Examining and Supporting Laypeople's Learning in Online Health Information Seeking

Yu Chi*
4th year Ph.D. Student
School of Computing and Information
University of Pittsburgh
Pittsburgh, PA, USA
yuc73@pitt.edu

ABSTRACT

It has long been understood that knowledge acquisition is an important component in the information seeking process [2, 18]. Further, empirical studies have demonstrated that learning is a common phenomenon in information seeking [8, 10, 20]. However, for users, especially laypeople, who must gain knowledge through their interactions with a search engine, the current general-purpose search engine does not sufficiently support learning through search. Health information seeking (HIS, hereafter) is a domain-specific search [14], where users who possess higher knowledge tend to have better strategies and performances in solving their search tasks [3, 21]. While learning clearly plays an important role in the HIS process, there has been little research in this area. Little is known about the factors that might enhance or impede such learning during online HIS. Therefore, this project aims at examining health consumers, especially laypeople's search as learning behaviors and performances. A mixed method design will be adopted, consisting of experimental-based studies and interviews. So far, we have conducted 24 user studies and semi-structured interviews, investigating the source selection behaviors in the HIS tasks with increasing levels of learning goals. The results of this phase of the study will be used to guide the following analysis and predict laypeople's knowledge levels in the HIS process and provide corresponding support.

CCS CONCEPTS

• **Information systems** → *Users and interactive retrieval*;

KEYWORDS

search as learning; health information seeking; source selection; system design

ACM Reference Format:
Yu Chi. 2019. Examining and Supporting Laypeople's Learning in Online Health Information Seeking. In *2019 Conference on Human Information Interaction and Retrieval (CHIIR '19), March 10–14, 2019, Glasgow, Scotland, Uk*. ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3295750.3298975

*This work is guided by my advisor: Dr. Daqing He, dah44@pitt.edu.

CHIIR '19, March 10–14, 2019, Glasgow, Scotland, Uk

ACM ISBN 978-1-4503-6025-8/19/03.
https://doi.org/10.1145/3295750.3298975

1 MOTIVATION

According to a national survey conducted by the Pew Research Center [5], 72% of Internet users in the USA have looked online for health information, many reporting that this information affected decisions they took about their health. However, not all online health consumers are equal.

In general, health information seeking (HIS) is a typical type of domain-specific search [14], in which high domain knowledge is required to conduct search successfully. Therefore, consumers who possess higher knowledge tend to have better strategies and higher success rates in solving HIS tasks [3, 21]. While, on the contrary, laypeople mostly rely on the general-purpose search engine as they are not familiar with domain-specific sources, but the lack of knowledge also restrains them from formulating the accurate queries[3, 21]. The current general-purpose search engine is not sufficient in supporting laypeople's health-related search.

The knowledge of a person is not fixed. Information seekers can gain domain-specific knowledge in their interactions with the online information sources. In information seeking behavior literature, learning has been primarily situated around the concept of knowledge acquisition, as stated in ASK (Anomalous State of Knowledge) model, in which Belkin [2] argued that information seeking is a process to resolve the anomaly between users' current states of knowledge and the problem they faced. Marchionini [18] claimed that beyond simple lookup search, people often engage in exploratory search tasks where learning and investigation could play essential roles. Besides, numerous empirical studies suggested that learning is a common phenomenon in people's search process [8, 20]. Learning can occur in a relatively long-duration search [20, 22]. It can take place in a single search session as well [6, 8, 12].

While an increasing population relies on online information seeking to solve their health-related questions, it is known that, in this kind of domain-specific information seeking, learning plays an important role as both a byproduct of the search process and as an influential factor determining the search performance [3, 21]. However, to the best of our knowledge, there is little existing work that combines learning and HIS. Therefore, our ***first motivation*** in this research project is to fill the gap of the literature to understand laypeople's behavior and performance from the perspective of learning achievement in HIS.

Additionally, in the assessment of performance, we are interested in how well the current HIS web environment can support laypeople in terms of learning. So we plan to examine the learning performance against the six cognitive levels in Bloom's taxonomy [1].

Bloom's Taxonomy[1] identified six levels of learning goals from lower to higher order of thinking skills. It is used in the search as learning studies for assessing the learning outcomes[8, 12, 23].

