Understanding Search Behavior of Low-Literacy & Low English Comprehension Users

Naman K. Gupta, Carolyn P. Rose

Language Technologies Institute, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15213, United States of America.
{nkgupta, cprose}@cs.cmu.edu

Abstract. This paper presents a user study based log analysis of the search behavior of low-literacy users in the developing world. It is first step towards Web search personalization for such users, which has been hindered by a paucity of appropriate data for inducing effective user models that target the real problems in information access for low literacy user population. We demonstrate the variations in the behavior of such users with the 'typical' user population by comparing the user actions during an Information Seeking session. The users deviated from the 'ideal' search behavior during the course of the information-seeking session with inefficient use of querying and subsequent navigation of the search results. We present an analysis that offers intuitions for more efficient information seeking practices for such users.

Keywords: Personalization, Low-literacy Users, Web-log Analysis, Information Seeking Task, Search strategies.

1 Introduction

Currently, Internet penetration [1] in the developed world has reached a point that much of the growth in Internet users will most probably be from the developing world, where a large majority of users, particularly from rural regions, are low literacy users. And Internet is going to be the defining Technology for Literacy and Learning [2, 3, 4] for such users. The needs of these new users are very different from the great majority of Internet users today [5], who are largely capable of using current search technology to meet their information seeking needs.

Imagine a student from a rural area in India or Africa with limited web experience and limited education level, or a foreign student with low English comprehension - just entering a school or university environment in the United States. For these users, the experience of using search technology is quite different than the effortless experience many of us have every day. For such users, their low comprehension of the language may act as a hindrance in formulating an effective query phrase. Even if relevant information is provided in response to their query, they may or may not recognize it as such. Long lists of search results may be overwhelming to them.

Many of the current models in the area of probabilistic retrieval [6], which are embedded in the popular search engines, build in the assumption that users are able to distinguish effectively between relevant and non-relevant documents by examining the text around the links that have been provided in response to their query, that they click on those links that meet their needs best, and that the search ends when they have found what they are looking for. We challenge all of these assumptions when dealing with inexperienced and low literacy populations, and show significant variations in search behavior with experienced and high-literacy users. We create models based on user search actions during the Information Seeking session to highlight the variations across the course of the session. This posits the need for a targeted effort to assist such populations to search efficiently and effectively. Understanding the search strategies and needs for support of such populations is unchartered territory, and arguably essential at this time as Internet penetration continues to expand into developing regions. The analysis of data presented in this paper contributes towards understanding the issues faced by this emerging market.

In the emerging area of personalization for Information seeking systems, significant progress has been made towards adapting the behavior of these technologies to the specific needs of particular user groups [7, 8, 9, 10]. Most of the personalization work has concentrated on effect of domain expertise and search experience on search behavior. In this paper, we concentrate on different user population of low-literacy level and low English Comprehension. Also, the field of personalization research has been hindered by a paucity of appropriate data for inducing effective user models that target the real problems in information access for needy populations, such as low literacy users. Fortunately, we have access to a large population of users who fit into our target user population, who have recently become part of a community where they have access to technology and support for their English communication skills. Our goal is to understand the search behavior of such users and to identify specific actions where they might be inefficient with respect to finding relevant Information. Then to negate the inefficient behavior either by means of providing a personalize search interaction or suggesting specific search behavior which fits particularly well for such users. But we are still relatively early in the process.

In the remainder of the paper we discuss related work in Modeling Search Behavior and Personalization. Then we present the experimental study we ran as part of this effort in Section 3. Section 4 gives the Search Log data description and describes the data analysis methods. In Section 5, we present the results that confirm our suspicion that the assumptions underlying current probabilistic models of search behavior are not valid for our target user population. Section 6 concludes with our intuition for more efficient information seeking practices of such users.

