A Web Search Model for Strategic Decision Making

Shailaja Venkatsubramanyan
Stephen K. Kwan
Department of Management Information Systems
College of Business, San Jose State University
San Jose, California, USA
(venkat_s@cob.sjsu.edu, kwan_s@cob.sjsu.edu)

DOI: 10.26573/2008.2.3.3
Volume 2, *Number* 3
September 2008, *pp*. 197-214

Search services are now ubiquitously employed in satisfying the information needs of managers and business analysts involved in strategic decision making. In this paper, we propose a model of a user's interaction with a search service in satisfying information needs and empirically evaluate the principal factors involved. Findings indicate that the information need type influences the search process more significantly than the specific search service being used. Consequently, managers and business analysts should pay particular attention to the types of information needs involved in a strategic decision.

Keywords: Search Engine, Information Need, Web Search Process

1. Introduction

With the growth of the World Wide Web, one technology that has become ubiquitous and indispensable is Web search. Search services are now widely employed in searching for information on the Internet and on enterprise intranets. Furthermore, search services are popular with strategic decision makers like managers and business analysts who use search services to satisfy information needs involved in a critical decision – examples include employee background searching and product sentiment analysis [2, 9, and 16]. Such a typical user would use these search services by entering queries appropriate to their information need, and perusing the most appropriate search results corresponding to their queries.

Prior studies have focused on information seeking from the user's perspective. These studies have tried to understand user's sense-making process, and the cognitive and affective aspects of the process of information seeking with the objective of having the information system reflecting those aspects [16]. A typical information search process is composed of the tasks of initiation of information need, selection of topic to be investigated, exploration of feelings of confusion, formulation of a sense of clarity, collection of information, and presentation or use of findings of search. For the purpose of our research, we assume that the search process begins when a user is faced with a decision problem that may consist of multiple information needs. For each information need, the user formulates a query and submits it to a search service to obtain search results. The user filters the search results to look for information relevant to the decision problem. Based on the filtering, the user may reformulate or refine the query and submit it to the same or a different search service, or abandon the search (See Figure 1). The goal of this research is to understand the principal factors that could be involved in the various

stages of the search process. From Figure 1, the three main sub-processes in the search process include formulating a query, running the query through a search service, and filtering search results. We conducted an experiment; collected data about the various factors discussed above, applied statistical analysis to analyze the data, and identified the principal factors that affect the search process. We used three popular search services – Google, Yahoo and MSN for our experimental analysis.

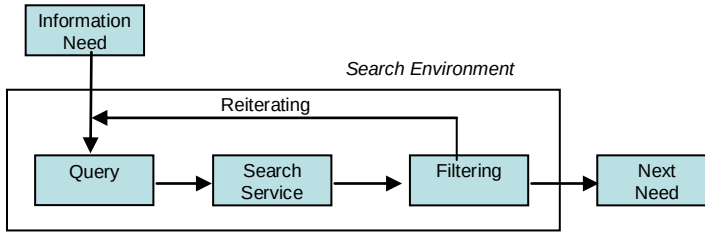


Figure 1 *Information Search Process*

The rest of the paper is organized as follows: Section 2 presents a model that serves as the theoretical framework for this research. Section 3 presents the experimental design and the results from the test, while conclusions and implications for managers and business analysts involved in strategic decision making are presented in Section 4.

2. Research Model

To get a deeper understanding of how users translate their information need into a web search, we propose a model (See Figure 2) based on the theoretical constructs developed by Kuhlthau [16] to capture the various factors at play. The various pieces in the model are outlined below:

(a) *Sub-processes*: The search process consists of three main sub-processes – query formulation, use of the search service, and filtering of search results. The process of refining a query is broad in scope and is difficult to objectively separate from a new search query and is treated as a new invocation of the search process based on statistical evidence [13, 19].

(b) *Factors*: Each sub-process could be influenced by several factors such as (i) the user profile, (ii) the information need type, (iii) the search service characteristics and (iv) the web search environment. Some of those factors could have sub-components. For example, user domain knowledge and user search experience are sub-components of the user profile factor [13].

(c) *Outputs*: The output of each sub-process of the search process is also distinct: the query formulation, use of the search service and search results filtering yield an input query, search results corresponding to the query and filtered search results respectively. The last output is instrumental in determining whether the information need is satisfied so that the user can refine the query or abandon the search process. Each of the search process outputs can be measured using surrogate measures such as query complexity, search result precision and ease of filtering. It is important to note that the output of each search sub-process is influenced by the output of the previous search sub-processes.

