

# Session-based Search Behavior in Naturalistic Settings for Learning-related Tasks

Souvick Ghosh

School of Communication and Information (SC&I)  
Rutgers University  
New Brunswick, NJ  
souvick.ghosh@rutgers.edu

Chirag Shah

Information School  
University of Washington  
Seattle, WA  
chirags@uw.edu

## ABSTRACT

In this research, we investigate the behavioral patterns exhibited in different search sessions as users attempt to complete search tasks of increasing cognitive complexity. The search tasks, which are exploratory in nature, have been designed using the Taxonomy of Educational Objectives, and are presented to the users hierarchically. We capture naturalistic search behavior of the users in real world (non-lab) setting using a Chrome browser plugin. The research analyzes the web log data of the users to assess if and how the web search behavior of the users changes over different search sessions. We also look at the different demographic factors like age and gender, educational factors like the academic background, read and write proficiency in English, and search skills to determine if these factors influence the web search behavior of the users. Our results indicate that search sessions have significant effects on the web search behavior of the users. Most of the web search behaviors differed significantly across search sessions. Of the secondary factors, gender showed significant effect on the query reformulations (measured using average edit distance between queries) and query length (measured using number of words per query) while year of study affected only the average query length. Search experience had significant effect on all the web search behaviors.

## KEYWORDS

Information seeking, Search Sessions, Naturalistic Search Environment, Search behavior, Task complexity

### ACM Reference Format:

Souvick Ghosh and Chirag Shah. 2019. Session-based Search Behavior in Naturalistic Settings for Learning-related Tasks. In *The 28th ACM International Conference on Information and Knowledge Management (CIKM '19)*, November 3–7, 2019, Beijing, China. ACM, New York, NY, USA, 4 pages. <https://doi.org/10.1145/3357384.3358098>

## 1 INTRODUCTION

In everyday world, we often encounter situations and experiences which do not fit into our existing knowledge structures. Such a

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).

CIKM '19, November 3–7, 2019, Beijing, China

© 2019 Copyright held by the owner/author(s). Publication rights licensed to Association for Computing Machinery.

ACM ISBN 978-1-4503-6976-3/19/11...\$15.00

<https://doi.org/10.1145/3357384.3358098>

gap (if our existing knowledge is insufficient) or anomaly (if the existing knowledge needs restructuring) is often a precursor to search activities, and the corresponding behavior can be explained as human information behavior [7, 18]. While human information behavior can be both active or passive, information seeking involves actively identifying relevant information sources to mitigate the knowledge gap.

For this research, we wanted to explore the information seeking activities and exhibited behavioral patterns in different search sessions when the users are presented with tasks of varying complexity levels. Task-based searching under laboratory settings is a popular approach in many information science researches. Task sessions are typically long and involves multiple queries as the user attempts to synergize the task and document contexts with his mental model. While researchers have investigated cross-sectional tasks interleaved with other tasks and sessions, there has been few studies which focus on the effect of task sessions in a naturalistic environment. We have used the Taxonomy of Educational Objectives [1, 2, 13] to design four tasks of varying cognitive complexity levels. However, in this research, we have not focused on the learning processes which occur during search tasks. Instead, we concentrate on how the different task sessions influence the web search behavior of the participants. We also explore the influence of different demographic factors like age and gender, educational factors like the academic background, the users' reading and writing proficiency in English, and search skills on the web search behavior of the users.

The search sessions have not been separated by the search tasks. Instead, we view the user study as an overall search task (whose description has been provided in Table 1) which contains four sub-tasks. Ideally, Task 1 would begin with Session 1 and Task 4 would end with Session n, where n is the maximum number of search session for a specific user. Any two search sessions which are separated by a time gap of more than 30 minutes are considered different. As the search tasks are designed in a hierarchical order, on a single topic, we envisage that as users complete the tasks during the course of the user study, we would observe visible patterns specific to the search sessions.

In this paper, we have tried to answer the following two research questions:

RQ1: *How does different sessions influence the web search behavior of the users when presented with tasks of different complexity levels, in a hierarchical order?*

RQ2: *How does different demographic (age and gender), educational (field of study, year of study, reading and writing proficiency in English), and experiential factors (search skills, and experience) influence the web-search behavior in different users?*

The rest of the paper is arranged as follows: Section 2 discusses the Background and Related work, and Section 3 explains the experimental methodology. Section 4 explains the data collection while Section 5 analyzes the results. The conclusion and future work have been discussed in Section 6.

