

1-1-2009

Using information noise to compute the economic benefit of a search service

Shailaja Venkatsubramanyan

San Jose State University, shailaja.venkatsubramanyan@sjsu.edu

Stephen K. Kwan

San Jose State University, stephen.kwan@sjsu.edu

Follow this and additional works at: https://scholarworks.sjsu.edu/mis_pub

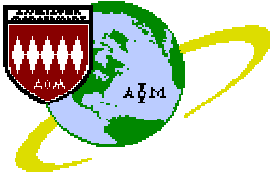


Part of the [Management Information Systems Commons](#)

Recommended Citation

Shailaja Venkatsubramanyan and Stephen K. Kwan. "Using information noise to compute the economic benefit of a search service" *Journal of Information Technology Management* (2009): 1-11.

This Article is brought to you for free and open access by the Management Information Systems at SJSU ScholarWorks. It has been accepted for inclusion in Faculty Publications by an authorized administrator of SJSU ScholarWorks. For more information, please contact scholarworks@sjsu.edu.



Journal of Information Technology Management

ISSN #1042-1319

A Publication of the Association of Management

USING INFORMATION NOISE TO COMPUTE THE ECONOMIC BENEFIT OF A SEARCH SERVICE

SHAILAJA VENKATSUBRAMANYAN
SAN JOSE STATE UNIVERSITY
venkat_s@cob.sjsu.edu

STEPHEN K. KWAN
SAN JOSE STATE UNIVERSITY
kwan_s@cob.sjsu.edu

ABSTRACT

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 type of information need, the quality of the search service, the ease of filtering results, the user's domain knowledge and search experience. Users are thus faced with the selection of a search service in order to minimize cost, reduce uncertainty, and maximize the benefits derived for their efforts. This research develops a model of the search process and considers the noise effects of querying, search and filtering of results to derive a benefit measure for evaluating the search service. A methodology for comparing search services based on the benefit measure is presented along with an empirical analysis using three popular search services to validate the methodology. Our analysis revealed that the economic benefit of a search service is determined more by the information need type than by the search service itself. For a particular information need type, the value is determined primarily by the ease of filtering in the search service interface.

Keywords: search service evaluation, information noise, information needs, information valuation

INTRODUCTION

With the growth of the World Wide Web, one technology that has become ubiquitous and indispensable is that of Web search. Search services are now widely employed in searching for documents on the Internet and on enterprise intranets. There are many commercial search services available to users. Users are thus faced with the task of comparing search services in order to minimize costs, reduce uncertainty, and maximize the benefits derived for their efforts.

The process used to search for information is composed of multiple steps. An example is Kuhlthau's

information search process or ISP [14]. Kuhlthau's ISP 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 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. Finally, the user either makes a decision based on

the search results or decides to abandon the search. (See Figure 1)

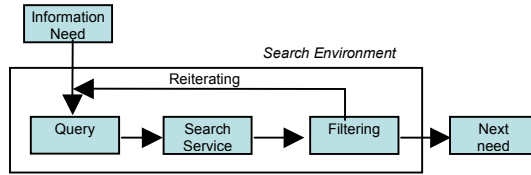


Figure 1: Search Process Model

Traditionally, precision and recall have been used to evaluate search. Other measures used include the stability of a search service over time, recall or precision over a subset of retrieved documents, and correlation between human and engine ranking [4, 10, 24]. However, all these measures focus on only the performance of the search service, and do not measure the noise or garbling introduced during various stages of the search process. First, user characteristics that may play a role in the search outcome are not taken into account. Noise may be introduced into the search process by the inability on the part of the user to provide good query terms. When users translate information needs to keyword queries, the quality of keyword and phrase construction could influence the results returned by the search service. Second, certain search services may be better at handling certain types of information needs than others. Lastly, users have to use their filtering skills to find useful documents. The filtering skills could depend on several factors such as the user’s domain knowledge and skill with searching, as well as the user interface of the search service.