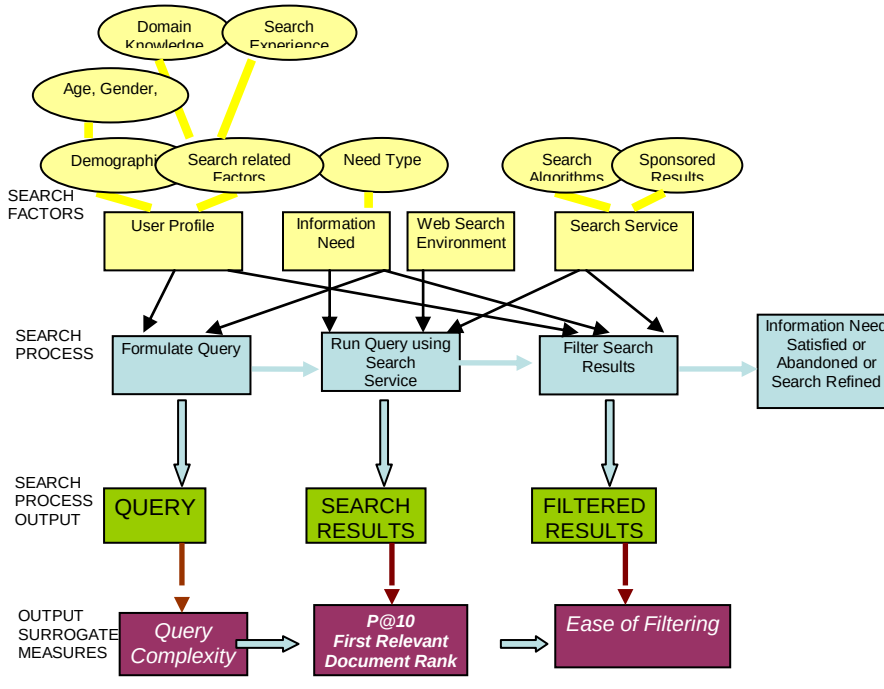


Figure 2 *Web Search Model*

Section 2.1 describes the search process with its various sub-processes and the output produced at each sub-process. Section 2.2 describes the main factors at play in the search process.

2.1 Search Process

The search process in the research model is broken down into three broad sub-processes [16] and we formulate the factors that affect these processes based on past literature [12, 19, and 27]. Those sub-processes are described below:

Formulate Query: The user faces a decision problem and needs information to help with the decision. Typically, a decision problem involves multiple information needs and the user proceeds to resolve these needs based on some strategy.

The output of this sub-process in the search process is a query formulated by the user. The query may consist of one or more keywords, or an advanced query consisting of keywords as well as operators such as "+", "-", or quotation marks. A primary statistical proxy for a user query is the notion of query complexity [19], where the query complexity is defined in terms of the number of words in a user query and the number of complex operators used in the query. The query complexity is influenced by the characteristics of the user such as prior knowledge of the decision domain, the information need type and experience using search services [19]. It is expected that given the same information need, different users will formulate queries with varying degree of complexity that produces different results of varying quantity and quality under the influence of the factors listed above. The

correctness and fineness of the query have direct effects on the quality of returned results from the search service.

Run Query Using a Search Service: Once the user's query is submitted to the search service, the search service is deployed to process the query, execute the underlying algorithms, and return results to the user.

The main factors affecting this sub-process are the general web search environment, information need type, the type of search service, and the complexity of the query formulated by the user in the previous search sub-process [13].

Each search service can be characterized by how it spiders the Web, how often it spiders the web, its indexing algorithm, its internal organization, and its ranking method. A review of past literature reveals that different search services could produce different results of varying quantity and quality for the same query [12].

The output of this search sub-process is a collection of search results. Measures such as precision and recall of relevant documents, stability of a search service over time, and correlation between human and service ranking are popularly used to evaluate the quality of a search service [5, 12, and 26]. In this research, we have used two measures – Precision at 10 or P@10 and the rank of the first relevant document (FRDR) to measure search service quality. The reason for including FRDR is that the traditional measures of precision and recall may not be appropriate for evaluating tasks that need high accuracy [23]. For instance, when a user is looking for a very specific answer to a specific question such as "what was the name of Abraham Lincoln's wife?", precision and recall may be less indicative of performance of a search service as compared to a measure that takes into account the rank of the first document in the search result set that answers the user's query.

Filter Search Results: When the user is presented with the search results, the user filters the results in order to evaluate the quality of the results as related to the decision problem. In other words, the user tries to find results relevant to the decision domain. The output of filtering search results in a set of results considered relevant by the user. Based on the information filtered from the search results, the user makes a strategic determination of the next information need to be satisfied (if any) so as to solve the decision-making problem. If the user is not satisfied with the results, he can refine his query and seek better results by going back to the query sub-process.

As stated earlier, the process of refining a query is broad in scope and is difficult to objectively separate from a new search query and is treated as a new invocation of the search process [16]. Besides, past research has shown that queries are infrequently refined in practice [13].

2.2 Search Process Factors

In this section, we describe the search process factors in greater detail.

Information Need Types

Information need refers to the type of information sought by the user in the search process. Belkin and Croft define an information need as 'a problematic situation where a person cannot attain some goals due to inadequacy of resources or knowledge' [3]. Kuhlthau defines an information need as the gap between the user's problem or topic and what the user needs to know to solve a problem [16].

Information needs have been classified in various manners by different researchers. Tague-Sutcliffe [25] classified information needs into categories such as quick reference questions, how-to-do questions, questions that involve collecting and synthesizing information about a topic, and doing a literature search for a project. These were based on the kind of information required for the user task or question for which information is sought, as well as whether there would be variation among users about expected results. Glover et al [10] suggested categories based on the kind of information sought. Categories include research papers, home pages of research organizations, topical current events, and introductory articles. Kelly et al. [15] categorized user needs into task oriented questions and fact oriented questions.