Our ***second motivation*** is to investigate the interplay between health consumers' behaviors and their learning outcomes. Previous work suggests that domain experts and novices behave differently in searching [21, 22], while on the other hand, there exist studies demonstrate one's knowledge level is predictable from her searching behaviors [17, 24]. So this motivated us to investigate the interplay between search behaviors and learning performances in the dynamic HIS interaction process. Is it possible to predict the laypeople's current learning performance based on the prior search behaviors? Besides, what following search behaviors will one adopt given the temporary learning performance? The goal is to alleviate the difficulties and provide the demanding facilities that would result in better search as learning experience in the design of the domain-specific search system.

In addition, our ***third motivation*** is inspired by the uniqueness of health information seeking. We are interested in examining whether searching for different health conditions, for example, acute vs. chronic or severe vs. mild conditions would influence the relationship between search behavior and learning performance.

In the most HIS tasks, health consumers', especially laypeople's domain knowledge is lack, so they are more likely to involve in search as learning process. A search system specifically supporting learning in the health-related domain is needed. Given the insufficient support for learning in the current search system, our final objective in this study is to provide design implications or improve ranking algorithm that enhances search as learning performance in the HIS process and evaluate its usability.

With more and more exploration in the search as learning research area, studying learning in the HIS is a promising research agenda. Overall, this research project aims to respond to an increasing demand for understanding the interplay between search as learning and health information seeking, and how the external factors such as health conditions would influence the behavior and the performance, thus providing demanding support.

2 RESEARCH QUESTIONS

Specifically, the research questions are defined as follows. Firstly, RQ1 and RQ2 are driven by the first motivation to understand the laypeople's behaviors and learning performance.

- RQ1: How do laypeople search to fulfill HIS information needs with different levels of learning goals?
- RQ2: To what extent, can web search support laypeople's search as learning for HIS information needs?

RQ3 is to examine the interplay between search behavior and learning performance, and it is broken down to two sub-questions.

- RQ3a: How does the former search behavior influence the learning performance at different stages of the HIS process? Does the influence vary when searching for different health conditions? If so, how?
- RQ3b: How does the temporal learning performance influence the following search behavior at different stages of the HIS process? Does the influence vary when searching for different health conditions? If so, how?

The answer to RQ1, RQ2, and RQ3 will help to design learning enhancing search systems. Therefore, the RQ4 is:

- RQ4: How can we support and enhance the search as learning in the system design? What is the usability of the support functions?

3 METHODOLOGY

We adopt a mixed method design of combining quantitative and qualitative approaches. Mixed methods research possesses the strengths that offset the weaknesses of both quantitative and qualitative research [9]. As shown in Figure 1, this research consists of three studies for the data collection: user study I and the semi-structure interview for collecting the data to answer RQ1, RQ2 and RQ3. Based on the results, user study II will be conducted to answer RQ4.

The major emphasis in the experimental-based user study is the behavioral data and the emphasis in the survey-based interview study is the subjective data. The data from these two studies will be compared with each other to identify agreements or divergences. The triangulation of the two separate studies helps to increase the validity as well as provide a better understanding of the research problem than either approach alone [13].

4 PROGRESS

To date, the data collection in the user study I and the semi-structured interview has been conducted. After initial data analysis, we have obtained following preliminary results.

4.1 Study 1: User Study I

We designed the controlled study around scenario-based health problems, and provided a live search system for the participants to search on. The study design was approved by the Human Research Protection Office (formerly IRB) of the University of Pittsburgh.

4.1.1 Health conditions and task design. We selected two health conditions with different severity levels: Severe Condition (SC) and Mild Condition (MC). SC was designed to be a more urgent and more complex health issue comparing to MC. This is because we want to explore whether health conditions with different severity levels would result in different search as learning behaviors and performances. As for the specific health issue, the former was multiple sclerosis and the latter was weight loss.

The design of the specific search task is guided by Table 1. The simulated tasks should be both close to real HIS information needs on the web and capable of reflecting learning with different levels of goals. Therefore, we map the classification of health search intentions established by Cartright et al. [4] through large-scale real search logs to three levels of Bloom's taxonomy [1]. Understand, analyze and evaluate represents three levels of six learning goals in Bloom's Taxonomy[1] from lower to higher order of thinking skills. Based on the mapping, we also propose specific search goals for each HIS intention. This framework was used to design the specific scenario-based search tasks in the user study I.

