

Describing Health Querying Behavior

Anushia Inthiran
School of Information Technology
Monash University Malaysia
anushia.inthiran@monash.edu

Saadat M Alhashmi
School of Business
University of Sharjah
United Arab Emirates
salhashmi@sharjah.ac.ae

Pervaiz K Ahmed
School of Business
Monash University Malaysia
pervaiz.ahmed@monash.edu

ABSTRACT

Laypeople use the Internet to search for health information. Limited research studies describe the query and reformulation patterns of laypeople when searching across tasks of varying levels of difficulty. In this paper, a small scale study is conducted to describe the health querying behavior of laypeople. Preliminary results indicate different querying patterns are demonstrated when searching across tasks of varying levels of difficulty. As task difficulty increased reformulation strategies remained the same. Implications for design are suggested to allow laypeople experience high querying efficacy.

CCS Concepts

• Information systems → Information Retrieval • Users and Interactive Retrieval

Keywords

Difficulty, health search, laypeople, query patterns, reformulation

1. INTRODUCTION

Results of a survey conducted in 2011 indicate top health topics searched for online are on food, drug safety and pregnancy [1]. In 2014, survey results indicate health searchers shared information online, used online means to track a health condition and read and watched someone else's experience about a health condition [2]. Whilst laypeople have access to plethora of health information online, performing a health search remains challenging. Laypeople do not have the necessary skills and expertise to perform a health search [3,4]. Laypeople experience challenges in the querying stage [3,4] and results viewing stage [5,6]. As a result, laypeople lament that they experienced low search satisfaction [7,8] and low search efficacy [8,9]. Whilst querying behavior of laypeople are well documented [3,4,5,6], there are limited studies which focus on describing qualitative changes in querying and reformulation patterns when searching across tasks of varying levels of difficulty. Instead of focusing on the entire search process, this study sets out to focus on querying patterns. The querying process is selected as it represents the searchers initial understanding of the problem and shows the development of this understanding through querying and reformulation patterns [3].

It is argued that categorizing querying and reformulation patterns when searching across tasks of varying levels of difficulty will enable rich understanding of the search session. Thus, when

SAMPLE: 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.

Conference '10, Month 1–2, 2010, City, State, Country.

Copyright 2010 ACM 1-58113-000-0/00/0010 ...\$15.00.

DOI: <http://dx.doi.org/10.1145/12345.67890>

querying and reformulation patterns for a difficult task is detected, search assisting features could intervene and provide personalised assistance. In order to describe querying and reformulation patterns when searching across tasks of varying levels of difficulty, a small scale experiment is conducted. Simulated situations are used to invoke the search process. MedlinePlus is used as the search domain. Querying and reformulation patterns are categorized based on user perceived task difficulty. The aim of this study is to describe qualitative querying and reformulation patterns when searching across tasks of varying levels of difficulty. The following research question will be tackled: Do qualitative querying and reformulation patterns differ when laypeople search across tasks of varying levels of difficulty?

2. LITERATURE REVIEW

In this section, a review of literature pertaining to i) laypeople's querying behavior and ii) querying patterns when searching across tasks of varying levels of difficulty is provided.

2.1 Laypeople's Querying Behavior

Laypeople typically enter queries that are characterized as short and simple [10,11,12]. However, newer studies indicate the average query length is between 4.82 to 5.33 words [3,6,10]. One explanation for the change in the average query length could be attributed to the use of natural language type queries [6]. In addition, terms such as 'what,' 'how', 'can', 'is', and 'does' frequently appear in queries [3,6]. Longer queries are issued only once whilst shorter queries tend to be issued repeatedly [10]. Longer queries cover a broader range of returned results thus shifting the focus of the searcher to viewing and evaluating results rather than issuing queries. The usage of long queries and natural language queries led to high querying effort, increased task completion time and left laypeople dissatisfied and frustrated with returned results [3]. Partial words and run together words [13] are commonly used when issuing a health query [3,7]. Laypeople also made spelling mistakes when issuing queries [10,11]. Most laypeople are able to initiate the search process, however others struggle to issue appropriate query terms for their search goal [11]. A trial and error approach was used when reformulating queries [11]. Other reformulation strategies include arbitrarily appending query operators to query terms [12]. Laypeople frequently experienced search failure and low search efficacy due to challenges faced in the querying stage [3,6].