In this paper, we use a classification system based on the granularity of the information need of the user. This is based on the typical usage model of web search services as empirically observed by other researchers [24]. Our classification system consists of the following types of information needs – (i) Atomic (one answer), (ii) One Page, (iii) Some of All Pages, (iv) All of the Pages, and (v) Meta Search (any related pages). The following table shows the various information needs with typical queries.

Table 1 *Information Need Types*

Information Need Type	Information Need Description	Example
Atomic (one answer)	A very short answer to a question	What is/are the telephone area codes for Tucson, AZ?
One page	A single document	Where is the webpage for WWW conference 2008?
Some of the pages	A selection of documents	Documents about US Policy on North Korea
All of the pages	Every document matching a criterion	All documents authored by Richard Feynman
Meta Search (any related pages)	Exploratory research	"I want to learn about RFID. What are the sub-topics?"

Web Search Environment

The web search environment is the source for information in the search process, and is represented in set-theoretic terms to illustrate the differences between what a user wants, asks for, and what the user actually gets. The set-theoretic approach has been used in describing a similar situation of uncertainty in information retrieval from a scientific database [18]. A discussion of the relevance of accessed information in the context of an efficient Internet search market can be found in [7]. The principal sets in this model are shown in Figures 3 and 4, with the caveat that the figures are for explanation and not drawn to scale.

In Figures 3 and 4, *R* (blue colored star shape) be defined as the real world phenomenon that is represented by the current content of the World Wide Web (WWW). The content of the WWW is continually changing as users add and remove

content. Z (purple colored oval) is the subset of content in the WWW that is of interest to the user's information need. Y (dotted box) is the set of content that has been indexed by the search service. The indexed content is an incomplete representation of *R* because there is new content added to *R* since the last indexing and it is unlikely that the search service will ever have complete coverage of the WWW. Furthermore, the indexed content may contain references that may no longer be in *R* as the referenced objects may have been removed since the last indexing. *W* (haze colored solid box) is the set of content within *Y* that is retrieved by the search service for the user's query. The set $Z \cap W$ represents the relevant content retrieved and is marked by green stripes inside *W* in Figure 4.

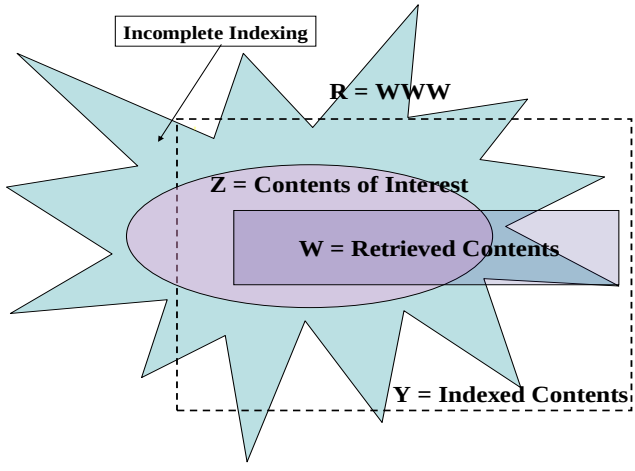


Figure 3 *Set diagram showing WWW contents in the search environment (Figure not to scale)*

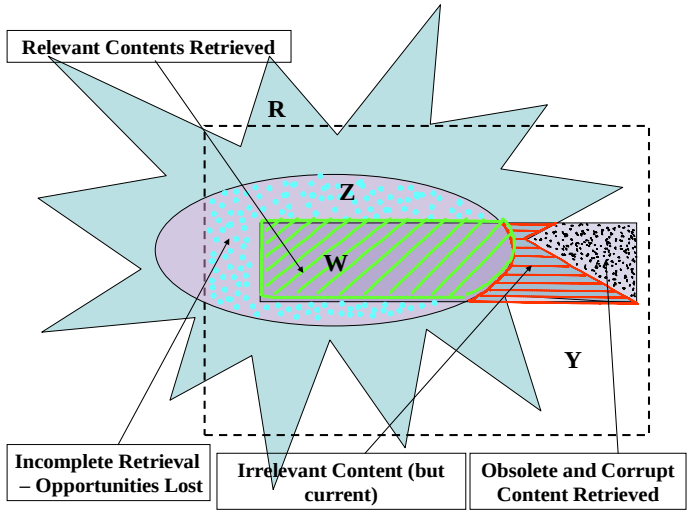


Figure 4 *Set diagram showing search contents and results (Figure not to scale)*

The set $(Z \cap Y) - W$ represents the set of relevant content that was not retrieved due to the inadequacy of the search process and is marked by blue dots inside *Z* in Figure 4. The set $W - Z$ represents content retrieved by the query that is not relevant to the decision problem. This content can be categorized into two types. First, the set $(R - Z) \cap W$ represents content retrieved that is current but not relevant and is marked by red stripes inside *W* in Figure 4. The second set $W - R$ represents content retrieved that is not relevant due to inconsistencies between the indexed content and real world phenomenon, and is represented by black dots inside Figure 4.

