Understanding the Influence of Topic Familiarity on Search Behavior in Digital Libraries

Masoud Davari
GESIS - Leibniz Institute for the Social Sciences
masoud.davari@gesis.org

Ran Yu
GESIS - Leibniz Institute for the Social Sciences
ran.yu@gesis.org

Stefan Dietze
GESIS - Leibniz Institute for the Social Sciences
stefan.dietze@gesis.org

ABSTRACT

Search in digital libraries is among the most frequent activities for people to acquire knowledge and find resources for academic purposes. In order to better support users in fulfilling their information need with personalized search results, a better understanding of the relation between users' search topic familiarity and their search behavior is a prerequisite. Eye tracking data as a reliable source for studying users' interaction in digital library search sessions has not been sufficiently explored due to the difficulty of automatically and accurately processing eye tracking logs. In this paper, we study the influence of topic familiarity on users' search behavior to extend the current understanding of user behavior while searching digital libraries. To gain novel insights, we apply a recently emerged tool for parsing eye tracking logs on a dataset collected from a lab study which captures user interactions as well as topic familiarity in digital library.

KEYWORDS

task difficulty, topic familiarity, eye fixation, user interaction, digital library

ACM Reference Format:
Masoud Davari, Ran Yu, and Stefan Dietze. 2019. Understanding the Influence of Topic Familiarity on Search Behavior in Digital Libraries. In *Proceedings of SIGIR 2019 Workshop on ExplainAble Recommendation and Search (EARS'19).* ACM, Paris, France, 4 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

Search is among the most frequent activities for acquiring knowledge and information in daily life. Digital libraries have been widely used by students and researchers for a range of academic activities. Like general purpose search engines, the scale of available resources in digital libraries has grown tremendously in the past years, where better retrieval approaches such as ranking search results according to users' prior knowledge [13] and learning characteristics are necessary for fulfilling users' information need. Understanding a user's prior knowledge on the search topic is the first step for recommending resources. Hence, understanding the relation between user behavior and their topic familiarity is a prerequisite for optimizing and personalize Web search result.

Previous works have studied the influence of task difficulty and familiarity on search behaviors in Web search environmentS. Hu et al. [9] studied the influence of users' topic familiarity and search skills on the query reformulation process in health information seeking. Arguello [1] proposed to use logistic regression to predict task difficulty in a search environment. Data was collected through a crowdsourcing platform, and the author used search tasks created by Wu et al. [12], which conduct task difficulty assessments on multiple dimensions. White et al. [11] investigated the difference between the behavior of domain experts and non-experts and found that the distribution of features such as the number of queries and query length differed across the levels of expertise. Gwizdka and Spence [5] showed that a searcher's perception of task difficulty is subjective and depends on the domain knowledge and some other individual traits.

Eye tracking data has been used for extracting search behavior features in different scenarios. For instance, Bhattacharya et al. [2] investigated the relationship between users' search and eye gaze behaviours and their learning performance based on a lab study (n=30). Eickhoff et al. [4] studies the relation between eye fixation and query formulation. However, the previous works either manually label the eye fixation on a small scale, or analyse the eye fixation on a lower granularity such as paragraphs. With the emerging of *Reading Protocol* [7]– a tool for parsing and visualizing eye tracking data at word level, we are able to automatically parse large-scale eye tracking data at higher granularity with high accuracy.

Supported by the *Reading Protocol*, we parsed a dataset collected from a lab study (n=25) in which users were asked to search for information about topics of different familiarity on Sowiport[1], a digital library that contains more than nine million multi-lingual bibliographic records for social science research [8]. Next to traiditional search behavior logs, we extracted users' eye fixation at term level throughout the search process. With this data, we were able to study the influence of task difficulty on user behavior on different aspects.

In this paper we analyse the relationship between topic familiarity and features extracted from search and browsing behaviour as well as eye tracking data. Our findings extend the current understanding of the influence of the topic familiarity on search behavior in digital libraries and can potentially be used for the automated detection of user topic familiarity in real time and hence support the personalisation of result rankings in digital libraries.

EARS'19, July 25, 2019, Paris, France

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM.
https://doi.org/10.1145/nnnnnnn.nnnnnnn

[1] http://sowiport.gesis.org/

2 PROBLEM & DATASET

In this section, we introduce the research questions that motivate our analysis, the notations used in the remainder of the paper and the dataset used for the analysis.