2.2 Querying Patterns based on Task Difficulty

Research studies indicate task difficulty influenced search behavior [14,15,16]. When searching on a difficult task, searchers tend to use more natural language queries, question type queries, query operators and demonstrate unsystematic query refinement in comparison to when searching on an easy task [15]. Users also issued more queries when searching on a difficult task in comparison to an easy task [15,17]. The process of issuing queries

within short intervals is also an indicator of high task difficulty [14]. However, task difficulty did not influence query length [18].

With reference to health searches, laypeople on average issued more queries and longer queries when searching on an easy task in comparison to a difficult task [19]. Boolean operators and medical terms were issued when searching on an easy task in comparison to when searching on a difficult task [19]. Query reformulation took place more often when searching on a difficult task [3]. Laypeople also re-executed queries often when searching on a difficult task [3]. Previous research studies do not provide information on qualitative querying and reformulation patterns when searching across tasks of varying levels of difficulty. This information would be useful to interactively provide personalized assistance.

3. EXPERIMENT METHODOLOGY

A user's search behavior captured during the search process is thought to be more representative of the user's true behavior [20]. Although a large number of previous studies have used user logs to study search behavior [21,22], it is argued that humans act and compensate in many ways while dealing with errors or failures during the search process [20]. Thus, it is necessary to conduct user based studies.

An exploratory survey was conducted on a convenience sample of 20 participants. Participants were laypeople who did not have any education and training in health and medical sciences discipline. Participants were recruited based on call for participation notices posted on university bulletin boards. Each participant was given two simulated situations to search for on MedlinePlus. A total of 40 search sessions was analyzed. MedlinePlus was used as the search domain as the information retrieval strategy used in this domain is text based matching. This allowed for the observation of 'true' interactive searching without participants being influenced by any background information retrieval strategy. A pilot test was conducted with four participants. Theoretical saturation was achieved with 40 search sessions. This means there was no more change in querying behavior amongst participants. All querying activities were logged using a key logging software called Morae. Queries were retrieved from the keylogging software and analyzed manually. There was no time limit for the search session. Participants ended the search session when they were able to locate desired information or wished to terminate the search. The experiment began with the pre-experiment interview, followed by the search process and ended with the post experiment interview.

3.1 Pre-Experiment Interview

The pre-experiment interview was used to obtain socio-demographic details, information on general search experience and health search experience.

3.2 Simulated Situations

Simulated situations are used in this study to invoke the search process as it provides a platform to study and compare search behaviour amongst participants using a standardised scenario [23]. Clinical scenarios are used as this type of health search is commonly performed by laypeople [24]. Scenarios were selected based on cases observed at a hospital. An independent physician verified the simulated situations for medical accuracy. Simulated situations are provided in Figures 1 and 2. Simulated situations are handed to participants one at a time. The order of the simulated situation was rotated for each participant.

Simulated work task scenario: Your colleague had undergone a health test and found out that his kidney is enlarged and there is a stricture. A procedure called URS&RPG was performed. After this procedure, he then experienced urine retention. He is in pain and has been told that surgery is required. You are concerned for your colleague and would like to use Medline Plus to provide him with some information.

Indicative Request: Find for an instance, information to provide to your colleague of his condition, why is he experiencing this condition? Try to help him locate possible alternatives to treat his condition. Or is surgery the only option?

Figure 1. Simulated Situation A.

Simulated work task scenario: Today morning after getting out of bed you noticed that you could not move your neck. You could not move it left or right. There is swelling on the left side of your neck. The swelling seems to be near lymphatic nodes. You want to find out what is wrong.

Indicative Request: Find for an instance, information to inform you of your condition, what can be done to relieve you of this condition and why are in pain.

Figure 2. Simulated Situation B.

3.3 Post-Experiment Interview

At the end of each search session, participants were to asked rate their perceived post-search task difficulty [25]. Participants rated their perception of task difficulty [26] using one of three options: easy, neutral (neither easy nor difficult) or difficult.

