



An exploration of risk information search via a search engine: Queries and clicks in healthcare and information security

Jingguo Wang ^{a,*}, Nan Xiao ^b, H. Raghav Rao ^{b,c,**}

^a Information Systems and Operations Management, College of Business, the University of Texas at Arlington, United States

^b Management Science and Systems, School of Management, State University of New York at Buffalo, United States

^c WCU Visiting Prof, SSME Department, Sogang University, Seoul, South Korea

ARTICLE INFO

Article history:

Received 28 April 2011

Received in revised form 8 September 2011

Accepted 18 September 2011

Available online 25 September 2011

Keywords:

Search engine

Information foraging

Search log

Session length

Query click rate

Markov Chain Monte Carlo simulation

ABSTRACT

The general public is increasingly using search engines to seek information on risks and threats. Based on a search log from a large search engine, spanning three months, this study explores user patterns of query submission and subsequent clicks in sessions, for two important risk related topics, healthcare and information security, and compares them to other randomly sampled sessions. We investigate two session-level metrics reflecting users' interactivity with a search engine: session length and query click rate. Drawing from information foraging theory, we find that session length can be characterized well by the Inverse Gaussian distribution. Among three types of sessions on different topics (healthcare, information security, and other randomly sampled sessions), we find that healthcare sessions have the most queries and the highest query click rate, and information security sessions have the lowest query click rate. In addition, sessions initiated by the users with greater search engine activity level tend to have more queries and higher query click rates. Among three types of sessions, search engine activity level shows the strongest effect on query click rate for information security sessions and weakest for healthcare sessions. We discuss theoretical and practical implications of the study.

© 2011 Elsevier B.V. All rights reserved.

1. Introduction

Nearly all problem solving and decision making relies on information search [71]. The growth of the Internet and the World Wide Web has provided easier access to information. One of the benefits that the Internet offers is the quantity and quality of individually customized information available with minimal effort and cost, the information that facilitates better decision making and makes the decision-making process more efficient. The arrival of search engines such as Google, Yahoo!, Bing, and AOL has significantly changed the way to forage for information. Millions of searchers regularly use modern search engines to find information on the Web for topics ranging from news to healthcare. Recent findings from Pew Internet show that almost half of all Internet users submit queries to and click on the results from search engines on a typical day [25].

One research stream in the area of information search focuses on user behavior of seeking for risk information. It attempts to understand how users seek for information related to some threats or hazards. Risk information seeking can be considered as part of individuals' reaction to uncertain risk and affects their consequent protection and mitigation actions [31, 68, 90]. Searching for healthcare information [29, 30, 80] and information security information [85] are two such examples that

draw attention from the research community. Compared with searches on other topics, users' search for healthcare and information security are generally considered to be protection-motivated with the purpose to assess, mitigate, or prevent some threat or risk that might directly affect or enhance one's wellbeing or computer assets [18, 28, 73, 86]. An understanding of risk information search could help promote the dissemination of risk information, raise the awareness of risk, and mitigate the impact of hazards. By exploring these two topics, we can better understand how users' information foraging behaviors might change in the context of different types of risks or hazards in their decisions ranging from making purchases of information security artifacts to finalizing treatment options for healthcare issues.

Using a search engine in pursuit of a search task, a user formulates and reformulates a series of queries. A search session can be defined from a contextual viewpoint as a series of interactions among a searcher, a Web system, and the content provided by that system within a specific period toward addressing a single information need [38]. During a search session, the user may take several actions including submitting a query, viewing results pages, clicking on URLs, viewing Web documents, and returning to the Web search engine for query reformulation. In a broad sense, the click is a searcher's or seeker's point of "meaningful connection with another person or with the world around" the seeker [10]. The goal for the user in a search session is to locate relevant information that addresses an information need [38].

Many researchers have analyzed search sessions with the goal of using the information about users' activities to improve the

* Corresponding author.

** Corresponding author at: Management Science and Systems, School of Management, State University of New York at Buffalo, United States.

E-mail addresses: jjwang@uta.edu (J. Wang), mgmtrao@buffalo.edu (H.R. Rao).

performance of Web search engines [38, 55, 74, 89]. The session level is considered to be the key for measuring the performance of search engines and understanding user behavior [38]. There are two main behavioral indicators reflecting users' interactivity with a search engine: (a) session length (i.e., the number of unique queries submitted by a searcher in a search session) and (b) query click rate (i.e., the average number of clicks per query in a search session), that are often explored. Session length and query click rate reflect the extent to which a user revises his query and interacts with the search engine for a task. However most prior investigations on these two indicators are descriptive in nature (see Table 1).

Guided by theories in information seeking, we explore (1) the extent to which users submit unique queries in a search session (2) how user search patterns in terms of session length and query click rate may change with search topics (information security, healthcare information, and other topics) and users' search engine activity level (i.e., users' tendency on a typical day to engage in searches with a search engine to find answers [16, 17]). We first draw from information foraging theory and employ the Inverse Gaussian (IG) distribution to characterize the number of queries per search session. As information seeking behavior can be considered as a function of task characteristics and individual characteristics [45], we propose that users' search pattern (in terms of session length and query click rate) is affected by search topics, users' search engine activity level, and their interaction. In order to test these suppositions, we model the mean parameter of an Inverse Gaussian distribution as the function of search topics, users' search engine activity level, and their interaction. We model the number of clicks in a session following a conditional Poisson distribution given session length, search topics, users' search engine

activity level, and their interactions. We test our hypotheses using a search log from a large search engine, spanning three months, for empirical validation. The search log allows us to better understand user search behavior without subjecting users to controlled lab experiments [89].