2 Related Work

Research on charactering user search behavior and building user models to personalize search is relevant to our work. Lot of work has been done for modeling the search behavior for 'typical' web search users. Agichtein et. al. [8] presented a generic user interaction model for the task of predicting web search preferences by observing their post-search behavior, incorporating click-through, browsing and query features. While their model improves performance with respect to search results relevant for the typical user, it is not applicable for users who do not have extensive web experience. Holscher et al. [7] showed the heterogeneous needs and capabilities of search-engine users, which have to be catered differently. They showed the ineffectiveness in query reformulation and navigation strategies of novice users during information seeking activities. We argue a similar inefficiency in Search behavior of low-literacy users.

Most such work [8, 12] involves analysis of large-scale query logs. But identifying query session for a particular information need is hard task in such logs. For the task of segregating queries for a specific information need, Metzler et al. [11] computed the similarity between two queries using lexical, stemming and probabilistic modeling methods. Downey's [12] work highlights the importance of applying more complex techniques for task of identifying session boundaries in query logs.

Teevan et al. [13] emphasized that searches are just the starting point for richer information interactions that evolve over the course of session. They highlight the distinction between the User's information need and his articulation of that need into a query. Downey [12] seeks to understand the relationship between the articulation of a goal (represented by the query) and the estimated information need (represented by the last URL visited or with most dwell time, during a search session). They also investigate search behavior across rare and common (frequency based estimation) queries, as well as target URLs. In our study, we present a well-defined information-seeking Task, which allows us to evaluate the articulation of the information need by analyzing the subsequent queries issued.

Zhang et. al [14] evaluated the relationship between domain knowledge, search behavior and search success. The performance of search specialist, domain specialist and novices was also compared in Marchionini et.al [15]. Most such research has shown differences in search behavior and success as a function of domain expertise and search experience. In this paper, we aim to understand the behavioral differences between low-literacy & low English comprehension users and more experienced educated users.

Kelly et al. [16] explains the difference between basic IR systems and information seeking support systems (ISSSs) and in particular highlights the need to build separate evaluation models for ISSSs, independent of the insufficient current models for basic IR evaluations. In IR systems, a single query (possibly reformulated) represents the users information need, but in ISSSs, users engage in multiple search session where they may enter many queries and review corresponding results. They suggest a need for more longitudinal designs that observe these activities over sustained periods and analyze the process of the information gathering. We define a similar information-seeking task, which elicits an extended search session spanning multiple queries with a more involved information need.

As highlighted above, most previous research has shown differences in search behavior and success as a function of domain expertise and search experience. In this paper, we aim to understand the behavioral differences between low-literacy & low English Comprehension Users and more Experienced Educated Users. Also most of the research, done on large-scale query logs, make an implicit assumption about the information need and also have unreliable estimates of a query session. In our research, we conduct a study with a well-defined information-seeking task to be completed in limited time duration. At the same, it has a large user base with 300 participants, which is much higher than similar user studies done in related research in Personalization [7, 17, 18].

3 Experiment

We conducted a large-scale user study with an elaborate information-seeking task, which provided us a well-defined search behavior and activity session.

3.1 Study Participants

We conducted this user study with 300 participants. They are college undergraduates with English as a 2^{nd} Language from rural areas in India, characteristic of our low-literacy target user population.

3.2 Experimental Procedure

The study was conducted in 6 sessions with 50 participants each over a period of 2 days. Each session extended for 2 hours. Initially the experimenter, giving a short self-introduction, explained the purpose and motivation behind the study. Then a brief walkthrough of the study was given to the participants. The experiment survey extended for 1 hour and 10 minutes duration: 10 minutes for completing a background information questionnaire, 10 minutes for installing a Search Activity Logging Toolbar and other browser configurations, 20 minutes for understanding the information seeking task and completing the Pre-search write-up. They were then given another 30 minutes for the Search activity and subsequent Task write-up. Once finishing the survey, the participants uploaded the log files recorded by the toolbar using the toolbar itself, and subsequently uninstalled it.

3.3 Experimental Task

The Experimental task itself was an exploratory information-seeking task based on the characteristics defined in [16, 19]. It had the following template:

Imagine that you are a new professor assigned to teach the course <*familiar/ unfamiliar Course Name*> for the first time to 11^{th} grade students, and you want to make sure the content is up-to-date with the latest <*technology/ literature*>. The specific topics you will be focusing on are < *Broad Topic/ Less Broad Topic/ Specific Topic*>. Write a brief content summary for the course curriculum with reference books to be followed during the course.