Users are thus faced with the choice of a search service to get the best possible results for their information needs while factoring in user ability as well as effectiveness of the search service. While most Internet-based search services are free to the general public, the value of a search service in satisfying an information need has a tangible economic value. Users are thus faced with making a decision about which search service to use so as to extract the maximum economic value for their information need. The decision is bound to depend on user characteristics that affect the search process, the type of information need, and search service characteristics. There is no standard or recommended manner in which search services can be compared in order to pick one that would be most appropriate for specific users with their individual information needs. In the case of intranets, more often than not, search services have to be purchased for a price. In such a situation, a

decision has to be made about which search service would be most useful for the needs of that organization. This paper describes a methodology for comparing search services within the context of certain types of user needs.

The key contribution of this research is a methodology for estimating and comparing the economic value of a search service based on a benefit measure. Here, we utilize the definition of value from the perspective of the benefit to the user. We derive the economic benefit of a noisy information structure to come up with a comparable benefit equation that can be used to rate search services. To validate our methodology, we conducted an empirical analysis using three popular search services – Google, Yahoo and MSN, and analyzed the data to estimate the overall value of a search service. Our analysis revealed that the economic benefit of a search service is determined more by the information need type than by the search service itself, and that there is no statistically significant difference between the qualities of the three search services. Within an information need type, the benefit is determined primarily by the ease of filtering in the search service interface.

The rest of the paper is organized as follows: The Research Model section presents a formalization of the model behind our research so as to provide a theoretical framework for our methodology. The empirical analysis presents the empirical analysis and the subsequent results, and conclusions are presented in the final section.

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) 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.

- 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.
- 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 at which point the user can reformulate the query or abandon the search process.

- d) Noise: Each of the search process outputs generates information noise due to the inherent characteristics of the process itself. For example, a non-ideal query may generate irrelevant search results that are hard to filter. As a result, we can say that the particulars of the query act as a source of the information.
- e) Measures: 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.

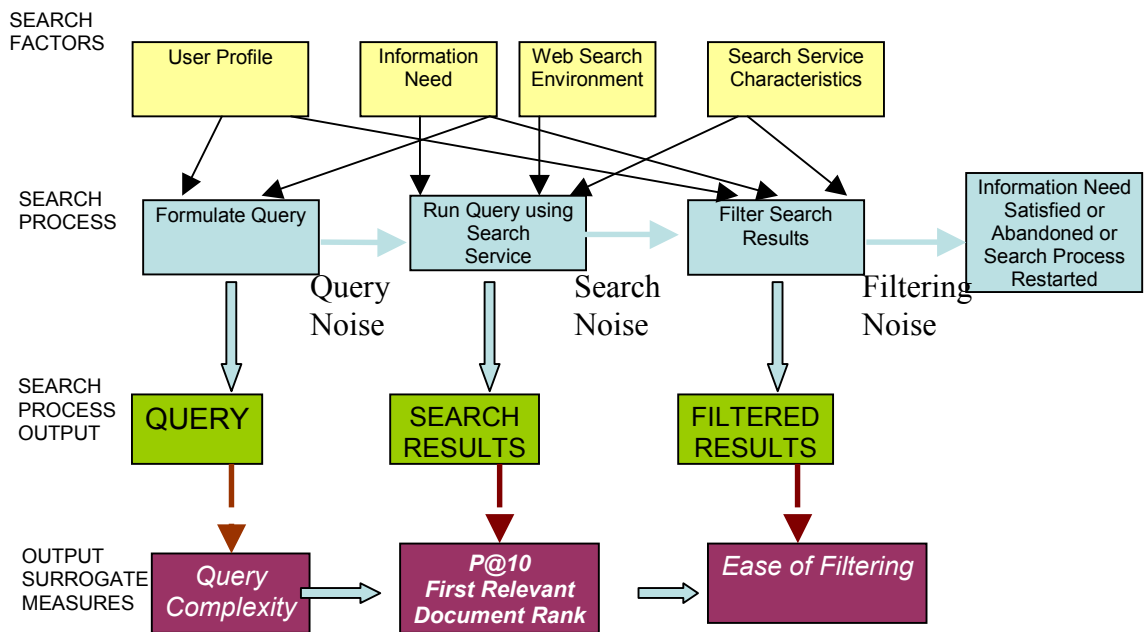


Figure 2: Web Search Model

Information Need

Before, we describe the search process, we elaborate on the concept of an information need as this concept drives the entire search process. Information need refers to the type of information sought by the user. Belkin et. al. define an information need as a problematic situation where a person cannot attain some goals due to inadequacy of resources or knowledge [2]. 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 [14].