2.1 Problem

This paper aims at extending the understanding of the relation between the user topic familiarity and their search behaviour. More specifically, we aim to answer the following research question: How are *search behavior*, *query reformulation* and *browsing behavior* of a user correlated with his/her topic familiarity? Results regarding each of the three aspects are shown in Section 3.2.1, 3.2.2 and 3.2.3 respectively.

Here we define the notations used in later sections. Let s be a search session of user u who is seeking for information. For each s, let Q be the set of queries $q_1, \cdots, q_n$ issued by u at the time points $t_1, \cdots, t_n$ throughout s, i.e. a query $q_p \in Q$ is executed at time point t_p. $K_p \subseteq K$ is the set of search terms $k_{p1}, \cdots, k_{pm}$ of the query q_p after removing stop words. Let T be the set of all seen terms $\tau_0, \cdots, \tau_l$ between t_0 and t_p. Then dur_{τ_i} is the "eye-fixation" duration of τ_i during the time period t_0 to t_p of session s. f_{τ_i} is the frequency of τ_i appeared in the viewed Web documents and search engine result pages (SERPs) between time t_0 and t_p.

An example taken from our experimental dataset is shown in Table 1. The user u=1 carried out a search session s with 2 queries, the query q_2 = "education inequality school start" issued at time t_2 = 2':02" consists of a set of query term K_2 = {"education", "inequality", "school", "start"}. From the time the user start the search session to the time t_2 = 2':02", the user have seen documents that contains a set of terms $T = \tau_0, \cdots, \tau_l$.

Table 1: Examples of query reformulation from our study

User(u)	Query term (q_p)	Time (t_p)
1	education inequality	2':02"
1	education inequality school start	3':03"
2	ethnische Bildungsungleichheit	3':39"
2	ethnische kompetenz	4':35"
2	education ethnic inequality	7':52"

2.2 Dataset

For this analysis we use data from a user study [10] which initially conducted with 30 participants but due to quality issue 5 were excluded. The remaining 25 participants are native German speakers with a Sociology background (with one exception of a psychologist). Among all participants, 1 was a postdoc, 10 held a master degree and 14 held a bachelor degree. 64% of the participants were female and 36% were male. All participants are within a age range of [23, 45] ($Mean$ = 28.6, SD = 4.12) and were recruited trough the IIRpanel [2].

The participants were provided with a mouse, a keyboard and a 22-inch monitor which was connected to a laptop. Participants' eye movements were captured by SMI iView RED 250 mobile eye trackers that were set to 60Hz sampling frequency. During the study, experimenters monitored the study from the observation room where the participants' eye-gaze behaviours were visualized using the Camtasia software[3]. All the participant were given a short introduction to Sowiport, and time to make themselves familiar with the portal before the study started.

During the study, participants were asked to perform the following tasks in two separate sessions (max. 30 minutes for each session):

(1) Use Sowiport to search for publications on a topic that they are well familiar with.
(2) Use Sowiport to search for publications on a topic they are not familiar with but are interested to learn more about it.

From the study, 50 sessions were collected in total. For each search session, the user interactions such as session duration, query terms and the visited Web documents were logged. Furthermore, the eye movements and the monitor screen activities were recorded by the software BeGaze. Participants looked at 2,344 web pages (SERPs and detailed views of Web documents which contain metadata about selected articles such as title, author and abstract) in total and result in 2.6 million rows of eye tracking data.

3 DATA ANALYSIS

3.1 Data Preprocessing

Standard approaches for analyzing eye gaze behaviour rely on manually parsing stored images and videos from eye tracking software or automatically parsing at a lower granularity (e.g. locate the gaze in a frame or on a paragraph). The *Reading Protocol* software [7] can determine the exact fixation duration and frequency of every fixated term. In this software, instead of images, original web pages are used as input and the output is a set of fixated terms including corresponding variables such as fixated terms, fixation frequency and fixation duration. In this study, (1) we use the variables of fixated terms obtained from *Reading Protocol* to analyze users' fixation behaviours on the term level. Figure 1 shows an example of user's fixation behaviour on an unfamiliar topic, (2) we use Sowiport log for analyzing session, query and browsing related behaviours, e.g. session duration, number of actions in each session and number of queries. In order to study the query formulation process, we identify new terms in each query and cross reference with the seen terms extracted by *Reading Protocol* to analyse the origin of the query terms. Stop words are removed from query terms and all the rest terms have been stemmed for the analysis.