The ideal search scenario for the user is one where $W \subseteq Z$. Here, the user receives perfect information about the real world phenomena that is relevant to his decision problem. In a realistic situation, the user will receive less than perfect information as a result of the inconsistencies in the web search environment.

User Profile

When users translate information needs to keyword queries, the quality of keyword and phrase construction could influence the results returned by the search service. The ability to form keyword queries could be affected by many factors related to the user's profile. Factors could be demographic such as age, gender and level of education, and search related such as domain knowledge about the subject, experience with search in general and experience with specific search services. Similarly, the user profile could also affect filtering skills to find useful documents in a set of search results. In our study, we have focused on the user's domain knowledge and search experience as factors that could affect the search process. We did not have enough diversity in our subject population to test demographic factors.

Search Service Characteristics

The nature and quality of the search service could also affect the search process. As seen in Figures 3 and 4, search services could vary to the extent of indexing, frequency of indexing, and retrieval algorithms. When presenting search results to users, search services could also vary in document ranking, user, and user interface features including the number and placement of sponsored results. Given all these factors, search services may vary in ability to handle certain information needs types. Given that the internals of search services are proprietary, our study attempted to see if there was a difference between three popular search services assuming that their basic algorithms vary in some manner. As far as user interface differences are concerned, all three interfaces used similar list interfaces [27]. A key difference in the user interfaces was in the number and placement of sponsored links. Our research took into account the number of sponsored results on the search results page for every search service.

3. Experimental Design

The goal of our experimental design is to measure and analyze the influence of the various factors on the search process. We include these factors in our experimental hypotheses in order to test their impact. An online instrument was created to collect data about users and their search experience.

We designed and conducted an experiment to collect 480 independent observations from subjects going through the search process. An experiment using 40 undergraduate students as subjects was conducted where each subject was familiar with the search service process, though none of them were aware of the internal workings of a search service. The subjects were divided into two groups for ease of data collection, and the same experiment was conducted on each of the two groups. There were no incentives provided to any of the subjects to participate in the experiment. While the subjects of this experiment are biased towards those with a higher level of education than the general population, our results are consistent wherever applicable to prior results [24]. In the experiment, the subjects used Google, Yahoo and MSN as representative search services since these are the leaders in terms of number of web pages indexed [11]. We used four different scenarios representing different information need types. In each independent iteration of the above experiment, a subject was asked to formulate a query given an information need scenario and an input search service. The iterations continued till all combinations of information need scenarios and search services were chosen for each of the subjects. As a result, there are 40*4*3 or 480 independent observations of subjects going through the search service process.

3.1 Research Hypotheses

In Section 2 we have developed a model depicting the various factors affecting each sub-process of a user's search experience: the query formulation, running the query through a search service, and the filtering of the results returned by search**.**

In the first stage, we hypothesized that user profile factors of *search service experience* and *domain knowledge*, compounded with the *information need type* of the decision problem would influence the query formulation process, and thus affect the *complexity* of the query formulated. We also wanted to find out whether these influences would be different across the search services being investigated.

In the second stage, we treated the search services as black boxes and test whether varying query complexity, information need type and search service type would produce results with different *quality*. The measures for quality used in this document are:

- First Relevant Document Rank (FRDR)**:** Rank of the first document containing the piece of information sought by the user. In other words, once the user reads this document, they can terminate their search
- Precision @ 10: Since research has shown that users tend to view only one or two pages of results returned by a search service, we have considered only the first 10 results for computing precision.

In the third stage, our assumption is that ease of filtering is impacted by user characteristics such as domain knowledge and experience with search as well as factors such as information need type, quality of search results, proportion of organic results and paid placements, and the design of the user interface. Our experimental design focuses on testing the impact of search results quality, domain knowledge, user search experience, number of sponsored results and type of search service on the ease of filtering. The reason for including sponsored results is that search results contain both organic or natural results as well as paid placements in the form of

sponsored links and advertisements [22]. The user will have to spend considerable time and effort in filtering out the relevant information from the irrelevant using experiential knowledge.

The following null hypotheses are formulated for the three stages:

1. **Hypotheses related to query formulation:**
   a. There will be no difference in mean query complexity among subjects across the information need types.
   b. There will be no difference in mean query complexity among subjects for the same information need type based on level of search service experience.
   c. There will be no difference in mean query complexity among subjects for the same information need type based on level of domain knowledge about the scenario.
2. **Hypotheses related to quality of search service performance**
   a. There will be no difference in mean search result quality across the information need types.
   b. There will be no difference in mean search result quality for the same information need type based on the search service.
   c. There will be no difference in mean the search quality for the same information need type based on query complexity.
3. **Hypotheses related to filtering of search results**
   a. There will be no difference in mean search result filtering ease across the information need types.
   b. There will be no difference in mean search result filtering ease for the same information need type based on the search service.
   c. There will be no difference in mean search result filtering ease for the same information need type based on search service experience.
   d. There will be no difference in mean search result filtering ease for the same information need type based on domain knowledge about the information need types.
   e. There will be no difference in mean search result filtering ease based on search result quality(P@10).
   f. There will be no difference in mean search result filtering ease based on search result quality(FRDR).
   g. There will be no difference in mean search result filtering ease based on the number of sponsored results.