4.1.2 System and procedure. Upon arrival, the participants were asked to fill in a questionnaire about their background (e.g., age, education). Then they were introduced with our experimental search system, on which they could freely search, click and view webpages. The system is wrapped around Google search API, and it returns

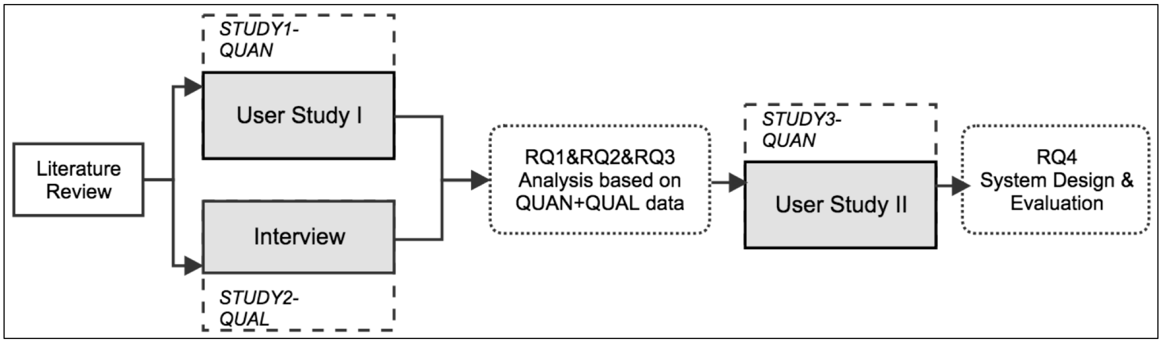

Figure 1: Data Collection Techniques and the Research Questions

Table 1: The Mapping Between Health Information Seeking Intentions and Depths of learning

Depths of Learning [1]	HIS Intentions [4]	Task Design Guidance
Understand	Evidence-based HIS	pursuit of details and relevance of signs and symptoms
Analyze	Hypothesis-based HIS	pursuit of content on one or more illnesses, and on the discrimination among different diseases under the uncertainty
Evaluate	Decision-based HIS	given specific situation of a condition, collect causes, treatments, tests and other information to make decision

Google search results with our behavioral logging functions for follow-up log analysis. During the introduction of our system, the participants were given a training task to get familiar with our system, and they were explicitly informed that their behaviors such as queries and clicks would be logged and analyzed afterward. Participants could highlight certain text in any webpage and click the "save to workspace" button to save it. They could save as many snippets as possible, which would be used for them to generate reports after the search. During the actual search, each participant were assigned to the two health conditions in different rotations to avoid sequential bias. The participants were given 21 minutes to complete the three subtasks with increasing levels of learning goal for each condition. The whole study took approximately 1.5 hours.

4.1.3 Participants and user study data. 24 college students (15 females and 9 males, 18-33 years old) were recruited as convenient samples from the University of Pittsburgh (17) and Carnegie Mellon University (7). 17 of them were undergraduates and the remaining were graduate. We applied two screening criteria to make sure all the participants were laypeople: 1) we excluded people from health or medicine related domains; and 2) we only recruited people who have conducted online health information seeking within the last 12 months [19]. Therefore, most of our participants have searched for health-related information very recently (in last week: n=15, 62.8%; last month: n=8, 33.3%; last six months: n=1, 4.2%). In total, we collected 48 complete search logs (144 search sessions) from 24 participants. Our data collection contains 5,298 clicked webpages, among which 965 (18.21%) were saved into the workspace.

4.1.4 Current findings: analysis of the source selection behavior. To start with RQ1 and RQ3, source selection during the whole HIS process has been examined. One preliminary study has been published in CHIIR'18 [7]. in which we examined what sources laypeople select (i.e., visit and adopt) to resolve their HIS needs, and how different health conditions affect the selection.