3.4 Data Analysis

Querying patterns was analyzed based on Rose and Levinson's [27] classification. Specifically, the informational directed (open and closed) and informational undirected classification method was used [27]. Information directed queries are queries relating to knowing something particular about a topic. Information directed closed queries are specific to a single question whilst open resembles open ended questions. Informational undirected queries are queries related to learning everything/anything about a topic [27]. Definitions for information directed queries and undirected queries are taken directly from Rose and Levinson [27]. To analyze query reformulation patterns, the semantic (meaningful words) analysis methods were used [3]. All queries issued in a search session were analyzed manually to describe querying and reformulation patterns based on user perceived task difficulty.

4. RESULTS

There were 10 male and 10 female participants. The average age is 33.7 years old (SD=9.6). Participants consisted of undergraduate students, postgraduate students and university staff members. The average general search experience was 10.3 years (SD=3.5). The average health search experience was 6.2 years (SD=3.5). A total of 5 participants found searching for simulated situation A easy, 5 found it neutral and 10 found it difficult. A total of 15 participants found searching for simulated situation B easy, 5 found it neutral and 5 found it difficult. Thus, a total of 20 search sessions were collected when searching on an easy task, 10 on a neutral task and 15 on a difficult task. Overall querying and reformulation patterns when searching across tasks of varying levels of difficulty are presented in the following sections. More than 70% of query sessions demonstrate these querying and reformulation patterns. Query terms in bold are identified as informational directed open queries, query terms italicized are identified as informational undirected queries.

4.1 Query Behavior Patterns

When searching on an easy task, participants demonstrate initiating the query with informational directed open queries and resorted to informational undirected queries. This indicates participants wanted to get answers to specific questions and then moved on to searching for broader aspects of a topic. Participants initiated the search session with specific queries and thereafter became more confident to explore broader and wider areas about the topic by issuing informational undirected queries.

- (i) [enlarged kidney], [URS], [URS kidney]
- (ii) [lymph nodes], [swollen lymph nodes], [stiff neck], [swelling neck]

When searching on a neutral task. Participants initiated the query with informational undirected queries and subsequently used informational directed open queries. This indicates participants started the search with broad and wide queries and then narrowed down query terms. This querying pattern suggests participants wanted broad information to be provided before focusing on specific aspects.

- (iii) [kidney treatment], [kidney enlargement], [kidney stricture treatment], [kidney stricture urine retention]
- (iv) [lymphatic nodes], [stiff neck swollen], [neck lymphatic nodes]

When searching on a difficult task participants began the search using informational directed open queries then moved to using informational undirected queries before ending the search session with informational directed open queries. This pattern suggests participants broadening and narrowing the search without much focus. Participants querying behavior indicate unsystematic movements (random refinement).

- (v) [urological surgical procedure], [enlarged kidney], [URS RPG stricture]
- (vi) [stiff neck medication], [lymphatic nodes], [swollen neck lymph nodes]

Based on querying patterns demonstrated when searching across tasks of varying levels of difficulty, a model based on a triangle is used to depict querying patterns. This model is provided in Figure 3 (Note: SS– start search, ES – end search). Figure 3a depicts querying behavior when searching on an easy task, Figures 3b and 3c depicts querying behavior when searching on a neutral and difficult task respectively. The activity of broadening the query is depicted as moving towards the base of the triangle. The activity of narrowing down the query is shown as moving towards the tip of the triangle.

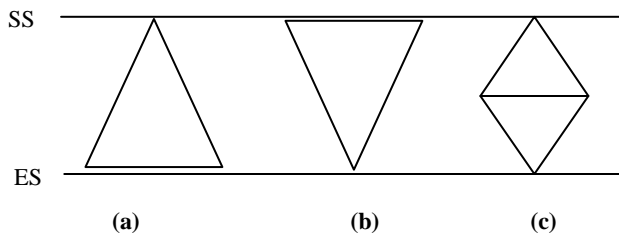


Figure 3. Models of qualitative querying patterns when searching across tasks of varying levels of difficulty.