3.2 Experiment Details

In the survey instrument, subjects were asked to enter experiential factors such as their major or discipline, their year in school, and a self appraisal of their experiences with each of the search services. For the purpose of this study, we took into account the four information need types described in Section 2 – Atomic, One Page, Some of the Pages, and All of the Pages. The reason we did not consider the Meta-search information need type is that it is difficult to come up with an objective measure for the goal of a meta-search that can be expressed succinctly.



Figure 5 *Survey Instrument (Screens 1-3)*

While multiple scenarios for particular information need type were used to evaluate the experimental hypotheses, we presented the results for one typical scenario in each information need type for reasons of brevity. The scenarios were:

Atomic: Which is the third most populous city in California?

One Page: You plan on visiting India and want to find the official web page that describes the US Government's recommendations for visiting India.

Some of the pages: You recently adopted a Labrador Retriever. You want to find titles of books that you could purchase to learn how to train your dog.

All of the pages: You want to know the differences between Hepatitis A, B, C, D.

In the subsequent screens, subjects were presented with these information need scenarios. They were then asked to rate their prior domain knowledge about the scenarios, and asked to construct queries for a chosen search. Users had to filter the results, and grade the ease of filtering on a scale of 1 to 7. They were also asked to enter the number of minutes spent on the filtering task.

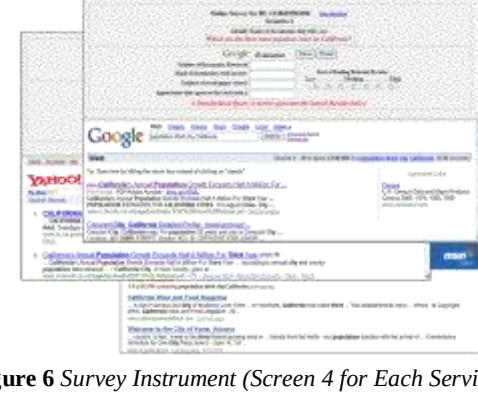


Figure 6 *Survey Instrument (Screen 4 for Each Service)*

3.3 Test Results

Before conducting the actual experiment, we performed a trial run so as to perform some analysis of variance tests on the results from the trial run. The statistical tests helped us verify that the data collected from the instruments supported the testing of the hypotheses. For example, there were differences in the query complexity measure among information need types but not among search services. The subjects demonstrated their unfamiliarity with the new MSN search service by rating it low in ease of filtering. In certain cases, the subjects wrote simpler queries when faced with certain information need types.

The trial run also provided valuable feedback to us about the instrument and the conduct of the experiment. For example, the subjects indicated that they did not understand clearly the ranking of documents on the WWW. Although the subjects were asked to evaluate the ease of filtering the results, it was difficult to record which documents retrieved were deemed relevant by the subjects. A surprising incident showed the unpredictability of the web. Some web sites hijacked the functionality of the browser (e.g., erasing the browsing history) and the subjects were unable to return to the evaluation page of the instrument.

Table 2 *Summary of ANOVA results on Hypotheses related to Query Formulation using 95% confidence interval [S = Significant, N.S = Not Significant]*

V_n	V_p	dF	F	Significance	Result
Query Complexity	Information need type	3	2.37	0.071	N.S.
	Search Experience	6	1.14	0.343	N.S.
	Domain Knowledge	7	1.05	0.397	N.S.
Search Service Performance (P@10)	Information need type	3	27.19	0.000	S.
	Search Service	2	1.08	0.342	N.S.
	Query Complexity	8	6.42	0.000	S.
Filtering of Search Results (Ease of Filtering)	Information need type	3	2.83	0.039	S.
	Search Service	2	4.92	0.008	S.
	Search Experience	6	3.68	0.004	S.
	Domain Knowledge	7	1.34	0.251	N.S.
	Search Result Quality (P@10)	14	1.81	0.039	S.
	Search Result Quality (FRDR)	9	1.36	0.207	N.S.
	Number of Sponsored Results	3	1.79	0.150	N.S.

Based on the feedback received from the trial run, we performed the actual experiment using the parameters described in Section 3, and analyzed the results to evaluate the various hypotheses regarding the model presented in this paper. On completion of the experiment, the measures of P@10 and FRDR were computed in

order to evaluate the performance of each service for each user query for each information need type.

We examined the hypotheses based on their category: query formulation, search service performance and filtering of search results. The results of the experiment were also compared to those derived from past studies [1, 27, 24]. Table 2 below provides a summary of the one-way ANOVA analysis results on all the hypotheses stated in the paper.