The contribution of our study is four fold. First, grounded in theories of information seeking, we find that Inverse Gaussian (IG) distribution provides a strong fit to the distribution of session length. Second, we quantitatively explore how a seeker's search behavior varies across search topics using a real-life search log. Most prior studies have either descriptively analyzed the popular search topics or terms [14, 37, 74, 78]; see, also, <http://www.google.com/trends>), or qualitatively compared the search behavior across different topics [6, 53, 79]. Our research also contributes to risk information seeking literature. We demonstrate the differences between the search for information security and healthcare while prior literature (e.g., Ng et al. [52]) considers both searches are similar in the sense that they are protection-motivated. Third, search engine activity level is an important user characteristic that can be observed based on users' historical behavior and further used to profile users. However, the question of how search engine activity level impacts search behavior in terms of session length and query click rate, for a given topic, is an open question that has not been explored before. We find users' search engine activity level is an important factor influencing user search behavior (in terms of session length and query click rate). Fourth we illustrate that there is an interaction effect between search engine activity level and search topics.

We organize this paper as follows. Section 2 reviews related literatures and introduces our research hypotheses. Section 3 describes the search

Table 1
Prior studies on session interactions.

Ref.	Log used	Session length	Query click rate
[69]	AltaVista search engine from August 1998 to 13 September 1998	The average number of queries per session is 2.02. 1 query (77.6%), 2 queries (13.5%), 3 queries (4.4%), >3 queries (4.5%)	N/A
[34]	Summary of multiple search engines including Fireball July 1998, Excite on March 10, 1997, and Alta Vista from 2 August to 13 September 1998	Average number of queries per user session for Excite: 1 query (67%), 2 queries (19%), 3 queries (7%), 4 queries (3%), more than 4 queries (4%) Average number of queries per user session for Alta Vista: 1 query (77.6%), 2 queries (13.5%), 3 queries (4.4%), more than 3 queries (4.5%)	Fireball: 59.51% sessions have 10 or less clicks, 40.47% sessions have >10 clicks; Excite: 58% sessions have 10 or less clicks, 42% sessions have >10 clicks; Alta Vista: 85.2% sessions have 10 or less clicks, 14.8% sessions have >10 clicks
[75]	Excite search engine on 16 September 1997	Average number of queries per user session is 4.86 Median number of queries per user session is 8 Average number of unique queries per user session is 2.52 Median number of unique queries per user session is 4	N/A
[75]	Excite search engine on 16 September 1997 and 20 December 1999.	Average number of queries per user session in 1997: 1 query (48.4%) 2 queries 20.8%; 3+ queries 30.8% Year of 1999: 1 query 60.4% 2 queries 19.8%; 3+ queries 19.8%	N/A
[74]	Excite search engine in September 1997, December 1999, and May 2001	Average number of queries per user session in 1997: 1 query (48.4%), 2 queries (60.4%), 3+ queries (55.4%); Average number of queries per user session in 1999: 1 query (20.8%), 2 queries (19.8%), 3+ queries (19.3%);2001: 1 query (30.8%),2 queries(19.8%), 3+ queries (25.3%);	N/A
[73]	Excite Web search engine collected in May 2001; FAST search engine Web queries submitted on February 6, 2001	Average number of queries per user session for Excite: 1 query (55.4%), 2 queries (19.3%), 3+ queries (25.3%) Average number of queries per user session for FAST: 1 query (53%), 2 queries (18.9%), 3+ queries (29%)	Click rate of Excite is 1.7 and click rate of FAST search engine is 2.2
[13]	Search engine of the Utah state government web site from March 1, 2003 to August 15, 2003	Mean number of queries per session 1.73 Median number of queries per session 1 Mean number of unique queries per session 1.25 Median number of unique queries per session 1	Average query click rate is 0.56
[35]	AlltheWeb.com on 6 February 2001 and 28 May 2002 submitted by European users	Average number of queries per user session in 2001: 1query (53%), 2 queries (18%), 3+ queries (29%) Average number of queries per user session in 2002: 1 query (59%), 2 queries(16%), 3+ queries 25%	Average number of clicks per session is 8.2
[35]	AltaVista search engine from 1998 and on September 8,2002	Average number of queries per user session in 1998: 1 query (77.6%), 2 queries (13.5%), 3+ queries (6.9%) Average number of queries per user session in 2002: 1 query (47.6%), 2 queries (20.4%), 3+ queries (32.0%)	N/A
[54]	NAVER search engine from 5 January to 11 January 2003	Average number of queries per session is 1.8 queries with a standard derivation of 2.03	N/A

متن کامل مقاله

دریافت فوری ←

ISIArticles

مرجع مقالات تخصصی ایران

- ✓ امکان دانلود نسخه تمام متن مقالات انگلیسی
- ✓ امکان دانلود نسخه ترجمه شده مقالات
- ✓ پذیرش سفارش ترجمه تخصصی
- ✓ امکان جستجو در آرشیو جامعی از صدها موضوع و هزاران مقاله
- ✓ امکان دانلود رایگان ۲ صفحه اول هر مقاله
- ✓ امکان پرداخت اینترنتی با کلیه کارت های عضو شتاب
- ✓ دانلود فوری مقاله پس از پرداخت آنلاین
- ✓ پشتیبانی کامل خرید با بهره مندی از سیستم هوشمند رهگیری سفارشات