4.2 Query Reformulation Patterns

Semantic changes (meaningful change in queries issued) indicate participants demonstrate two reformulation patterns when searching on an easy task. The first pattern indicates the reformulation strategy started with switching topic and ended with specialization. The second strategy begins with specialization and ends with parallel movement (reformulated query has partial overlap with previous query) [3]. When searching on a neutral task two reformulation strategies were demonstrated. The first strategy demonstrates parallel movement and ending with specialization. The second strategy demonstrates switching topic. When searching on a difficult task, the switching topic reformulation pattern is demonstrated. Reformulation patterns indicate participants demonstrate varying reformulation strategies when searching on an easier task. On the other hand, reformulations strategies did not change (switching topic only) as task difficulty increased. Table 1 provides information on reformulation strategies supplemented with examples.

Table 1. Reformulation patterns demonstrated when searching across tasks of varying levels of difficulty

Task Difficulty	Reformulation Pattern	Example
Easy	Switching Topic → Specialization	(i)
	Specialization → Parallel Movement	(ii)
Neutral	Parallel Movement → Specialization	(iii)
	Switching Topic	(iv)
Difficult	Switching Topic	(v), (vi)

4.3 Design Suggestions

Results indicate laypeople demonstrate different querying and reformulation patterns when searching across tasks of varying levels of difficulty. Based on results of this study, two recommendations are proposed to allow laypeople to experience high querying efficacy. These recommendations are: i) to develop an algorithm to detect and classify query and reformulation patterns and ii) to provide personalized query recommendations. As shown in Figure 3, participants demonstrate different querying patterns when searching across tasks of varying levels of difficulty. Differences in querying patterns enable categorization of querying behavior based on task difficulty. Thus, querying patterns can be classified into categories to infer and detect task difficulty. The development of an algorithm will enable interactive detection of task difficulty based on query patterns. As seen in Table 1, as tasks become more difficult, reformulation strategies remain the same. Laypeople demonstrate the switching topic pattern as task difficulty increased. In this case, there is also a need for health domains to detect this reformulation pattern with the use of an algorithm. Once the algorithm detects a query or reformulation pattern for a difficult task, health domains could provide relevant assisting features.

Health domains could assist participants who demonstrate the switching topic pattern by providing interactive query suggestions to help laypeople remain focused on the search goal. Interactive query suggestions would enable laypeople to issue relevant queries and also provide ideas to laypeople to construct efficient queries. Similarly, when the switching topic pattern goes on for too long before moving to specialization as shown in example (i) in Table 1, algorithms on health domains should be an the alert to intervene with interactive query suggestions. Thus, health

domains are able to personalize assisting features as and when required.

5. CONCLUSION

This study reports on qualitative query and reformulation patterns when searching on tasks across varying levels of difficulty. A model for querying patterns was developed and information on reformulation patterns was described. In future work the design suggestions will be built onto a health domain and a large scale study will be conducted to evaluate the design suggestions. It is noted that results cannot be generalized as data was analyzed based on user perceived perception of task difficulty.

6. ACKNOWLEDGMENTS

The authors thank participants of this survey.