Source selection is found to vary in the information seeking process and would result in different learning outcomes [7, 16]. While, given the uniqueness of HIS, source selection becomes an even more important type of behavior. This is because health consumers may use the information found online to make critical health-related decisions. Besides, different types of online sources might serve different roles in the learning process, because understanding health-related information often requires users to master a certain amount of domain-specific knowledge, particularly when encountering the resources full with domain-specific terminologies. The selected online health information sources were classified on the website level as well as webpage level for the analysis. We plotted the source selection behavior in search tasks with different levels of learning goals, and found that, in both health conditions, the participants visited more sources in the tasks with higher levels of learning goals, especially in the evaluate task, but significant results were only found in the mild condition. We also analyzed the changes in the types of online sources. The results indicated the laypeople tend to select more health-specific webpage, such as WebMD.com, when they lack sufficient knowledge to solve their HIS task. Additionally, laypeople will select more diverse types of sources with the increase of the levels of the learning goals, i.e., in the evaluate tasks. Consequently, search engines are also employed more frequently in the higher learning level tasks.

However, the relationship between source selection and the learning performance remains unknown and will be further analyzed.

4.2 Study 2: Semi-structured interview

4.2.1 Interview Design. Though the user study collects some subjective data, such as self-assessment of their learning achievements, we conducted another survey-based study, where the participants could express opinions in their own words, and we could directly elicit their HIS experiences in the real practice. The goal of the interview was to learn about laypeople's own HIS process and the factors that may involve in the process. A semi-structured interview protocol guided the interviews, consisting of two sections. In the first section, the participants were asked to express their general views about online HIS, including the barriers they encountered.

After these general conversations, the interview processed to the section designed to elicit the participants' real HIS experiences based on the principles of Critical Incident Technique (CIT) [11]. If they were glad to share, they were prompted to start from their most impressive or most recent HIS experience.

To take full advantage of CIT, we designed a list of questions and prompts to help the participants to recall and unfold the incidents from memory. In general, the conversation was led by a list of questions: *"What motived your HIS? How did you search for it? Could you recall what kind of websites did you select, and Why? Did you encounter any difficulty? How did you feel before and after the search?"* and *"Did your HIS involve any decision-making activities and Why?"*

4.2.2 Current findings: characterizing the HIS process and discovering the barriers relates to learning. 24 interview audio clips were transcribed and the data analysis was conducted through conventional content analysis, an inductive process [15]. The Participants reported that knowledge plays a critical role in their HIS experience, as lack of knowledge is a major difficulty. Participants were asked to identify the major barriers they confronted when conducting online HIS. Among the 30 relevant comments, the most commonly reported barrier is the 1) "lack of knowledge" about the disease (18 out of 30), accounting for more than half of the comments. 7 comments were grouped into the criticism of the quality of the information from the Internet - 2) "hard to find credible source". The rest 5 comments reflected that the difficulties come from - 3) "intrinsic complexity of the health condition." Furthermore, we characterized the diverse online HIS process with four critical aspects: information needs, search starting-points, emotion changes, and decision-making. The study demonstrated the importance of studying learning in HIS process. Future analysis will investigate the learning goal in each incident shared by the participants, and the corresponding behaviors and search process.

5 FUTURE PLAN

The future research plan includes three main phases:

(i) So far, the initial data analysis only centered on the source selection behavior in response to HIS tasks with different levels of learning goals. For the next step, we plan to explore other types of search behaviors and assess the performance of the participants with the rich behavioral logs, as well as the questionnaire responses. We asked the participants to write a short summary in answer to the search tasks after each search session and they also self-assessed their knowledge levels. This data is available for the assessment of the participants' learning performance.

(ii) With the results and findings, we plan to upgrade the current search system with either ranking algorithm or design features in support of learning for HIS. For example, is it beneficial to recommend different types of learning sources to the consumers based on the prediction of her topic knowledge level? Also, is it worth to assess the severity of the health condition searched by the consumer and support the HIS accordingly?

(iii) After employing the upgraded system, We plan to design user study II and evaluate the usability of the system.

REFERENCES

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

[2] Nicholas J Belkin. 1980. Anomalous states of knowledge as a basis for information retrieval. *Canadian journal of information science* 5, 1 (1980), 133–143.