Information needs have been classified in various manners by different researchers. Tague-Sutcliffe [23] 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. [9] 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. [13] 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 [22]. Our classification system consists of the following types of information needs – (i) Atomic (one answer), (ii) One Page (jewel), (iii) Some of All Pages, (iv) All of the Pages, and (v) Meta Search (any related pages). The following table shows the various information need types with examples that illustrate a typical query that might be used to satisfy the information need.

Table 1: Information Needs

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 2005?
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?"

The Search Process

The search process that is the foundation of our model is based on the statistical decision model from Marschak [18,19] and later applied to a computing environment in [15]. In the model, the decision making process is divided into inquiring, communicating and deciding sub-processes with costs associated with each. Marschak also developed the concept of informative-ness based on the noise in “information structures”. In our analysis, the concept of information structures is equivalent to modern search services which transform events of the environment into search results.

The search process (Figure 1) has the following steps:

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

Query: For every information need, the user formulates a query for the search service. This may be a simple query consisting of one or more keywords, or an advanced query consisting of keywords as well as operators such as “+”, “-”, or quotation marks. 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 and experience using search services [17]. It is expected that given the same decision making scenario, different users will formulate queries with varying degree of complexity that

produce different results of varying quantity and quality under the influence of the factors listed above. In our model, this step is one of the sources of noise. In other words, the quality of the query could potentially enhance or reduce the quality of the output in terms of informative-ness to the decision maker.

The correctness and fineness of the query have direct effects on the quality of returned results from the search service. For example, a query might not be directing the search service correctly and result in retrieving not relevant results. Another example is that due to the fineness (or not so fineness) of the query, the results miss relevant content. The effect of the query on the quality of the results in the form of noise is represented as η_Q .

Search Service: Once the user’s query is submitted to the search service, the search service is deployed to process the query, and execute the underlying algorithms to return results to the user. Each search service can be characterized by how it spiders the Web contents, how often it performs the spidering, 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 [10]. The search service quality is thus another source of noise and is represented as η_S .

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 [4, 10, 24]. In this research, we have used the measure of Precision at 10 or P@10 as a indicator for the accuracy of a search service [21]. Precision at 10 refers to the proportion of documents relevant to the user's need in the top 10 results presented by the search service.

Filtering: When the user is presented with the 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. Depending on user characteristics such as those described in the Query step, the filtered results could vary from user to user, and thus contribute as another source of noise.

The results of the query being returned to the user contain both organic or natural results as well as paid placements in the form of sponsored links and advertisements [20]. The user will have to spend time and effort in filtering out the relevant information from the irrelevant using experiential knowledge as well as the specific information need.

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 quality of search results, proportion of organic results and paid placements, and the design of the user interface.

The effect of filtering in detracting the user from relevant results in the form of information noise is represented in the economic model as η_F in the Benefit Analysis section.

Deciding: 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 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 step.

Benefit Analysis

In this paper, we utilize the definition of value from the perspective of the user. In this section, we outline the methodology for estimating and comparing the economic value of a search service. The concept of value used in this paper is based on the measure of worth that is based purely on the utility derived from the consumption of a product or service [3]. Utility derived value allows products or services to be valued based on outcome instead of demand or supply theories that have the inherent ability to be manipulated. For example, the real value of a book sold to a student who pays \$50.00 at the cash register for the text and who learns nothing from the

content is essentially zero. However; the real value of the same text purchased in a thrift shop at a price of \$0.25 and provides the reader with an insight that allows him or her to earn \$100,000.00 in additional income is \$100,000.00 or the extended lifetime value earned by the consumer. This definition of economic value is more in alignment with the search service domain as opposed to classic economic definitions of value based on cost of input and demand-supply parameters.

In our analysis, we first postulate the benefit of a decision in a noise-less information space, then take noise into account and finally, adapt the benefit equation to the search process.