## 2 BACKGROUND AND RELATED WORKS

Web search behavior has been investigated in many research studies, but from various perspectives. While some of them focus on the task sessions, others concentrate on task complexity. A common approach to exploring web search behavior is either through careful feature engineering of browser log data or through observation and interview of the participants engaged in search tasks.

Several research works ([15], [4], [12], [10], [9], [17]) focused on understanding search behavior in different task types and argued that different facets of a search task – task type, the task product, task complexity and task level (document vs segment) – strongly influence the search behavior of the information seeker. Borlund (2003) [3] explored how simulated work task situations can be used for evaluation of IIR models. These simulated situations are developed using realistic scenarios and call for alternative measures of performance. It only helps in judging situational relevance but can be used for both real-world interpretation and control. However, such tasks and evaluation metrics should be developed with the target user in mind. Ghosh et al. (2018) [8], Brennan et al (2014) [4], and Kelly et al. (2015) [12] designed the search tasks on different levels of cognitive complexity and tailored the situations to their users. The above researches indicate significant correlations between cognitive ability, task complexity, task workload, and search behaviors.

It is common for researches to focus on the search patterns for different search tasks [6, 14, 15] or search sessions [10, 11]. While more studies have concentrated on task-based evaluation, session based evaluation could provide some useful insights to similar work-task scenarios. Jiang and Ni (2016) [11] explored how different factors affect the users' query reformulation strategies during task-based search sessions, while Jiang et al. (2014) [10] focused on the influence of session time on the search behavior of users. Other studies analyzed the impact of domain expertise on search behavior [16], or evaluated the usefulness of online and offline metrics in assessing web search behavior [5].

## 3 EXPERIMENTAL METHODOLOGY

While the details of experimental methodology and data collection can be found in our previous work ([8], [17]), we provide a brief overview in the following sections.

### 3.1 Participants

We recruited a total of 31 undergraduate students, including 10 males and 21 females. There were 10 sophomores, 8 juniors, and 13 seniors, while freshmen were excluded from this study. The academic background of the students was diverse, with 18 students from STEM and 13 from Arts and Humanities. Only two of the participants identified themselves as non-native English speakers. The proficiency scores of the participants in reading and writing English were 4.8 and 4.6 respectively, and their average search

expertise score was 4.16, where all scores were measured on a 5-point Likert scale (1=Novice, 5=Expert).

The study was conducted for a period of two weeks, during which the participants were instructed to perform four search tasks, all of which were based on a common topic and were related to learning. The total time spent to complete this study was five hours on average, as reported by the participants. An exit interview was conducted at the end of the study.

### 3.2 Task Design

Our tasks have been designed using the Taxonomy of Educational Objectives [1, 2, 13] which was developed to promote thinking, and not in-rote learning in classrooms. The cognitive process dimension consists of *Remember*, *Understand*, *Apply*, *Evaluate*, *Analyze*, and *Create* levels, *Remember* being the lowest level and *Create* being the highest level. While the lowest levels involve additions of new information to our existing knowledge structures, the middle levels involve reshaping and accommodating knowledge, while the highest levels may require deleting and restructuring pre-existing knowledge. The importance of the taxonomy lies in the fact that it views educational goals as not only recalling facts but creating a more comprehensive understanding and application of the learned concepts. The design of the task questions was based on the search task development work by Wu et al. (2012) [19] and Kelly et al. (2015) [12].

### 3.3 Tasks

Overall Task Assignment:

The participants were asked to visualize themselves as leaders of student organization and perform four tasks [8]. They were instructed to write a brief report on cyberbullying which will inform schoolteachers, educators, and decision makers while designing new policies and strategies to tackle cyberbullying.

