completely comprehensive. Yet, identifying the three principal origins was important as it led to practical ideas that generated the implications of this study.

The first origin of uncertainty on the user-side concerned *limited knowledge of the information need or topical area*. The type of uncertainty related to this origin included *availability of information* within the dimension of *Resource – World Wide Web sites* (see Table 5). In this type of uncertainty, the subjects showed limited knowledge of availability of related resources in the World Wide Web. This

origin of uncertainty suggested ideas to provide appropriate information services to users with limited knowledge of the above areas. Possible interventions may include suggesting a specific method of searching to locate appropriate resources or alternative methods to search for different resources.

The second origin of uncertainty on the user-side concerned *limited knowledge of the functions and/or features of the information retrieval system*. The type of uncertainty related to this origin was *choice of search engine* (see Table 5). The subjects showed limited knowledge of the following area in this type of uncertainty: different functions and features of Web search engines to determine the best search engine for the information search. This origin of uncertainty suggested the need to improve information retrieval 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 concerned *limited retrieval mechanisms for utilizing the characteristics of the resource*. The types of uncertainty related to this origin included: *search (& display) mechanism, characteristics of information available, and quality of information source* (see Table 5). 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 limited aspects of the resource characteristics during the retrieval process. Such limitations influenced the users' uncertainty.

Table 5. Origins of Uncertainty and Types of Uncertainty Related

Origins of Uncertainty		Types of Uncertainty Related
User-side	Limited knowledge of an information need or a topical area	Availability of information
	Limited knowledge of the functions and/or features of information systems	Choice of search engine
System-side	Limited retrieval mechanisms for utilizing the characteristics of resources	Search (& display) mechanism; Characteristics of information available; Quality of information source

retrieval 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 retrieval 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.

Web search engines have the following limitations within each type of uncertainty for the utilization of resource characteristics: retrieval of irrelevant sites based on the occurrence of search terms in the text of sites within the type of uncertainty, *search (& display) mechanism*; limited mechanisms for indicating technicality and specificity of information searched within the type of uncertainty, *characteristics of information available*; and limited mechanisms for source credibility and accuracy of information searched within the type of uncertainty, *quality of information source*.

The limitations mentioned for this origin related to the restricted use of resource characteristics during the retrieval process. For instance, as an example of *search mechanism*, if the system had further utilized the related characteristics of resources, the retrieval of irrelevant sites, based on the appearance 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.

Improving information retrieval systems in the following aspects would facilitate users' search process: retrieval of irrelevant sites relying on the occurrence of search terms (*search [& display] mechanism*); mechanisms for indicating technicality and specificity of searched information (*characteristics of information available*); and mechanisms for source credibility or accuracy of searched information (*quality of information source*). Many of these limitations of information

6. CONCLUSION

Analyzing uncertainty in this study was useful for generating implications of the research findings to improve information retrieval systems and services. One meaningful attempt made for further research would be to apply a similar approach to different information-seeking contexts. Other related components in addition to 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.