First, we compute the benefit of a decision in a noise-less information space. Let us assume X is the state of states in the decision-maker's environment, W the set of messages received by the decision maker, and A the set of possible actions to be taken by the decision maker. For every state x in X , the prior probability of being in the state is represented as $\pi(x)$, and the decision maker's strategy on the action taken on receiving a message w in W is given by the function $\alpha(w) = a$ where a is a member of A . The benefit of taking an action a in state x is given by the function $\beta(a, x)$. As the information structure is noise-less, the message w generated in a state in a state x is given by the function $\eta(x) = w$. With all these above assumptions, the benefit equation for the decision maker's environment can be represented as the summation of the benefits at every state x in the decision maker's environment:

$$B = B(\eta, \alpha; \pi, \beta)$$

$$(A.1) = \sum_x \pi(x) \beta(\alpha(\eta(x)), x)$$

Next, we adapt this equation to a noisy environment. The most important change to the equation A.1 comes in the calculation of the probability of being in state x . Due to the noisy environment, there is now a conditional probability of the message w being generated given state x and this is represented by $p(w|x)$. If the joint probability of a message w being generated while in state x is given by $p(x, w)$, the probability of being in state x is given by the equation: $\sum_w p(x, w) = \sum_w \pi(x) p(w|x)$.

If we assume that η is a noisy function that determines the message delivered to the user given an event, then $\eta_{xw} = p(w|x)$ where $\sum_w \eta_{xw} = 1$ as the sum of all conditional probabilities must equal 1 for the state x . Consequently, the benefit equation takes the form:

$$\begin{aligned}
 (A.2) \quad B &= B(\eta, \alpha; \pi, \beta) \\
 &= \sum_x \sum_w \pi(x) p(w|x) \beta(\alpha(w), x) \\
 &= \sum_x \sum_w \pi(x) \eta_{xw} \beta(\alpha(w), x)
 \end{aligned}$$

Blackwell's theorem [15] tells us that we can use the η matrix to compare two information spaces.

Specifically, the theorem states that $\eta = [\eta_{xw}]$ is more informative than $\eta' = [\eta'_{xw}]$ if and only if there exists a Markov matrix $Q = [Q_{ww'}]$ such that $\eta' = \eta \cdot Q$.

Finally, we adapt the benefit equation to the web search process. It was shown in our research model that the query, search and filtering sub-processes contribute to the information noise emanating from the search process. That is, they contribute to the make-up of the information space described above. In effect, we can formulate the equation below using the terms η_Q, η_S and η_F defined above,

$$(A.3) \quad \eta = \eta_Q \bullet \eta_S \bullet \eta_F$$

Now, we can compute the η matrix for each search service and use Blackwell's theorem to compare the search service for a benefit assuming that the remaining variables π, x, α, β are constant in the decision making scenario.

EMPIRICAL ANALYSIS

The goal of our empirical analysis is to measure and analyze the influence of the various factors on the search process as postulated in our research model. 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 [22]. 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 [9]. 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 type and an input search service. The iterations continued till all combinations of information need types 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.

Experiment Details

In the survey instrument (see Figure 3), subjects were asked to enter experiential factors such as their major or discipline, their year in school, and a self appraisal of their experience with each of the search services (Google, Yahoo and MSN).



Figure 3: Survey instrument (screens 1-3)

For the purpose of this study, we took into account the four information need types described in the research model section – 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.

One scenario was constructed for each information need type as using multiple scenarios for a particular information need type did not provide any significant additional analytic value:

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, 1 being very difficult to filter and 7 being very easy to filter. They were also asked to enter the number of minutes spent on the filtering task.



Figure 4: Survey instrument (screen 4 for each service)

Test Results

Before conducting the actual experiment, we performed a trial run so as to perform sanity checks on the results from the trial run. 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 clearly understand the concept of evaluating the precision of a search query. 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 high-jacked the functionality of the browser (e.g., erasing the browsing history) making it difficult to return to the evaluation page of the instrument. Based on the feedback received from the trial run, we performed the actual experiment and analyzed the results to evaluate the economic benefit for each of the search services for the various information need types.