The slots in the template were filled in differently for each condition based on the experimental manipulation described below. To ensure their understanding of the Task statement, the Participants were asked to mention the characteristics of the students, which seem relevant to them for their assigned search task: Age, Gender, Educational Background, Medium of Instruction in School, Experience with Computers, Experience with Internet/Searching, Personal Interests, Others factors.

Before accessing any information online, they were asked to prepare a Pre-search write-up based on prior knowledge. Then using any search engine, they were told to prepare a Post-search write-up having all the information required for the given search task. No distinction was made in what search engine was used.

3.4 Experimental Manipulation

The difficulty of an information-seeking task is expected to have an effect on search strategy and task success. We operationalized the task difficulty as a combination of how familiar the topic is – Topic Familiarity, and what the level of specificity is with which the information need is formulated – Specificity. The experiment was a 3X2 factorial design, where the Specificity is a 3 level between subject factors – High, Medium, Low and the Topic Familiarity is 2 level between subject factors – High, Low. This design, as defined in Table 1, allows us to avoid order effects and

Table 1. 3x2 Factorial Design with Specificity and Topic Familiarity as variables.

3x2		**Topic Familiarity**	
		Low	**High**
Specificity	**Low**	Any 2-3 topics on Foundational Computer Science (1A)	Any 2-3 Topics on World History in the 20^{th} Century (2A)
	Medium	Broad Topics - Computer Hardware and Operating System (1B)	Broad Topics – World Wars and US-Russian Cold War (2B)
	High	Blue Ray discs and Unix Operating System (1C)	Watergate Scandal and Collapse of Soviet Union (2C)

confounds from interaction between Topic and Specificity. These 6 variations were defined in the 6 Experiment sessions with 50 participants each. The instructions across all the 6 sessions were same just the necessary variations in the Information-seeking task statement according to the above factors.

3.5 Tools and Materials

The following Tools and Materials were used for the experiment:

- A 4 page Web-based survey[1] designed using www.surveymonkey.com. The survey included following question types - Background Information, Instructions for Installing Logging Toolbar, Search Task statement, Pre-Search and Post-Search Write-ups and instructions for uploading Search activity logs.
- Firefox browser compatible with both Windows and Ubuntu systems was used for the experiment.
- Lemur Query Log Toolbar[2] was used to log all Search based activities performed during the Experiment.

4 Data Description and Processing

Completely-anonymized logs generated by the above toolbar were used along with the Survey Data. The details of the Log Data and Processing methods are as follows:

4.1 Data Description

Survey Data. We collected a total 360 survey responses over the 6 study sessions. This included spurious responses filtered out during Pre-Processing described below. These surveys contained the following details:

- Background Information – Unique ID, Type of High School, Medium of Instruction in School and University, Experience and Frequency with Computers, Frequency of using Search Engines.
- Student Characteristics deemed relevant for the Search Task by the Participants
- Pre-Search and Post-Search Write-ups
- Self-reported Topic Familiarity and Search Task Difficulty.

Activity and Search Log Data. We collected 280 Activity and search logs using the Lemur Toolbar. These logs contained the following event details:

- Search Related – Details (Query string, timestamp) of all queries issued. Details (Result rank, URL, timestamp) of results clicked from results
- Viewed Pages – Details (URLs, content, Time on Page, timestamp) of all the pages viewed.
- Browser Events – Details (RClick, Add/Close New Tab/Window, Copy, Scroll events) of any browser activity during the experiment. This allows us to build a sequence of events during the Search session.

---

[1] www.cs.cmu.edu/~nkgupta/SearchStudy/
[2] http://www.lemurproject.org/querylogtoolbar

Gold Standard Data. We collected 6 Survey and Search Logs, one for each of the 6 conditions from 6 high literacy graduate students at a top-tier US University.

4.2 Data Pre-Processing

The incomplete responses in the Survey data were removed giving a total of 305 responses. This further reduced to 296 responses after removing double submissions from some participants.