- (1) Task 1 (Remember and Understand): What is cyber bullying? How is it similar or different to other types of harassments (e.g. cyberbullying vs. traditional bullying)? What are some long-term/short-term risks faced with cyberbullying?
- (2) Task 2 (Apply): In 2010, Rutgers University has witnessed the tragic incident of Tyler Clementi, whose case raised concerns about cyberbullying. Find out more about this case, and possibly some other cases. What does this case(s) show you about some common characteristics of cyberbullying?
- (3) Task 3 (Analyze): Having heard some of the recent reports on cyberbullying, what seems to be the main cause of the bullying behavior online? How much are technology and use of electronic communication associated with cyber bullying? Why?
- (4) Task 4 (Evaluate): How effective are some of the currently available strategies to mitigate cyberbullying at schools and university campuses? Why? Which strategy/method do you think is best and why?

## 4 DATA COLLECTION

In this study, we collected a wide array of data – demographic information, web log data, pre- and post-search questionnaires, online diary, the final report, and the exit interview. However, for

this research, we will be using the following types of information: (1) Demographic Data: The background information about the participants which included demographic details like age and gender, their academic background, web search experiences and language proficiency; and (2) Web search data: The online behavior of the participants was captured using the Chrome browser plugin Coagmento. Once enabled (only during tasks), the plugin collected information about the search queries, web pages visited and saved, and session information along with timestamps. We also cleaned the data in consultation with the participants to make sure that it contained search results relevant to the tasks. This enabled us to assess the query reformulation and diversity patterns, dwell time, page visit patterns, and session-specific search behavior of the participants.

## 5 DATA ANALYSIS

The descriptive statistics of the predictor and outcome variables have been presented in Table 1 and 2. Both the tables report the values for the mean, standard deviation, and the Shapiro-Wilk score. All the  $p$ -values were less than 0.05. Query diversity helps us assess the uniqueness of the query while average edit distance is the measure of query reformulations. All the feature means were obtained by aggregating them over the particular search session. The Shapiro-Wilk test was used to see if a distribution of data points significantly differs from a normal. The assumption of normality was violated for all the variables (as  $p < 0.05$ ).

### 5.1 Effect of Search Sessions (RQ1)

**Table 1: Predictor Variables**

Variable	Mean	Std. Dev.	W
Age	21.03	3.03	0.74
Search Skills	4.12	0.68	0.79
Search exp.(in yrs.)	9.52	4.31	0.90
No. of searches/day	8.90	2.13	0.66
Read proficiency	4.81	0.39	0.48
Write Proficiency	4.65	0.48	0.61
Search Sessions	6.5	3.45	0.94

**Table 2: Outcome Variables.**

Variable	Mean	Std. Dev.	W
No. of pages	12.62	28.40	0.51
No. of unique pages	9.77	21.26	0.52
Dwell Time	120.58	1289.06	0.06
Avg. Dwell Time	10.35	75.37	0.11
No. of Queries	1.90	3.39	0.61
Query Diversity	5.08	8.75	0.62
No. of words	2.16	3.13	0.66
Edit Distance	12.89	28.61	0.42

As there are more than one outcome variables, we used multivariate analysis of variance (MANOVA) for statistical analysis. MANOVA not only helps in comparing the different groups (based on one or more input variables) but also in determining the relationship between the outcome variables. For answering our first

research question, we have grouped our data based on the session to which they belong. There was a maximum of 12 session for the task. Hence, we have created 12 groups based on the session information. To test the assumption of Multivariate Normality, we grouped the data by sessions and checked if the dependent variables collectively had multivariate normality between groups. The results of the test are reported in Table 3. All the  $p$ -values were less than 0.05, so, we perform a robust analysis of our data.

The interaction between the outcome variables and the search session can be observed in Table 4 where we compare the means of the outcome variables for the different sessions. This table shows how the mean of predictor variables vary between sessions. It provides an insight on which variables are influenced by the search sessions. To perform our robust multivariate analysis, we have used Choi and Marden's robust test, which is an extension of Kruskal-Wallis test and based on the ranked data. From the output of the test, we can say that  $H(88) = 889.85$ ,  $p < 0.05$ . This suggests that there were significant effects of the session on the web-search behavior of the participants. We used orthogonal contrasts to differentiate each session with the rest of the sessions. Our results show that there is significant difference ( $p < 0.05$ ) between sessions 1 and the rest, session 2 and the rest, and session 3 and the rest in the number of pages, the number of unique pages visited, and the number of queries. For query diversity, each of the sessions 1, 2, 3, 4, 5, and 7 differed significantly from the rest of the sessions. Similarly, sessions 1, 2, 3, 4, 5 were each different from the rest in the average number of words used in the query. For the average edit distance (which measures query reformulations), each of the sessions 1, 2, 3, and 4 were unique and significant differences were observed against the rest. Univariate analysis confirmed that all the outcome variables except dwell time and average dwell time differ significantly with changing sessions.