Estimation of Noise Parameters

The next stage of our experiment was to identify the noise parameters for the search sub-processes of querying, searching and filtering. For each of these sub-processes, we measure noise by calculating the difference between a sub-process metric with respect to the ideal sub-process metric.

First, we derived an estimation measure for the query sub-process noise. We did this by identifying the query with the highest surrogate measure in terms of search result quality. In this paper, for each information need type, we use the query with the highest P@10 and anoint the query to be the ideal one. After this, we use information theory to compute the noise between the

query submitted by a user with a particular information need and the ideal query. Specifically, we use the Damerau-Levenshtein distance which is a string metric that finds the difference between two strings by giving the minimum number of operations needed to transform one string into the other where an operation is an insertion, deletion, or substitution of a single character.

Next, we derived an estimation process for the search sub-process noise. We did this by assuming that the ideal P@10 for each user query in an information need type is perfect and is represented by the numerical representation of 1. Then, for each user query submitted to a search service in an information need type, we estimate the search sub-process noise for the query as the Euclidean distance between the P@10 of the user query in the search service and the ideal P@10. Finally, we compute the search sub-process noise for a search service in an information need type as the average of the search sub-process noises for all the user queries submitted to the search service in the information need type.

Finally, we derived an estimation process for the filtering sub-process noise. We did this by assuming that the ideal ease of filtering for a user queries for an information need type is perfect and is represented by the numerical representation of 7 (as determined from the scale used in the experiment). Then, for each user query submitted to a search service in an information need type, we estimate the filtering sub-process noise for the query as the Euclidean distance between the ease of filtering of the user query and the ideal ease of filtering. Finally, we compute the filtering sub-process noise for a search service in an information need type as the average of the filtering sub-process noises for all the user queries

submitted to the search service in the information need type.

Based on the above estimation, we estimate the noise of a sub-process for a particular search service and an information need type as the average of the noise of all user queries in the sub-process for the search service and the information need type. Then, using Equation A.3, we come up with the total noise for the search process for a particular search service and an information need type as the product of the sub-process noises for the search service and the information need type.

Results

Table 2 below shows the estimation of the various noise parameters from our experiment. The participants in the study used the same queries for all the three search services as a result of which the noise parameter η_Q was independent of the search service and was related to the information need type in question. In this experiment, the Atomic information need type demonstrated significantly less query noise due to the specificity of the information need.

Table 2: Estimation of querying, searching, filtering and total noise for Google, Yahoo and MSN for each information need type

					η
Atomic	Google	0.45	0.74	1.67	0.56
	Yahoo	0.45	0.66	2.13	0.63
	MSN	0.45	0.77	2.80	0.97
One-page	Google	0.75	0.84	1.60	1.00
	Yahoo	0.75	0.86	1.20	0.77
	MSN	0.75	0.93	2.13	1.49
Some of the pages	Google	0.72	0.50	1.13	0.42
	Yahoo	0.72	0.41	1.67	0.49
	MSN	0.72	0.58	2.33	0.97
All of the pages	Google	0.63	0.59	1.87	0.70
	Yahoo	0.63	0.56	0.53	0.19
	MSN	0.63	0.59	1.20	0.45

The noise parameter η_S is a measure of the search service quality and is a function of both the search service and the information need type. As can be seen, the noise is higher for specific information need types such as Atomic and One-page – this is not surprising as search services are tailored for more generic information need types such as Some of the pages and All of the pages. Furthermore, both Google and Yahoo appear to demonstrate lower noise, but the result is not statistically significant.

The noise parameter η_F is a measure of the ease of filtering of search results and is predominantly determined by the search service interface. The MSN interface (in our experiments) was accompanied by a lot of sponsored advertisements that increased the difficulty of filtering search results. Consequently, the filtering noise is statistically higher for MSN than the other two search services.

Overall, statistical analysis reveals that the information noise η and therefore the economic benefit is

determined more by the information need type than the search service. For a given information need type, the information noise (and economic benefit) is determined primarily by the ease of filtering in the search service interface. As a result, for a specific type of information need, Google and Yahoo demonstrate significantly lower information noise.