3.3.1. Hypotheses related to Query Formulation

First, we examined whether query complexity had any relation with the information need type. Prior literature states that users tend to use simple and ambiguous queries when searching leading to a large number of results. For example, Spink et al [24] report that the vast number of users issue queries that are short in length. In our experiment, we assumed that the query complexity was measured by the number of words as very few relational operators were used (only one query used an operator among all the 480 queries used in the experiment). A one-way ANOVA analysis found no statistically significant relation (95% confidence interval) between the information need type and the query complexity. On further analysis, a pair-wise difference analysis using the Tukey test indicated that there is a statistically significant difference in query complexity between an Atomic information need type and an All Pages information need type. This indicates that while there is a strong bias towards shorter queries, more general information need types can influence higher query complexity. It is interesting to note that the mean query complexity in all these information need types (4 to 5 words) is higher than that reported in the Spink study (2 to 3 words).

Second, we examined whether query complexity is influenced by search service experience of the users. For an initial assessment, we chose the All Pages information need type as that particular information need type has the highest standard deviation of query complexity among all the information need types. The ANOVA analysis found no statistically significant relation between search service experience and query complexity. An assessment of other information need types also revealed no statistically significant relation between query complexity and search service experience.

Third, we examined whether query complexity is influenced by the domain knowledge of the users. The range of domain knowledge was rated from 0 to 7 in our experiments. In this particular experiment, we noticed that the domain knowledge was low with respect to the One Page, Some Pages and All Pages information needs types. We found the most significant variance in domain knowledge in the Atomic information need type because of the specifics of the information need type. Overall, we found that domain knowledge does not have a significant relationship with query complexity but discovered relationships with particular information need types. In particular, we focused on specific information need types and found a statistically significant relation between domain experience and query complexity for the Atomic and All Pages information needs, highlighting the influence of the information need type on query complexity (Table 3).

Analysis revealed that the query complexity of users with intermediate knowledge (range of 2-5) is different from advanced (range of 6-7) or inexperienced (range of 1-

2). This indicates that users with intermediate domain knowledge are likely to use more complex queries.

Finally, the queries formulated for each of the search services by the users are statistically identical for every information need type, indicating that users are oblivious of the inner workings of a search service, and expect that the same queries will work for all search engines.

Table 3 *Summary of ANOVA results on Hypotheses related to effect of domain knowledge on query complexity using 95% confidence interval [S = Significant, N.S = Not Significant]*

V_n	V_p	dF	F	Significance	Result
Query Complexity	Domain knowledge (Information need type 1)	7	7.68	0.000	S.
	Domain knowledge (Information need type 2)	4	1.50	0.217	N.S.
	Domain knowledge (Information need type 3)	4	0.64	0.638	N.S.
	Domain knowledge (Information need type 4)	4	6.34	0.000	S.

3.3.2. Hypotheses Related to Quality of Search Service Performance

In this section, we examined the factors that influenced search result quality. As mentioned before, we measured search result quality using two factors: Precision@10 and First Relevant Document Rank (FRDR).

First, we found that the P@10 search result quality metric were deeply influenced by the information need type in question. The ANOVA analysis indicated that both the metrics were different for the various information need types. On a deeper examination using the Tukey test, we found that P@10 was statistically higher for the Some Pages and All Pages information need types compared to that for the Atomic and One Page information need types. This was consistent across all search engines. The conclusions were also not affected if FRDR is used in place of P@10.

In case of the One Page information need type, this is the expected result since the user is looking for a specific page. However, with the Atomic need, the user is looking for an answer to a very specific question, and that can be found on multiple pages. The P@10 for that need was significantly lower than that for other information need types. One possible reason for this is the use of the ranking algorithms used for web search in all the search services used in this survey (Google, MSN, Yahoo). In these search services, the ranking is done using criteria such as location and frequency of keyword occurrence, and the number of pages on the Web that point to a particular page [1]. This ranking approach may be more suitable for locating a class of documents about a topic as opposed to atomic pieces of information, and this possibly explains the difference in search result quality.

We also did a similar analysis with FRDR and found that FRDR has a statistically significant relation with the type of information need. In particular, FRDR is much worse for the One Page information need type. This is also not surprising as this is

the only information need where the correct search result is exactly one page and is thus incongruous with the general ranking approach used in search services.

Next, we examined all the other factors that can influence search result quality but found that the specific search service, or search service experience did not have a statistically significant relation with search result quality. While a visual inspection of plots could have possibly indicated a better search result quality from the Google search service, a statistical analysis did not reveal a wide disparity between the search services. A possible explanation for this is that the page rank algorithm has matured enough in the past five years so as to eliminate potential advantages in performance for a single search service provider. In fact, there was no statistical relation between search provider and search result quality (both P@10 and FRDR) for the same information need type. This means all search engines handle the different information need types with more or less the same level of effectiveness.

Domain knowledge did not seem to have a significant relation with search results quality except in the case of the One Page information need type where intermediate domain knowledge seemed to have a relation with the search result quality (P@10 and FRDR).

Finally, we examined whether the query complexity for a particular information need type influenced search result quality. While query complexity does not share a statistically significant relation with search result quality, an examination of individual information need types reveal statistically significant relations for Atomic, One Page and Some of the Pages information need types (Table 4). This again highlights the role of the information need type in determining search service performance.