### 5.2 Influence of Secondary Factors (RQ2)

To explore the effects of the predictor variables on the outcome, we have performed one-way ANOVA using one predictor variable at a time. As both the assumptions of normality and homogeneity were violated, we performed a robust variation of ANOVA. We noticed that age and search experience in years have significant influence on all the outcome variables (the web search behaviors). We also observe significant effect of gender on number of words/query and average edit distance, and year of study on number of words/query. All significant effects have been measured at  $p < 0.05$ .

For all the predictor-outcome pairs which show significant effects, post-hoc tests using Bonferroni methods confirm that search experiences above and below 12 years produce significant differences in query behavior (number of queries, number of words/query, query diversity, and average edit distance) and page behavior (number of unique pages). However, for age ranges, we observed no difference in the web search behavior of the users. Males and females differed significantly in the number of words per query and the average edit distance between successive queries. Similarly, juniors differed significantly from seniors and sophomores in terms of number of words per query. Once again, difference of means was considered significant only for  $p < 0.05$ .

**Table 3: Multivariate Shapiro-Wilk test for normality.**

Shapiro-Wilk	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12
W-value	0.25	0.36	0.21	0.42	0.29	0.26	0.29	0.25	0.18	0.20	0.18	0.21
If p-value <0.05?	Yes											

**Table 4: Group mean of outcome variables for each search session.**

Variable	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12
No. of pages	43.9	17.3	14.5	11.9	11.6	5.6	11.7	6.6	6.2	11.9	9.8	0.4
No. of uniq. pg	34.5	14.2	10.6	9.3	9.4	4.4	9.7	4.6	4.4	8.3	7.5	0.3
Dwell Time	790.1	29.9	69.2	71.2	29.5	16.0	6.2	74.9	21.1	26.1	9.3	8.4
Avg. Dwell T.	57.5	6.2	26.5	8.5	5.9	7.8	0.3	3.7	3.4	2.9	0.2	1.4
No. of Queries	4.2	3.8	2.8	2.2	2.6	1.1	2.2	0.9	0.7	1.0	1.0	0.1
Query Div.	11.0	8.6	7.1	6.6	7.9	3.0	4.7	2.2	2.3	3.5	3.6	0.5
No. of words	3.2	3.1	3.8	3.5	2.9	1.9	1.7	1.4	1.7	1.3	0.8	0.5
Edit Distance	35.1	17.9	23.5	20.0	15.9	11.5	7.4	4.6	5.6	7.9	3.0	2.4

## 6 CONCLUSION AND FUTURE WORK

To answer our first research question, that is, if and how sessions influence the web search behavior of the participants, we conducted a MANOVA on the ranked data and observed significant effects of the session on the web-search behavior of the participants,  $H(88) = 889.85, p < 0.05$ . Univariate analysis confirmed that all the outcome variables except dwell time and average dwell time differ significantly with changing sessions. Our second research question gave some insight on the potential factors which could influence the web search behavior of users apart from the session number. Age and search experience produced significant effects on all outcome variables (although no significant differences were observed between different age groups). Gender showed significant effect on the average edit distance (query reformulations) and number of words per query (query length) while year of study affected only the average query length. To establish the reliability of these results, further analysis needs to be done using more users and different task topics. Using data collected in a naturalistic setting, this research investigates the differences in users’ web search behavior for different search sessions. The findings of this study should help in better understanding the influence of search sessions, demographic, and academic factors on the users’ web search behavior. It should also inspire the community to design better ranking algorithms and to promote user satisfaction and more personalized search experiences.

## 7 ACKNOWLEDGEMENT

This work is supported by the National Science Foundation (NSF) grant IIS-1717488.

## REFERENCES

[1] Lorin W Anderson, David R Krathwohl, Peter W Airasian, Kathleen A Cruikshank, Richard E Mayer, Paul R Pintrich, James Rath, and Merlin C Wittrock. 2001. A taxonomy for learning, teaching, and assessing: A revision of Bloom’s taxonomy of educational objectives, abridged edition. *White Plains, NY: Longman* (2001).