7. REFERENCES

- [1] Fox, S (2014). Health Fact Sheet, Pew Research Centre, Available at <http://www.pewinternet.org/fact-sheets/health-fact-sheet/>
- [2] Fox, S. (2011). Health Topics, Pew Research Centre, Available at <http://www.pewinternet.org/2011/02/01/health-topics-2/>
- [3] Zhang, Y., Wang, P., Heaton, A. and Winkler, H. (2012). Health Information Searching Behavior In MedlinePlus and the Impact of Tasks, Proceedings of the 2nd ACM-SIGHT International Health Informatics Symposium,200-210
- [4] Zeng, Q.T., Kogan, A., Ash, N., Greenes, R.A. and Boxwala, A.A. (2002). Characteristics of consumer terminology for health information retrieval, Methods of Information in Medicine, Issue 41,289-298
- [5] Nadkarni, P. M. (2000). Information retrieval in medicine: Overview and applications, Journal of Postgraduate Medicine, Vol. 46, 116-122
- [6] Yang, C.C., Winston,F., Zarro,M.A. and Kassam-Adams,N. (2011). Proceedings of the iConference 2011, 267-272.
- [7] Liu, R.L.and Lu, Y.L. (2009). Context based online medical terminology navigation, Expert Systems with Applications, Available at doi:10.1016/j.eswa.2009.06.038
- [8] Eysenbach, G. and Kohler, C. (2002). How Do Consumers Search For And Appraise Health Information On The World Wide Web? Qualitative Study Using Focus Groups, Usability Tests, and In Depth Interviews, British Medical Journal, 2002, Vol. 24,573-577
- [9] Can, A.B. and Baykal, N. (2007). MedicoPort: A medical search engine for all, Computer Methods and Programs in Biomedicine, Vol. 86,73-86
- [10] Jadhav, A., Andrews, D., Fiksdal, A., Kumbamu, A., McCormick, J.B., Misitano, A., Nelsen, Ryu, E., Sheth, A., Wu Stephen and Pathak, J. (2014). Comparative Analysis on Online Health Queries Originating From Personal Computers and Smart Devices on a Consumer Health Information Portal, Journal of medical Internet Research, 16 (76),e160
- [11] Toms, E., and Latter, C.(2007). How consumers search for health information, Health Informatics Journal, 13(3),223-235
- [12] Silience, E., Briggs, P., Fishwick, L., and Harris P (2004). Trust and mistrust of online health sires, in the proceedings of CHI 2004, 663-670
- [13] McCary, A.T. and Tse T. (2003). Understanding Search Failures in Consumer Health Information Systems, AMIA Annu Symp Proc. 2003, 430-434
- [14] Liu, C., Liu, J. and Belkin, N. (2014).Predicting Task Difficulty at Different Search Stages, Proceedings CIKM, 569-578
- [15] Aula, A., Khan R.M., Guan, Z. (2010). How Does Search Behavior Change as Search Becomes more difficult?, Proceedings of the 28th Human Factors in Computing Conference CHI 2010, 35-44
- [16] Liu, J. Chang, S.K., and Cree, C. (2015). Exploring search task difficulty reasons in different task types and user knowledge groups, Inf. Process. Manage. 51(3), 273-285
- [17] Kim, J (2008). Perceived difficulty as a determinant of web search performance, Journal of Information Research, Vol. 13, No. 4, Paper 379 Available at: <http://informationr.net/ir/12-4/papers379.html>
- [18] Liu, J. and Belkin, N. (2011). Search task difficulty: the expected vs. the reflected. SIGIR 2011, 1183-1184
- [19] Inthiran, A., Alhashmi, S.M., and Ahmed, P.K. (2012). The affects of task difficulty on medical searches, Proceedings of the WWW, 12,533-534
- [20] Kelly, D. (2009). Methods for Evaluating Interactive Information Retrieval Systems with Users, Foundations and Trends in Information Retrieval, Vol. 3, Nos. 1-2 2009, Now Publishers
- [21] White, R.W. and Drucker, S.M. (2007) Investigating behavioral variability in web search, Proceedings of the 16th international conference on World Wide Web, 21-30
- [22] White, R.W. and Morris, M. (2007), Investigating the querying and browsing behavior of advanced search engine users, Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval, 255-262
- [23] Borlund P. (2003). The IIR evaluation model: A framework for evaluation of interactive information retrieval systems, *Information Research* 8,(3), paper no 152 Retrieved from: <http://information.net/ir/8-3/paper152.html>
- [24] Gonzalez-Gonzalez A.I., Dawes, M., Sanchez-Mateos, J., Riesgo-Fuertes, R., Escortell-Mayor, E., Sanz-Cuesta, T. and Hernandez-Fernandez, T. (2010). Information Needs and Information Behavior of Primary Care Physicians, *Annals of Family Medicine*, Vol.5, Issue 4, 345-352
- [25] Li, Y. and Belkin, N.J. (2008). A faceted approach to conceptualizing tasks in information seeking Information Processing and Management, Vol. 44,1822-1837
- [26] Wildemuth, B., Freund, L., and Toms, E. (2014). Untangling search task complexity and difficulty in the context of interactive information retrieval studies, Journal of Documentation, 70(6), 1118-1140
- [27] Rose, D.E., and Levinson, D. (2004). Understanding user goals in web search, WWW,13-19