3.2 Preliminary Results

3.2.1 Topic Familiarity and Search Behavior. We investigate the relation between the topic familiarity and the user search activities, in particular, the session and query related features. The extracted feature values and the corresponding standard deviations (SD) are shown in Table 2.

Based on the statistical result, the average session duration of familiar tasks is 6.1% longer than on unfamiliar tasks. In previous

[2] https://gesis.org/iirpanel

[3] https://www.techsmith.com/video-editor.html

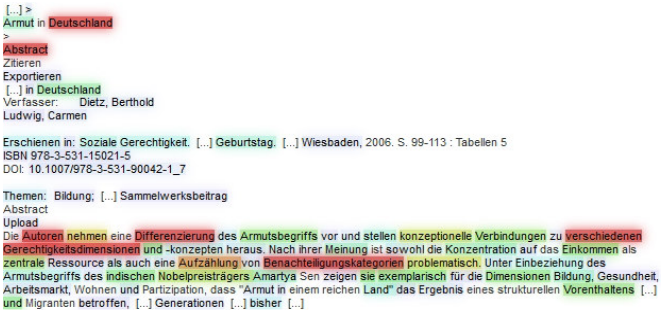

Figure 1: Eye behaviour of the first participant in an unfamiliar task obtained from *Reading Protocol*.

Table 2: Search behavior features of different task types.

Feature	Mean(familiar)			Mean (unfamiliar)		
session duration	1224.259	±	412.584	1153.434	±	410.341
# queries	7.37	±	3.59	5.48	±	2.41
session duration per query	298.78	±	286.473	345.108	±	267.557
# terms per query	2.21	±	0.724	2.044	±	0.649
# new terms per query	1.579	±	0.734	2.218	±	0.724
# actions	62.069	±	23.234	55.448	±	25.236

studies, the relation between the session duration and domain expertise have been found to vary on different domains [11]. In our case, Sowiport focuses on research topics in the social sciences domain, for which has not been studied to the best of our knowledge. Based on the result, in the case of searching in digital libraries for social sciences topics, users tend to perform longer sessions when they are more familiar with the search topic.

With respect to the number of queries in each session, we found that on average, users executed 7.37 (SD=3.59) queries on familiar topics and 5.48 (SD = 2.41) on unfamiliar topics. Combining the session duration and the number of queries, the result shows that on average, the browsing time in between queries on unfamiliar topics is 15.5% longer than on familiar topics.

The average length of the queries (#term per query) of the two different types of tasks are similar, both close to 2. It is intuitive that short queries are typically used in digital libraries, as the search is more resource oriented and keywords driven. With respect to the new terms in queries, the average number for unfamiliar tasks is 5.9% higher than familiar tasks. This is consistent with the observation reported in [9] that users with a higher topic familiarity tend to make less spelling errors and prefer to use specific terms for search (i.e. the important keywords are identified earlier in the session). When taken into consideration of all interactions (including query and all types of page visits) per session, familiar topics again shown to have more actions than unfamiliar topics.

A possible explanation for the findings above, i.e. longer search duration, more queries, more actions on familiar topics, would be that in our exploratory study setup the session duration is more dependent on users' engagement since we do not require the users to actually complete a particular task. The users dedicated more to the topics that they were familiar with than topics that they were encouraged to explore.

3.2.2 Topic Familiarity and Browsing Behavior. As found by previous works [5, 11], the prior knowledge of users on the search topic directly influences the selection of the Web resource and their browsing behaviour. In this section, we investigate the browsing behavior of users in digital library and compliment previous works with a higher granularity eye fixation analysis.

Table 3: The average number of document views before each term fixation.

Metrics	familiar	unfamiliar
All documents	1.748 ± 2.415	1.631 ± 2.106
SERP	0.736 ± 1.494	0.775 ± 1.381
non-SERPs	1.011 ± 1.579	0.865 ± 1.292

In our experimental dataset, we found that on average users fixated 23,020 and 16,774 terms for familiar and unfamiliar tasks respectively. Despite the average session duration of familiar tasks is 6.1% longer (Table 2), the 37.2% more fixated terms in familiar tasks suggest that the information gain is much larger for users when searching for topics that they are familiar with.