CONCLUSION

This paper presents a model and a methodology to allow users to compare search services. Search services may exhibit different behaviors depending on the information need, the quality of the search service, the ease of filtering results, the user’s domain knowledge and search experience. To achieve this goal, we outline a methodology for estimating and comparing the economic value of a search service. In this paper, we utilize the definition of value from the perspective of the benefit to the user and derive the economic benefit of a noisy

information structure to come up with a comparable benefit equation that can be used to rate search services. Finally, we use empirical analysis with three popular search services Google, MSN and Yahoo to validate the methodology. The key results from the analysis are:

1. The query noise η_Q and the search noise η_S are more related to the information need in question, and there is no statistically significant variation between the search services for search noise.
2. The filtering noise η_F is predominantly determined by the search service interface and the design of the MSN interface (in our experiments) increased the difficulty of filtering search results.
3. Overall, statistical analysis reveals that the information noise η and thus the economic benefit are determined more by the information need type than the search service. For a given information need type, the information noise (and the economic benefit) is determined primarily by the ease of filtering in the search service interface.

REFERENCES

- [1] Arasu, A., Cho, J., Garcia-Molina, H., Paepcke, A., & Raghavan, S. "Searching the Web". *ACM Transactions on Internet Technology (TOIT)*, 1(1), 2-43, 2001.
- [2] 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.
- [3] Brooks, C., S. P. Burke, and G. Persaud (2005), Autoregressive Conditional Kurtosis, *Journal of Financial Econometrics*, 3(3), 399{421}.
- [4] 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.
- [5] Chau, M. "Posters: Applying web analysis in web page filtering", *Proceedings of the 4th ACM/IEEE-CS joint conference on Digital libraries*, June 2004.
- [6] Choi, S., Stahl, D., and Whinston, A.B. *The Economics of Electronic Commerce*, MacMillan Technical Publishing, 1997.
- [7] 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.
- [8] 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.
- [9] 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.
- [10] Hawking, D., Craswell, N., Bailey, P., and Griffiths, K. "Measuring search engine quality. Information Retrieval", *Journal of Information Retrieval*, 4 (1), 33-59, 2001.
- [11] 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.
- [12] 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.
- [13] 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.
- [14] Kuhlthau, C. C. (1991) "Inside the Search Process: Information Seeking from the User's Perspective." *Journal of the American Society for Information Science*, 42 (5), 361-371, 1991.
- [15] Kwan, S. K. *A Job Contractor Function in Computing System Scheduling*, PhD in Management dissertation, Graduate School of Management, UCLA, 1982.
- [16] 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.
- [17] 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.
- [18] Marschak, J. "Economics of Inquiry, Communicating, Deciding", *American Economics Review*, 58, 1-18, May 1968.

- [19] Marschak, J. and Radner, R. *Economic Theory of Teams*, Cowles Foundation Monograph 22, Yale University Press, 1972.
- [20] 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.
- [21] 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.
- [22] 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.
- [23] 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.
- [24] Vaughan, L. "New measurements for search engine evaluation proposed and tested", *Information Processing & Management*, 40 (4), 677-691, May 2004.
- [25] Venkatsubramanyan, Shailaja and Perez-Carballo, Jose. "Techniques for organizing and presenting search results: A survey", *Journal of Information Science and Technologies*, 4(2), 2007.

Systems from the University of Arizona. She is currently working on research projects related to knowledge management, information retrieval, search engine valuation, web log analysis and impact of online information on impression formation.

AUTHOR BIOGRAPHY

Stephen K. Kwan is a professor of Management Information Systems at the College of Business, San José State University. He was the founding chair of the department and had served as the Senior Associate Dean of the College. He is currently working on the development of a Service Science, Management and Engineering (SSME) program at his university, and the promotion of education and research about standardization in US universities. He received a B.S. and M.S. in Computer Science from the University of Oregon, and a Ph.D. in Management from UCLA. His research has been published in *European Management Journal*, *Information Systems*, *Data Base*, *Queueing Systems: Theory and Applications* and *ACM Transactions on Office Information Systems*. He is also a contributor to the Standards Edge series.

Shailaja Venkatsubramanyan is an Assistant Professor in the Management Information Systems (MIS) department in the College of Business at San Jose State University. She has a Ph.D. in Management Information