

The Effects of Topic Familiarity on Information Search Behavior

Diane Kelly
SCILS
Rutgers University
New Brunswick, NJ 08901 USA
+1 732 932 9547
diane@scils.rutgers.edu

Colleen Cool
GSLIS
Queens College, CUNY
Flushing, NY 11367 USA
+1 718 997 3788
ccool@qc.edu

ABSTRACT

We describe results from a preliminary investigation of the relationship between topic familiarity and information search behavior. Two types of information search behaviors are considered: reading time and efficacy. Our results indicate that as one's familiarity with a topic increases, one's searching efficacy increases and one's reading time decreases. These results suggest that it may be possible to infer topic familiarity from information search behavior.

Categories and Subject Descriptors

H.3.7 [Information Storage and Retrieval]: Digital Libraries – user issues. H.1.2 [Models and Principles]: human factors.

General Terms

Human Factors.

Keywords

Personalization, topic familiarity, user modeling, information seeking behavior.

1. INTRODUCTION

A challenge in digital libraries and IR, in general, is the development of effective techniques for model acquisition. In order to build a user model, the intermediary function, whether human or computer, must somehow elicit information from the user about his/her interests and knowledge. Human search intermediaries are good at eliciting knowledge from users about their information needs [c.f. 3]. Studies about the search intermediary and user interaction have demonstrated consistently that search intermediaries attempt to characterize users among numerous facets, including knowledge of the topic. Knowledge of the user's familiarity with a topic is used by the search intermediary to select appropriate retrieval strategies and information objects for individual users. The underlying assumption is that different objects will be appropriate for different users depending upon their level of familiarity with the topic. In a physical library with human search intermediaries, this type of personalization is possible, but how do we offer this in a digital library environment, where interactions are typically limited to user and computer?

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. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

JCDL '02, July 13-17, 2002, Portland, Oregon, USA.
Copyright 2000 ACM 1-58113-513-0/02/007...\$5.00.

In information filtering and retrieval, systems employing user modeling attempt to replicate some aspects of the user modeling function that are performed by human search intermediaries [c.f. 2]. These systems typically require the user to actively participate in the construction of the user model. However, there are a number of drawbacks with techniques for user model construction. For instance, many users are generally unwilling to explicitly participate in user model construction. Additionally, models are often incomplete in that they do not account for changes in the user's knowledge states. There has been substantial effort devoted to obtaining information for user modeling implicitly, primarily through the observation of "normal" human search behavior [c.f. 5, 6]. Behaviors most extensively investigated as sources for implicit feedback have been reading, saving and printing. It has been suggested that contextual factors, such as topic familiarity and task, affect the rate of occurrence of these behaviors [4]. While it is commonly acknowledged that topic familiarity is an important factor influencing information seeking, no one has suggested how to automatically capture knowledge about topic familiarity or how to use this to tailor retrieval. Our current work focuses on identifying information search behaviors that might be directly related to topic familiarity. We begin our investigation with the examination of two such behaviors, reading time and efficacy.

2. METHOD

A secondary analysis of data was employed for the method of this study. The data for this study was extracted from trace files generated during a TREC-8 Interactive Track Study [1]. We include only a brief overview of the method for this study in this report. For a more detailed review, interested readers are referred to [1].

2.1 Participants & Procedures

A total of 36 volunteers, recruited from the university community, participated in the original study. The general search task required participants to find, label and save documents that identified different examples of the type of information specified by 6 topics. Participants were allotted 20 minutes per topic. The experimental retrieval system used Inquery 3.1.1 with its default values for indexing and retrieval [7]. The searching was performed on a database consisting of the full text of articles in the Financial Times during the years 1991 – 1994. All searches were logged.

2.2 Measures

Participants completed post-search questionnaires, which assessed participants' *familiarity* with each search topic on a 1 to 5 point Likert scale, where 1 was not at all, 3 was somewhat and 5 was extremely familiar. *Time* spent reading a document was derived from the search logs. *Efficacy* was measured as the ratio of the number of saved documents to the total number of viewed docu-

ments. Thus, higher efficacy scores indicate more efficient searching.

3. RESULTS

The means and standard deviations for time spent reading documents with respect to familiarity are displayed in Table 1. On average, participants spent less time reading documents retrieved for topics for which participants had a high degree of familiarity and more time reading those for which participants held a low degree of familiarity. Participants spent, on average, a similar amount of time reading documents for the topics for which familiarity level was 2, 3, and 4. A one-way ANOVA was conducted to evaluate the relationship between reading time and topic familiarity. The differences between these means was not significant.