We further investigate where and when these terms were fixated after each query. Table 3 presents the average number of different types of page views before the fixation of each term since the last query is issued. This metric indicates the attention distribution of users in the browsing process after each query.

According to the results, prior to the fixation of each term, users viewed 1.748 documents on average for familiar tasks and 1.631 for unfamiliar tasks. The standard deviation metrics shows that the majority of the fixated terms are towards the beginning of after issuing a query, however, the fixation span long after each search. While out of the total view, users have less view on SERPs for familiar tasks. This again indicate that the users having higher tasks familiarity act faster when deciding the relevance of the resource to their information need and selecting resource from the SERPs. The same observation has been drawn in a cognitive study with eye tracking measures [3].

Table 4: Term fixation duration and term frequency of search sessions on familiar and unfamiliar topics.

Metrics	familiar	unfamiliar
$\frac{\lvert K \cap T \rvert}{\lvert K \rvert}$	0.607	0.512
$f_\tau, \tau \in K$	47.362 ± 55.483	40.228 ± 73.916
$f_\tau, \tau \notin K$	5.339 ± 1.369	5.2 ± 2.12
$dur_\tau, \tau \in K$	589 ± 507 ms	402 ± 292 ms
$dur_\tau, \tau \notin K$	247 ± 93 ms	242 ± 96 ms

3.2.3 Topic Familiarity and Query Formulation. Table 4 shows the percentage of fixed query terms among all fixated terms ($\frac{\lvert K \cap T \rvert}{\lvert K \rvert}$), average fixation frequency (f_τ) and average fixation duration (dur_τ)

of each terms seen by users. $\tau \in K$ represents the terms that are used in future queries.

In *familiar* tasks, 60.7% of the search terms have been fixated prior to the query term acquisition. In *unfamiliar* tasks, only 51.2% of query terms have been fixated in the session. The frequency of seen query terms is significantly higher than other non-query terms for both familiar and unfamiliar tasks. This provides more evidence for the results reported in [4, 6] showing that users discover keywords to query for throughout the browsing process. When comparing between different types of tasks, the sessions of familiar tasks have higher fixation frequency on query terms compared to the sessions of unfamiliar tasks. This observation can be explained by the finding in [9], that the domain experts focus more on specific key terms and use these terms more frequently in different queries. This also aligns with the finding in Section 3.2.1 that fewer new terms are used in the queries in familiar tasks.

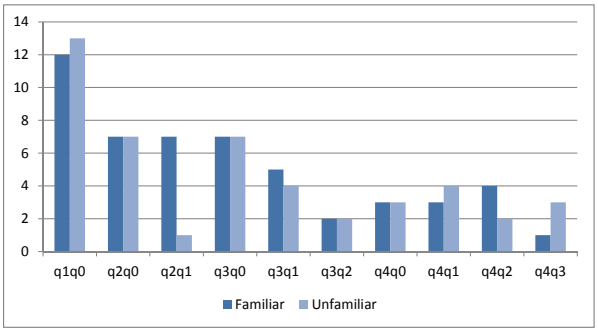

Figure 2: The first fixation time point of query terms for familiar and unfamiliar tasks.

The fixation duration shows a similar trend, where users fixated longer on the terms used in future queries for familiar topics. Similar to the average frequency of seen terms, there is no significant difference of the average term fixation duration of all terms between familiar and unfamiliar tasks. This again supports the assumption that it is easier for experts to find the sought-after information.

Furthermore, by investigating the time point at which a query term first was seen throughout a session we are able to draw more insights on users acquire query terms in tasks of different familiarity. Figure 2 present the number of the terms used in query q_i that were seen immediately after the query q_j ($j < i$) and before q_{j+1} (labeled on the x-axis as $qiqj$). Based on the result, up to the fourth query (corresponding to $q3qj, j = 0, 1, 2$ in figure) users search for familiar tasks tend to use terms that they saw more recently (e.g. $q2q1$ of familiar task is significantly larger than of unfamiliar task), while users searching for unfamiliar topics used more terms they discovered throughout the session. The reason might be that users having a higher topic familiarity may search for different aspects of the same topic to deepen their knowledge [9] rather than repeatedly searching for similar information. However, starting from the fifth query (corresponding to $q4qj$ on the x-axis in the figure), this observation no longer holds as users' familiarity changes throughout the search session and the focus may differ compared to the first a few queries.