[2] Benjamin S Bloom, DR Krathwohl, and BB Masia. 1956. *Taxonomy of Educational Objectives* (New York, McKay). *Google Scholar* (1956).

[3] Pia Borlund. 2003. The IIR evaluation model: a framework for evaluation of interactive information retrieval systems. *Information Research. An International Electronic Journal* 8, 3 (2003).

[4] Kathy Brennan, Diane Kelly, and Jaime Arguello. 2014. The effect of cognitive abilities on information search for tasks of varying levels of complexity. In *proceedings of the 5th information interaction in context symposium*. ACM, 165–174.

[5] Ye Chen, Ke Zhou, Yiqun Liu, Min Zhang, and Shaoping Ma. 2017. Meta-evaluation of online and offline web search evaluation metrics. (2017).

[6] Michael J Cole, Chathra Hendahewa, Nicholas J Belkin, and Chirag Shah. 2015. User activity patterns during information search. *ACM Transactions on Information Systems (TOIS)* 33, 1 (2015), 1.

[7] Brenda Dervin. 1992. From the mind’s eye of the user: The sense-making qualitative-quantitative methodology. *Qualitative research in information management* 9 (1992), 61–84.

[8] Souvick Ghosh, Manasa Rath, and Chirag Shah. 2018. Searching as learning: Exploring search behavior and learning outcomes in learning-related tasks. In *Proceedings of the 2018 Conference on Human Information Interaction & Retrieval*. ACM, 22–31.

[9] Souvick Ghosh and Chirag Shah. 2017. Information seeking in learning-oriented search. *Proceedings of the Association for Information Science and Technology* 54, 1 (2017), 682–684.

[10] Jiepu Jiang, Daqing He, and James Allan. 2014. Searching, browsing, and clicking in a search session: changes in user behavior by task and over time. In *Proceedings of the 37th international ACM SIGIR conference on Research & development in information retrieval*. ACM, 607–616.

[11] Jiepu Jiang and Chaoqun Ni. 2016. What Affects Word Changes in Query Reformulation During a Task-based Search Session?. In *Proceedings of the 2016 ACM on conference on human information interaction and retrieval*. ACM, 111–120.

[12] Diane Kelly, Jaime Arguello, Ashlee Edwards, and Wan-ching Wu. 2015. Development and evaluation of search tasks for IIR experiments using a cognitive complexity framework. In *Proceedings of the 2015 international conference on the theory of information retrieval*. ACM, 101–110.

[13] David R Krathwohl. 2002. A revision of Bloom’s taxonomy: An overview. *Theory into practice* 41, 4 (2002), 212–218.

[14] Chang Liu. 2012. *Personalizing information retrieval using interaction behaviors in search sessions in different types of tasks*. Rutgers The State University of New Jersey-New Brunswick.

[15] Jingjing Liu, Michael J Cole, Chang Liu, Ralf Bierig, Jacek Gwizdka, Nicholas J Belkin, Jun Zhang, and Xiangmin Zhang. 2010. Search behaviors in different task types. In *Proceedings of the 10th annual joint conference on Digital libraries*. ACM, 69–78.

[16] Sophie Monchaux, Franck Amadiou, Aline Chevalier, and Claudette Mariné. 2015. Query strategies during information searching: Effects of prior domain knowledge and complexity of the information problems to be solved. *Information Processing & Management* 51, 5 (2015), 557–569.

[17] Manasa Rath, Souvick Ghosh, and Chirag Shah. 2018. Exploring online and offline search behavior based on the varying task complexity. In *Proceedings of the 2018 Conference on Human Information Interaction & Retrieval*. ACM, 285–288.

[18] Thomas D Wilson. 2000. Human information behavior. *Informing science* 3, 2 (2000), 49–56.

[19] Wan-Ching Wu, Diane Kelly, Ashlee Edwards, and Jaime Arguello. 2012. Grannies, tanning beds, tattoos and NASCAR: Evaluation of search tasks with varying levels of cognitive complexity. In *Proceedings of the 4th Information Interaction in Context Symposium*. ACM, 254–257.