7. REFERENCES

- [1] S. Artandi, "Information concepts and their utility," *Journal of the American Society for Information Science*, vol. 24, 1973, pp. 242-245.
- [2] C.E. Shannon and W. Weaver, The mathematical theory of communication, *University of Illinois Press, Urbana, IL*, 1949.
- [3] B. Whittemore and M.C. Yovits, "A generalized conceptual development for the analysis and flow of information," *Journal of the American Society for Information Science*, vol. 24, 1973, pp. 221-231.
- [4] S. Taylor, "Questions-negotiation and information seeking in libraries," *College and Research Libraries*, vol. 29, 1968, pp. 178-194.
- [5] N.J. Belkin, "Anomalous states of knowledge as the basis for information retrieval," *Canadian Journal of Information Science*, vol. 5, 1980, pp. 133-143.
- [6] B. Dervin, An overview of sense making research: Concepts, methods, and results to date, *School of Communication, University of Washington, Seattle, WA*, 1983.
- [7] P. Ingwersen, Information retrieval interaction, *Taylor Graham, London*, 1992.
- [8] P. Ingwersen, "Cognitive perspective of information retrieval interaction: Elements of a cognitive IR theory," *Journal of Documentation*, vol. 52, no. 1, 1996, pp. 3-50.
- [9] D. Gerhan, "When quantitative analysis lies behind a reference question," *Reference and User Services Quarterly*, vol. 39, no. 2, 1999, pp. 166-185.
- [10] P. Ingwersen, "Search procedures in the library analyzed for the cognitive point of view," *Journal of Documentation*, vol. 38, no. 3, 1982, pp. 165-191.
- [11] G. Michell and P. Dewdney, "Mental models theory: Applications for library and information science," *Journal of Education for Library and Information Science*, vol. 39, 1998, pp. 275-281.
- [12] N.D. Stevens, "The importance of the verb in the reference questions," *The Reference Librarian*, vol. 22, 1988, pp. 241-244.
- [13] M. Bates, "Subject access to online catalogs: A design model," *Journal of the American Society for Information Science*, vol. 37, 1986, pp. 357-376.
- [14] C.J. van Rijsbergen, "Information, logic, and uncertainty in information science," *Proc. CoLIS 2 – Second International Conference on Conceptions of Library and Information Science: Integration in Perspective*, 1996, p. 1-10.
- [15] C.C. Kuhlthau, "Accommodating the user's information search process: Challenges for information retrieval system designers," *Bulletin of the American Society for Information Science*, Feb/Mar. 1999, pp. 12-15.
- [16] L. Rodriguez, "Uncertain relevance judgments: A quantitative and qualitative study," *Proc. Great Lakes Conference*, 2000, p.17-25.
- [17] A. Spink, "Partial relevance judgments during interactive information retrieval: An exploratory study," *Proc. The 60th ASIS Annual Meeting*, 1997, pp. 111-122.
- [18] A. Spink and H. Greisdorf, "Users' partial relevance judgments during online searching," *Online and CD-ROM Review*, vol. 21, 1997, pp. 271-279.
- [19] A. Spink, H. Greisdorf and J. Bateman, "From highly relevant to not relevant: Examining different regions of relevance," *Information Processing and Management*, vol. 34, 1998, pp. 599-621.
- [20] C. C. Kuhlthau, "A principle of uncertainty for information seeking," *Journal of Documentation*, vol. 49, no. 4, 1993, pp. 339-355.
- [21] C. C. Kuhlthau, Seeking meaning: A process approach to library and information services, *Ablex Publishing Corporation, Norwood, NJ*, 1993.
- [22] T. D. Wilson, D. Ellis and N. Ford, Uncertainty in information seeking: A research project in the Department of Information Studies, *University of Sheffield, Sheffield, UK*, 2000, Available at: <http://www.shef.ac.uk/~is/research/imrg/uncerty.html>, accessed 28 Nov. 2008.
- [23] T. D. Wilson, D. Ellis, N. Ford and A. Foster, "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," *Library and Information Commission Research Report* vol. 59, 1999, Available at: <http://www.shef.ac.uk/~is/publications/unis/>, accessed 28 Nov. 2008.
- [24] T. D. Wilson, N. Ford, D. Ellis, A. Foster and A. Spink, "Information seeking and mediated searching. Part 2. Uncertainty and its correlates," *Journal of the American Society for Information Science and Technology*, vol. 53, no. 9, 2002, pp. 704-715.
- [25] A. Pollock & A. Hockley, "What's wrong with Internet searching," *D-Lib Magazine*, vol. 3, 1996, 1-5, Available at: <http://www.dlib.org/march97/bt/03pollock.html>, accessed 28 Mar. 2009.
- [26] B. J. Jansen, A. Spink and T. Saracevic, "Real life, real users, and real needs: A study and analysis of user queries on the web," *Information Processing and Management*, vol. 36, no. 2, 2000, pp. 207-227.
- [27] W. Lucas and H. Topi, "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*, vol. 53, no. 2, 2002, pp. 95-108.
- [28] M. Gordon and P. Pathak, "Finding information on the World Wide Web: The retrieval effectiveness of search engines," *Information Processing and Management*, vol. 35, 1999, pp. 141-180.
- [29] P. Wang, M. W. Berry and Y. Yang, "Mining longitudinal Web queries: Trends and Patterns," *Journal of the American Society for Information Science and Technology*, vol. 54, no. 8, 2003, pp. 743-758.
- [30] N. Ford, D. Miller and N. Moss, "Web search strategies and approaches to studying," *Journal of the American Society for Information Science*, vol. 54, no. 6, 2003, pp. 473-489.
- [31] L. T. Su, "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*, vol. 54, no. 13, 2003, pp. 1193-1223.
- [32] M. A. Fitzgerald and C. Galloway, “Relevance judging, evaluation, and decision making in virtual libraries: A descriptive study,” *Journal of the American Society for Information Science and Technology*, vol. 52, no. 12, 2001, pp. 989-1010.
- [33] G. K. Rousseau, B.A. Jamieson, W. A. Rogers, S. E. Mead and R. A. Sit, “Assessing the usability of on-line library systems,” *Behavior and Information Technology*, vol. 17, 1998, pp. 274-281.
- [34] R. Fidel, R. K. Davies, M. H. Douglass, J. K. Holder, C. J. Hopkins and E.J. Kushner, “A visit to the information mall: Web searching behavior of high school students,” *Journal of the American Society for Information Science*, vol. 50, no. 1, 1999, pp. 24-37.
- [35] P. Wang, W. B. Hawk and C. Tenopir, “Users’ interaction with World Wide Web resources: An exploratory study using a holistic approach,” *Information Processing & Management*, vol. 36, no. 2, 2000, pp. 229-251.
- [36] A. Strauss and J. Corbin, Basics of qualitative research: Grounded theory procedures and techniques, *SAGE Publications, Newbury Park, CA*, 1990.
- [37] A. Strauss and J. Corbin, Basics of qualitative research: Techniques and procedures for developing grounded theory (2nd ed.), *SAGE Publications, Newbury Park, CA*, 1998.



Yang-woo Kim

He received the B.S. in Business Management from University of Maryland, Asian Division, Korea in 1991 and M.S. in Library and Information Science from Emporia State University, USA in 1998. He also received Ph.D. in information science

from Rutgers University, USA in 2004. Since 2005, he has been with Division of Knowledge and Information, Hansung University. His main research interests include information-seeking behavior and contents development.