Out of these logs, only 200 logs had Search Related information. This might have happened in cases where people did not use any search engine in performing the task, used other search engines than the specified (Google, Bing, Yahoo). Some Participants used the default Firefox Welcome Google search page, which was not logged by the Toolbar.

4.3 Data Processing

For each user response including the Gold Standard responses, we build 4 different unigram Language models [6, 21] with commonly used Laplace Smoothing [20]. Language models capture the distribution of words used by a user or population. Language models can be compared using metrics that measure how different their associated word distributions are, and thus can be used to rank users according to how different or similar they are to the Gold Standard Users. The models were built from the page content viewed as a result of different user actions during the search session. We defined 4 basic types of User actions – Query (Q), ClickResults (R), ClickResults+Navigation (R+RN) and DirectPages+Navigation (D+DN). The description of the 4 models computed for each user as well as the Gold Standard users is as follows:

- AllSearchResultsModel – includes the content from all the top 10 search results returned in response to each of the queries issued by the user. This is to evaluate the relevance of the queries compared to the ones issued by Gold Standard Users.
- ClickedResultsModel – includes the content from all the results that were clicked by the user. This is to evaluate the user's ability to choose a relevant result from the results page.
- ClickedResults+NavModel – includes the content from the above clicked results pages along with the subsequent navigated pages. This is to evaluate the user's ability to effectively navigate through the Clicked results to find the relevant information.
- DirectPages+NavModel – includes the content from the Pages viewed directly and subsequent navigation. This is to evaluate the user's prior knowledge about a source of information about the Topic.

A variable referring to the names of these 4 models is referred to in the remainder of the paper as Model-Label. Language models for each user in a study session were compared using KL divergence [22] with corresponding Gold standard Language model for that session. KL divergence measures the difference between two distributions. In this context, it is used as a way of evaluating how similar the behavior of the user is to that of the corresponding Gold Standard User for the condition. Since the task defined in this user study is an elaborate Information seeking task, the usual methods of IR evaluation are not applicable here [16]. So we make an assumption about the 'perfect' search behavior for the highly educated and experienced Gold standard Users in our study. This is reasonable assumption in this context and it allows us to compare the Search behavior patterns in the two User populations and identify relevant variations during such an information-seeking task.

Subsequently, a more detailed analysis of the Interactions during the User Search session was done. The search session was divided into 3 equal parts in order to understand the impact of different user actions in each of these time partitions defined as – Early, Middle, Late. The above 4 models were created for each of the time partitions and compared to the previously built Gold Standard User models.

The above 4 user action instances were also divided by the length of time spent on the pages corresponding to the action instances. They were divided into 3 partitions defined by the median and 3rd quartile of the time duration of all the actions instances in the study for each of the action types. Table 2 describes the partition time ranges. These partitions allow us to identify 'ideal' time duration of interaction for each of these actions as compared to the Gold Standard User models.

Table 2. Description of the partition ranges for time spent on user actions instances.

Action Type	1st Partition (sec)	2nd Partition (sec)	3rd Partition (sec)
Query (Q)	(0, 7.0]	(7.0, 16]	(16.0, above)
Results (R)	(0, 9.0]	(9.0, 23.0]	(23, above)
Results+Nav (R+RN)	(0, 9.0]	(9.0, 22.0]	(22.0, above)
Direct+Nav (D+DN)	(0, 9.5]	(9.5, 21.5]	(21.5, above)

5 Analysis and Discussion

The participants of the study were clustered, using K-means, into 4 groups based on their Search behavior information. The Search behavior for a participant was defined by the amount of time spent on each above defined User actions, % of his time spent on that particular action, number of such user action instances and % of instances spent on that particular action during its Search session. Table 3 describes the features of the 4 clusters.