Table 4 *Summary of ANOVA results on Hypotheses related to effect of query complexity on search service performance using 95% confidence interval [S = Significant, N.S = Not Significant]*

V_n	V_p	dF	F	Significance	Result
Search Service Performance (P@10)	Query Complexity for Information need type 1	6	3.2	0.010	S.
	Query Complexity for information need type 2	4	3.53	0.013	S.
	Query Complexity for information need type 3	4	12.11	0.000	S.
	Query Complexity for information need type 4	5	0.19	0.964	N.S.

3.3.3. Hypotheses related to Filtering of Search Results

Finally, we examined how filtering of search results is affected by various factors. Filtering measures the user's ability to disambiguate between relevant and irrelevant search results and is determined by the design of the user interface. Empirical evidence has shown that users prefer certain user interfaces more than others [27].

We first examined whether there was any variation in ease of search result filtering across the information need types. The ANOVA analysis revealed that there are

differences in the ease of filtering across information need types; a deeper analysis using the Tukey test revealed that the Some Pages and All Pages information need types are easier to filter than the Atomic and One Page information need types. This is expected because the distribution of search result relevance is more skewed in the latter information need types and makes the task of locating the correct information harder.

Next, we examined whether the search service had any impact on the ease of filtering. Here again, a similar analysis showed that the Google and Yahoo search services provided easier use of filtering than the MSN search service. Further analysis revealed that the number of sponsored advertisements were proportionately higher in the MSN search service as compared to the rest. Therefore, we found a strong correlation between the use of sponsored advertisements and the lack of ease of filtering. Looking more closely at individual information need types (Table 5), we found a statistically significant relationship in the case of the 'Some of the Pages' information need type because the information need type has a higher commercial value and therefore more sponsored links.

Table 5 *Summary of ANOVA results on Hypotheses related to effect of the effect of the number of sponsored results on the ease of filtering search results using 95% confidence interval [S = Significant, N.S = Not Significant]*

V_n	V_p	dF	F	Significance	Result
Filtering of Search Results (Ease of filtering)	Number of Sponsored Results (for information need type 1)	2	2.97	0.06	N.S.
	Number of Sponsored Results (for information need type 2)	1	2.37	0.129	N.S.
	Number of Sponsored Results (for information need type 3)	1	5.75	0.02	S.
	Number of Sponsored Results (for information need type 4)	1	0.00	0.99	N.S.

We did find that the quality of search results had a statistically significant relation with the self reported number of minutes on the task. The higher the P@10 metric for a particular information need type, the greater is the number of minutes spent on filtering results for the query. This seems to indicate that when more documents appear to be relevant, the time spent on filtering is greater. We did not find any significant effect of search service experience or domain knowledge on the ease of filtering except in one case where there was a pronounced impact of search service experience on the ease of filtering for the Atomic page information need type. This is not unexpected as this particular information need type demonstrates a highly skewed distribution of relevance among search results and makes the task of locating the correct information harder.

4. Conclusion and Managerial Implications

Search services are now ubiquitously employed in searching for documents on the Internet and on enterprise intranets. Search services may exhibit different behavior depending on the information need type, the quality of the search service, the ease of filtering results, the user's domain knowledge and search experience. The goal of this research is to understand the principal factors that are involved when a user with an information need uses a search service. In this paper, a formal research model including the various sub-processes of the search process, the factors influencing and the output of each sub-process has been presented.

An experiment was conducted in order to perform a realistic comparison among three popular search services so as to evaluate whether the user's query complexity, search service result quality and ease of search result filtering is influenced by factors such as the information need type, the search service, the user's domain knowledge and experience with the search service. We conducted an experiment using three popular search services – Google, Yahoo and MSN, and analyzed the data to outline the primary factors that affect the various stages of the search process. The principal findings of this particular experiment are that:

- More general information need types tend to generate higher query complexity and users with intermediate knowledge write more complex queries, and query complexity does have a statistically significant effect on search result quality. We also found that domain knowledge influences query complexity and search result quality for certain information need types.
- The type of information need has a statistically significant effect on the quality of search results. The search service quality is higher for more general information needs than that for specific information needs because of the bias of the popular page rank service towards page popularity.
- In the case of "Some of the Pages" information need, the use of sponsored advertisements has a strong correlation with a lower ease of filtering for search results because of the higher commercial value.

Overall, we found that the information need type influences every sub-process of the search process and is statistically more significant than the search service itself.

4.1 Managerial Implications

Managers and business analysts are often tasked to strategic decision making in order to satisfy the business goals of their organization. A key aspect of decision making requires the satisfaction of information needs so as to provide a firm basis for the validity of a decision [16]. Based on the wealth of information available on the Web and on Intranets, managers and business strategists use search services to satisfy the information needs – for example, they search social networking sites to get a holistic picture of potential candidates for employment [9] or use search tools to detect sentiment about products and services [2]. A key implication of the findings of the paper indicate that an orderly decomposition of the decision making process into information needs is critical for the decision. If the decision making process is decomposed to very specific information needs, then the probability of the information need being satisfied is reduced as the Web and intranets cater to more generic information needs. Conversely, a decision making process composed of

generic information needs has a higher probability of satisfying the information needs in question. If it is not possible to decompose a decision making process into generic information needs, specialized search tools or external consultants should be considered.