Table 1. Means and Standard Deviations for Time and Efficacy with respect to Topic Familiarity.

Familiarity Level	Time (seconds)		Efficacy	
	M	SD	M	SD
1	23.46	10.71	.32	.15
2	21.50	8.95	.38	.18
3	20.90	11.24	.42	.21
4	21.44	15.34	.45	.16
5	16.57	7.32	.43	.15

The means and standard deviations for efficacy with respect to familiarity are also displayed in Table 1. For most cases, the more familiar participants were with the topic, the higher the efficacy score. A one-way ANOVA was conducted to evaluate the relationship between efficacy and familiarity. The differences between the mean efficacy and familiarity was significant, $F(4, 206)=2.9$, $p<.05$. A Dunnett's C follow-up test was conducted to evaluate pair-wise differences among the means. The Dunnett's C test identified significant differences between the mean efficacy scores for familiarity of 1 and 3, 1 and 4, and 1 and 5.

4. DISCUSSION & CONCLUSIONS

The general results from our study indicate that some information search behaviors, like reading time and efficacy vary with respect to topic familiarity. Although we found no statistically significant difference between reading time for topic familiarity, the general trend suggests that the more familiar one is with a search topic, the less time one spends reading. The mean time spent on topics where participants rated their familiarity as 2, 3 or 4 was similar (21.50, 20.90, and 21.44, respectively). The mean time for participants whose familiarity was 1 was higher (23.46), while the mean time for those whose familiarity was 5 was somewhat lower (16.57). Efficacy was found to increase with topic familiarity; the more familiar participants were with a topic, the more efficient their searching. Post-hoc tests found significant differences be-

tween efficacy and familiarity of 1 and 3, 1 and 4 and 1 and 5. We understand that the scale which was used to measure familiarity may not be calibrated optimally for this type of application. However, these results are encouraging, for they suggest that some differences in reading time for topic familiarity exist.

We acknowledge several limitations to our study. For instance, we understand that there are several variables that effect reading time which we have not taken into account such as individual reading speeds. Our study was a controlled laboratory experiment where participants were required to complete search tasks in a specified time frame. In more naturalistic studies, valid methods of measuring reading time may not be as obvious. Our next step is to develop appropriate methods for measuring reading time and topic familiarity in natural searching contexts. The ultimate goal is to use information search behaviors to infer topic familiarity. We believe that this can lead to personalized search interactions and tailored retrieval in digital library environments.

5. ACKNOWLEDGMENTS

We thank both the members and participants of the Rutgers TREC-8 research project. This study was funded in part by NSF Grant #99-11942.

6. REFERENCES

1. Belkin, N. J., Cool, C., Head, J., Jeng, J., Kelly, D., Lin, S., Park, S. Y., Savage-Knepshield, P., & Sikora, C. (2000). Relevance feedback versus local context analysis as term-suggestion devices: Rutgers' TREC-8 interactive track experience. In D. Harman, & E. Voorhees (Eds.), *TREC-8, Proceedings of the Eighth Text Retrieval Conference*. Washington, D. C.: NIST, 565-574.
2. Croft, W. B. & Thompson, R. H. (1987). I³R: A new approach to the design of document retrieval systems. *Journal of the American Society for Information Science*, 38, 389-404.
3. Ingwersen, P. (1982). Search procedures in the library analyzed from the cognitive point of view. *Journal of Documentation*, 38, 165-191.
4. Kelly, D. & Belkin, N.J. (2002). Reading time as an implicit measure of relevance in an information seeking task. Manuscript submitted for publication.
5. Konstan, J.A., Miller, B.N., Maltz, D., Herlocker, J.L., Gordon, L.R., & Riedl, J. (1997). GroupLens: Applying collaborative filtering to usenet news. *Communications of the ACM*, 40(7), 77-87.
6. Morita, M. & Shinoda, Y. (1994). Information filtering based on user behavior analysis and best match text retrieval. In *Proceedings SIGIR*, (Dublin, Ireland, July, 1994), ACM, 272-281.
7. Xu, J. & Croft, W. B. (1996). Query expansion using local and global document analysis. In *Proceedings of SIGIR* (Zurich, Switzerland, August, 1996), ACM, 4-11.