Table 3. User Cluster Description

	Cluster 1 Queriers	Cluster 2 Directors	Cluster 3 Contemplators	Cluster 4 Navigators
Number of Queries	11.8 (6.5)^{A}	4.9 (4.7)^{B}	1.8 (3.2)^{C}	7.4 (5.6)^{D}
%age Time on Search Page	.27 (.17)^{A}	.13 (.26)^{B}	.16 (.16)^{B}	.15 (.11)^{B}
%age Time on Results Page	.39 (.14)^{A}	.09 (.11)^{B}	.4 (.16)^{A}	.69 (.15)^{C}
%age Time on Navigated from Results	.1 (.17)^{A}	.01 (.04)^{B}	.07 (.13)^{A}	.07 (.12)^{A}

%age Time on Direct URL Page	.22 (.14)[A]	.42 (.27)[B]	.34 (.17)[C]	.07 (.07)[D]

Cluster 1 users spent 66% of their time either entering queries or examining the results of their queries, so we have named them Queriers. They had the largest number of queries, which was significantly higher than that of any of the other clusters.

Cluster 2 users spent 75% of their time either directly navigating to pages or navigating from those pages, so we have named them Directors.

Cluster 3 users only entered an average of 1.8 queries in their whole session, which represented 16% of their time. However, they spent 40% of their time navigating from the results pages from these queries. Because of the concerted effort they placed into analyzing the results from those few queries, we have named them Contemplators. Like Directors, they also spent a fair amount of time (34%) on direct navigation.

Cluster 4 users spent the same percentage of their time on queries, but they did it faster, and thus they entered 7.4 queries on average, in contrast to the significantly lower 1.8 queries on average for Contemplators. They also spent a correspondingly significantly larger percentage of their time on navigation from these query results pages. Thus, we have named them Navigators.

Overall, Directors were the least successful in terms of our KLD based evaluation criterion. Their KLD score was marginally higher on average than that of each of the other clusters ($F(3,534) = 2.5$, $p = .06$). We find a significant interaction between Assessment and Cluster such that the difference between the Directors and the other users is only for the Query models ($F(9,534) = 4.7$, $p < .0001$), and not significant in the other cases. Thus, perhaps it is not surprising that Directors spend the majority of their time on direct navigation and navigating from those pages. It may be that they were frustrated by their unsuccessful query attempts and thus resorted to direct navigation instead.

We then replicated this analysis on 3 time partitions within the session – Early, Middle, Late as describes earlier. In the separate models, we computed for these time partitions, we see that although the interactions between Assessment and Cluster remain consistent across partitions, the main effect of cluster is significant for the 'Late' time partition and not significant for the first two. This gives some indication that Directors get more off track over the course of the session.

We also broke down our analysis by length of action. For each action type, we computed a histogram of times. The distribution was Zipfian and skewed towards shorter times. Therefore, we took median and below times to be short, median to 3rd quartile times to be medium, and larger than 3rd quartile times to be long. We then recomputed our models for this break down in terms of time length of actions. We wondered whether quicker actions would be less precise. And indeed, this was born out in our data. We found that the interaction between Cluster and Assessment was consistent with the results from the whole model for all of these time based models, but whereas there was no main effect of cluster for the models corresponding to long duration actions, there was a marginal effect for models corresponding to medium duration actions, and a significant effect for models associated with short duration actions. What we conclude from this was that Directors showed consistently poor query behavior. But their behavior in other respects was also poor on average in comparison with their counterparts in other clusters, particularly late in the session and while executing actions quickly.

We didn't find other differences in behavior across clusters to make predictions about the ultimate success of the actions in the context of the search task. Note that it is Cluster 1, the Queriers, that really stands out in terms of time spent focused on queries and query results, however the quality of their queries does not really stand out from that of their counterparts in other clusters. Thus, what we conclude is that the problematic behavior to look out for during a search session for this user population is direct navigation, since it appears that users who have trouble with their queries resort to this behavior, but are not able to use it effectively.

6 Conclusion and Future Work

In this paper we have presented an experimental study in which we have explored the specific needs of low literacy users in the developing world conducting a search task.

Our analysis a significant cluster of users was inefficient in Query formulation and subsequent navigation through Search results. And had to resort to Direct navigation which was again ineffective. Also users got more and more off track during the search session. Other support for distinguishing relevant information from irrelevant information may also be necessary.