In summary, results of both metrics support the finding that future query terms come from the seen terms during the session. For search in digital libraries, the more familiar a user is to a topic, the less effort it required for the user to find the right resource and to formulate the next query.

4 CONCLUSIONS

In this paper, we presented a preliminary analysis on the influence of domain familiarity on search behavior. We extend the current understanding of this topic by investigating the user search interactions in digital libraries, and study the eye fixation behaviour at a high granularity by making use of the eye tracking data parsing tool *Reading Protocol*. The observations are only partially consistent with the previous studies in Web search environments, which suggests that search behavior in digital libraries is different from general Web search. A more in-depth analysis of user search behaviors specifically in digital library environments is required in order to support the optimization and personalisation of search results. Our findings could in the future facilitate search result reranking and user interfaces optimization towards users topic familiarity and prior knowledge within the search task.

REFERENCES

[1] Jaime Arguello. 2014. Predicting Search Task Difficulty.. In *ECIR*, Vol. 14. 88–99.
[2] Nilavra Bhattacharya and Jacek Gwizdka. 2019. Measuring Learning During Search: Differences in Interactions, Eye-Gaze, and Semantic Similarity to Expert Knowledge. In *Proceedings of the 2019 Conference on Human Information Interaction and Retrieval*. ACM, 63–71.
[3] Julien Dampuré, Christine Ros, Jean-François Rouet, and Nicolas Vibert. 2012. How word familiarity facilitates visual search for verbal material. *Applied Cognitive Psychology* 26, 2 (2012), 271–288.
[4] Carsten Eickhoff, Sebastian Dungs, and Vu Tran. 2015. An eye-tracking study of query reformulation. In *Proceedings of the 38th international ACM SIGIR conference on research and development in information retrieval*. ACM, 13–22.
[5] Jacek Gwizdka and Ian Spence. 2006. What can searching behavior tell us about the difficulty of information tasks? A study of Web navigation. *Proceedings of the Association for Information Science and Technology* 43, 1 (2006), 1–22.
[6] Matthias Hagen, Martin Potthast, Michael Völske, Jakob Gomoll, and Benno Stein. 2016. How writers search: Analyzing the search and writing logs of non-fictional essays. In *Proceedings of the 2016 ACM on Conference on Human Information Interaction and Retrieval*. ACM, 193–202.
[7] Daniel Hienert, Dagmar Kern, Matthew Mitsui, Chirag Shah, and Nicholas J Belkin. 2019. Reading Protocol: Understanding what has been Read in Interactive Information Retrieval Tasks. In *Proceedings of the 2019 Conference on Human Information Interaction and Retrieval*. ACM, 73–81.
[8] Daniel Hienert, Frank Sawitzki, and Philipp Mayr. 2015. Digital library research in action–supporting information retrieval in sowiport. *D-Lib Magazine* 21, 3/4 (2015).
[9] Rong Hu, Kun Lu, and Soohyung Joo. 2013. Effects of topic familiarity and search skills on query reformulation behavior. In *Proceedings of the 76th ASIS&T Annual Meeting: Beyond the Cloud: Rethinking Information Boundaries*. American Society for Information Science, 66.
[10] Dagmar Kern, Daniel Hienert, Katrin Angerbauer, Tilman Dingler, and Pia Borlund. 2019. Lessons Learned from Users Reading Highlighted Abstracts in a Digital Library. In *Proceedings of the 2019 Conference on Human Information Interaction and Retrieval*. ACM, 271–275.
[11] Ryen W White, Susan T Dumais, and Jaime Teevan. 2009. Characterizing the influence of domain expertise on web search behavior. In *Proceedings of the second ACM international conference on web search and data mining*. ACM, 132–141.
[12] 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.
[13] Ran Yu, Ujwal Gadiraju, Peter Holtz, Markus Rokicki, Philipp Kemkes, and Stefan Dietze. 2018. Predicting User Knowledge Gain in Informational Search Sessions.. In *SIGIR*, Kevyn Collins-Thompson, Qiaozhu Mei, Brian D. Davison, Yiqun Liu, and Emine Yilmaz (Eds.). ACM, 75–84. http://dblp.uni-trier.de/db/conf/sigir/sigir2018.html#0001GHRKD18