[3] Suresh K Bhavnani. 2002. Domain-specific search strategies for the effective retrieval of healthcare and shopping information. In *CHI'02 Extended Abstracts on Human Factors in Computing Systems*. ACM, 610–611.

[4] Marc-Allen Cartright, Ryen W White, and Eric Horvitz. 2011. Intentions and attention in exploratory health search. In *Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval*. ACM, 65–74.

[5] Pew Reseach Center. 2013. Health Fact Sheet. *Pew Res. Cent. Internet Sci. Tech* (2013).

[6] Yu Chi, Shuguang Han, Daqing He, and Rui Meng. 2016. Exploring knowledge learning in collaborative information seeking process. In *CEUR Workshop Proceedings*, Vol. 1647.

[7] Yu Chi, Daqing He, Shuguang Han, and Jie Jiang. 2018. What Sources to Rely on:: Laypeople's Source Selection in Online Health Information Seeking. In *Proceedings of the 2018 Conference on Human Information Interaction&Retrieval*. ACM, 233–236.

[8] Kevyn Collins-Thompson, Soo Young Rieh, Carl C Haynes, and Rohail Syed. 2016. Assessing learning outcomes in web search: A comparison of tasks and query strategies. In *Proceedings of the 2016 ACM on Conference on Human Information Interaction and Retrieval*. ACM, 163–172.

[9] John W Creswell and Vicki L Plano Clark. 2017. *Designing and conducting mixed methods research*. Sage publications.

[10] Carsten Eickhoff, Jaime Teevan, Ryen White, and Susan Dumais. 2014. Lessons from the journey: a query log analysis of within-session learning. In *Proceedings of the 7th ACM international conference on Web search and data mining*. ACM, 223–232.

[11] John C Flanagan. 1954. The critical incident technique. *Psychological bulletin* 51, 4 (1954), 327.

[12] 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.

[13] Lisa Guion. 2002. Triangulation: Establishing the validity of qualitative studies. *University of Florida-IFAS Extension* (2002).

[14] Allan Hanbury. 2012. Medical information retrieval: an instance of domain-specific search. In *Proceedings of the 35th international ACM SIGIR conference on Research and development in information retrieval*. ACM, 1191–1192.

[15] Hsiu-Fang Hsieh and Sarah E Shannon. 2005. Three approaches to qualitative content analysis. *Qualitative health research* 15, 9 (2005), 1277–1288.

[16] Chang Liu and Xiaoxuan Song. 2018. How do Information Source Selection Strategies Influence Users' Learning Outcomes'. In *Proceedings of the 2018 Conference on Human Information Interaction&Retrieval*. ACM, 257–260.

[17] Jingjing Liu, Chang Liu, and Nicholas J Belkin. 2016. Predicting information searchers' topic knowledge at different search stages. *Journal of the Association for Information Science and Technology* 67, 11 (2016), 2652–2666.

[18] Gary Marchionini. 2006. Exploratory search: from finding to understanding. *Commun. ACM* 49, 4 (2006), 41–46.

[19] Susan L Perez, Debora A Paterniti, Machelle Wilson, Robert A Bell, Man Shan Chan, Chloe C Villareal, Hien Huy Nguyen, and Richard L Kravitz. 2015. Characterizing the processes for navigating Internet health information using real-time observations: a mixed-methods approach. *Journal of medical Internet research* 17, 7 (2015).

[20] Pertti Vakkari, Mikko Pennanen, and Sami Serola. 2003. Changes of search terms and tactics while writing a research proposal: A longitudinal case study. *Information processing & management* 39, 3 (2003), 445–463.

[21] 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.

[22] Barbara M Wildemuth. 2004. The effects of domain knowledge on search tactic formulation. *Journal of the american society for information science and technology* 55, 3 (2004), 246–258.

[23] Mathew J Wilson and Max L Wilson. 2013. A comparison of techniques for measuring sensemaking and learning within participant-generated summaries. *Journal of the American Society for Information Science and Technology* 64, 2 (2013), 291–306.

[24] Ran Yu, Ujwal Gadiraju, Peter Holtz, Markus Rokicki, Philipp Kemkes, and Stefan Dietze. 2018. Predicting User Knowledge Gain in Informational Search Sessions. *arXiv preprint arXiv:1805.00823* (2018).