This study is a pilot effort contributing some new insights towards modeling low literacy information seeking behavior on the web. In our current work we are preparing to conduct a even larger study with 6,000 users with even less computer experience and lower literacy than the users from this study.

Acknowledgments. We thank the administrative team at our collaborating university for managing the logistics of the Study.

References

1. Google User Statistics, http://www.comscore.com/Press_Events/Press_Releases/2009/1/Global_Internet_Audience_1_Billion
2. Leu, Donald J., Zawilinski, L., Castek, J., Banerjee, M.: What is New about New Literacies of Online Comprehension. In: Rush, S.Leslie, Eakle, A. Jonathan, Berger, A. (eds.) What Research Reveals for Classroom Practice, Secondary School Literacy.
3. Sutherland-Smith, W.: Weaving the literacy Web: Changes in reading from page to screen. In: The Reading Teacher, Vol. 55, No. 7, pp. 662-669 (2002)
4. Henry, L.A.: Information Search strategies on the Internet: A critical component of New Literacies. In Webology, Vol. 2, No. 1(2005)
5. Dass, R.: Study on Information Needs of Rural Consumers & Exploring Alternatives: A study in the Indian Context. Proc. of Intl. Conf. on System Sciences (2009)
6. Song, F., Croft, B.: A general language model for information retrieval. In Proc. ACM SIGIR, pp. 279-280 (1999)

7. Hölscher, C., Strube, G.: Web search behavior of Internet experts and newbies. Computer Networks: The International Journal of Computer and Telecommunications Networking, vol. 33 no. 1-6, pp. 337-346 (2000)
8. Agichtein, E., Brill, E., Dumais, S.T.: Improving web search ranking by incorporating user behavior. In Proceedings of SIGIR, 19-26 (2006)
9. Duggan, G.B., Payne, S.J.: Knowledge in the head and on the web: Using topic expertise to aid search. In Proc. SIGCHI, pp. 39-48 (2008)
10. White, R.W., Dumais, S.T., Teevan, J.: Characterizing the influence of Domain Expertise on Web Serach Behavior. In Proc. WSDM (2009)
11. Metzler, D., Dumais, S.T., Meek, C.: Similarity measures for short segments of text. In Proceedings of ECIR (2007)
12. Downey, D., Dumais, S.T., Liebling, D., Horvitz, E.: Understanding the relationship between searchers' queries and information goals. In Proceedings of CIKM 449-458 (2008)
13. Teevan, J., Alvarado, C., Ackerman, M.S., Karger, D.R.: The Perfect Search Engine Is Not Enough: A Study of Orienteering Behavior in Directed Search. In Proc. of ACM CHI (2004)
14. Zhang, X., Anghelescu, H.G.B., Yuan, X.: Domain Knowledge, search behavior and search effectiveness of engineering and science students. Inf. Res., 10(2), 217 (2005)
15. Marchionini, G., Lin, X., Dwiggins, S.: Effects of search and subject expertise on information seeking in hypertext environment. In Proc. ASIS, pp. 129-142 (1990)
16. Kelly, D., Dumais, S., Pedersen, J.: Evaluation challenges and directions for information seeking support systems. In IEEE Computer, pp. 60-66 (2009)
17. Bhavnani, S.K.: Strategy Hubs: Domain portals to help find comprehensive information. JASIST, 57(1), pp. 4-24 (2005)
18. Kelly, D., Cool, C.: The effects of topic familiarity on information search behavior. In Proc. JCDL, pp. 74-75 (2002)
19. Kules, B., Capra, R.: Designing Exploratory Search Tasks for User Studies of Information seeking supoort systems. 9th ACM/IEEE Joint Conference on Digital Libraries. JDCL (2009)
20. Zhai, C., Lafferty, J.: A study of smoothing methods for language models applied to ad hoc information retrieval. In Proc. ACM SIGIR (2001)
21. Chen, S.F., Goodman, J. An empirical study of smoothing techniques for language modeling. In Proc. ACL. pp 310-318 (1996)
22. Kulback, S.: The Kulback-Leibler distance. The American Statistician. 41: 340-341 (1987)