5. References

1. Arasu, A., Cho, J., Garcia-Molina, H., Paepcke, A., & Raghavan, S. (2001). "Searching the Web". *ACM Transactions on Internet Technology (TOIT)*, 1(1), 2-43.
2. Barre, I. (2006). Amateur reviews changing approach of small business. San Francisco Chronicle, September 3, 2006. http://www.sfgate.com/cgi-bin/article.cgi?file=/c/a/2006/09/03/BUG45KT0RH1.DTL&type=printable
3. Belkin, N. et. al. "Information filtering and information retrieval: two sides of the same coin?", *Communications of the ACM*, 35(12), 29 – 38, December 1992.
4. Brooks, C., S. P. Burke, and G. Persand (2005), Autoregressive Conditional Kurtosis, *Journal of Financial Econometrics*, 3(3), 399{421}.
5. Buckley, C., and Voorhees, E. M. "Evaluating evaluation measure stability", *Proceedings of the 23rd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, July 2000.
6. Chau, M. "Posters: Applying web analysis in web page filtering", *Proceedings of the 4th ACM/IEEE-CS joint conference on Digital libraries*, June 2004.
7. Choi, S., Stahl, D., and Whinston, A.B. The Economics of Electronic Commerce, MacMillan Technical Publishing, 1997.
8. Eastman, C. M., and Jansen, B. J. "Coverage, relevance, and ranking: The impact of query operators on Web search engine results", *ACM Transactions on Information Systems (TOIS)*, 21(4), October 2003.
9. Finder, A. (2006). When a Risqué Online Persona Undermines a Chance for a Job. New York Times, June 11, 2006. http://select.nytimes.com/gst/abstract.html?res=F30712F739550C728DDDAF0894DE404482.
10. Glover, E. J., Lawrence S., Birmingham, W. P., and Giles, L. C. "Architecture of a metasearch engine that supports user information needs", *Proceedings of the Eighth International Conference on Information and Knowledge Management*, November 1999.
11. Gulli, A., and Signorini, A. "Posters: The indexable web is more than 11.5 billion pages", *Special interest tracks and posters of the 14th International Conference on World Wide Web*, May 2005.
12. Hawking, D., Craswell, N., Bailey, P., and Griffiths, K. "Measuring search engine quality. Information Retrieval", *Journal of Information Retrieval*, 4 (1), 33-59, 2001.
13. Jansen, B. J., Spink A., Bateman J., and Saracevic T. "Real life information retrieval: a study of user queries on the Web", *ACM SIGIR Forum*, 32(1), 1998.

14. Kauffman, R. J. and Walden, E. A. "Economics and Electronic Commerce: Survey and Directions for Research", *International Journal of Electronic Commerce*, 5 (4), 5-116, Summer 2001.
15. Kelly D., Yuan X., Belkin N. J., Murdock V., and Croft B. W. "Poster session: Features of documents relevant to task- and fact- oriented questions", *Proceedings of the Eleventh International Conference on Information and Knowledge Management*, November 2002.
16. Kuhlthau, C. C. "Inside the Search Process: Information Seeking from the User's Perspective." *Journal of the American Society for Information Science*, 42 (5), 361-371, 1991.
17. Kwan, S. K. "A Job Contractor Function in Computing System Scheduling", PhD in Management dissertation, Graduate School of Management, UCLA, 1982.
18. Kwan, S. K., Olken, F. and Rotem, D. "Uncertain, Incomplete and Inconsistent Data in Scientific and Statistical Databases", in Motro, A. and Smets, P. (ed.) Uncertainty Management in Information Systems: From Needs to Solutions, Kluwer Academic Publishers, Norwell, 1997.
19. Lucas, W. and Topi, H. "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, November 2001.
20. Marschak, J. "Economics of Inquiry, Communicating, Deciding", *American Economics Review*, 58, 1-18, May 1968.
21. Marschak, J. and Radner, R. Economic Theory of Teams, Cowles Foundation Monograph 22, Yale University Press, 1972.
22. Moran, M and Hunt, B. Search Engine Marketing Inc. Driving Search Traffic to Your Company's Web Site, IBM Press, Upper Saddle River, New Jersey, July 2005.
23. Shah, C., and Croft, B. W. "Opening session: Evaluating high accuracy retrieval techniques", *Proceedings of the 27th Annual International Conference on Research and Development in Information Retrieval*, July 2004.
24. Spink, A., Jansen, B. J., Wolfram, D., & Saracevic, T. (2002, 2002/03). "From e-sex to e-commerce: Web search changes". *Computer*, 35, 107-109.
25. Tague-Sutcliffe, J. "Measuring the informativeness of a retrieval process", *Proceedings of the 15th annual International ACM SIGIR Conference on Research and development in information retrieval*, June 1992.
26. Vaughan, L. "New measurements for search engine evaluation proposed and tested", *Information Processing & Management*, 40 (4), 677-691, May 2004.
27. Venkatsubramanyan, Shailaja and Perez-Carballo, Jose. "Techniques for Organizing and Presenting Search Results: A Survey", *Journal of Information Science and Technology (JIST)*, 4(2), 2007.