The Impacts of Consumer's Health Topic Familiarity in Seeking Health Information Online

by Ira Puspitasari

The Impacts of Consumer's Health Topic Familiarity in Seeking Health Information Online

A Solution Based on Consumer's Perspective in Health Information Search

Ira Puspitasari
Information System Study Program
Faculty of Science and Technology, Universitas Airlangga
Surabaya, Indonesia
ira-p@fst.unair.ac.id

***Abstract*—Consumers or non-medical professionals are progressively going online to seek health information. Despite the increasing number of health information search online, acquiring the correct and relevant information based on consumer's understanding remains a problem. The information retrieved from the Internet may not fit consumer's understanding because the consumer's familiarity with health topic varies.**

To improve the accuracy of health information search results, this paper investigates the impact of consumer's familiarity on the search behaviour using language models approach. A user experiment was conducted with 60 participants searching on the topic tasks of dengue fever, diabetes mellitus, and gastroesophageal reflux disease. The participants also rated their familiarity with health task topics on the scale of 1 (not familiar at all) to 4 (familiar). This rating categorized the participants into four familiarity groups (F1, F2, F3, and F4). The data analysis involved the transcription of search task data into the sequence of search activities to identify unique search activity patterns between familiarity groups.

The results showed that the familiarity with health topics affected health information search behaviour. There were unique search patterns exhibited by groups of participants with different familiarity. In the query stage, participants with less familiarity issued more modified queries than the participants with higher familiarity. In the decision stage, familiar participants were likely to achieve higher search efficacy than less familiar participants. When locating the potential relevant search result, participants in the higher familiarity groups tended to be more successful than the participants in the lower familiarity groups. By analysing the search behaviour, health information search systems could predict the consumer's familiarity to present more relevant and understandable results.

***Keywords*—health information search; familiarity with health topic; search activity; sequence of search activities**

I. Introduction

The growing numbers of consumer health informatics researches have changed the way the consumers engage with health information. The use of Internet-derived health information is rapidly increasing. Among the contributing factors are the growing awareness of the need to equalize relationships between healthcare professionals and lay people [1], the increasing awareness to be more proactive and responsible for own and families health [2], and the pressure costs of healthcare system.

A number of consumer-centric health systems have been built to facilitate the rapid use of health information searches on the Internet, such as Health Information Query Assistant [3], Consumer Health Vocabularies initiatives [4], SimpleMed and MeshMed [5], and MedSearch [6]. Despite the increasing customer-centric effort, searching for health information remains difficult for most consumers. Most people are not familiar with medical terminology and understand some health terminologies differently with medical professionals. For example, a heart attack often mistakenly as the heart does not beat according to lay people, while to the doctor, a heart attack means there is damage to the heart muscle. The consumer's familiarity with health topic also varies. A person may be knowledgeable about tropical diseases, but not familiar with mental diseases, while another person may have the opposite familiarity. The terminology *dengue fever* (ICD-10-CM A90) is more familiar in Asian and Latin America countries because of its frequent occurrences in those regions [7]. This familiarity diversity may cause serious impacts since the information presented does not match the consumer's familiarity.

Most studies of health information seeking by consumers focus on building the bridge between consumers and medical professionals. To improve the quality of health information search, providing a *non-intrusive personalization* based on consumer's familiarity is necessary. The aim of this paper is to do the initial step in developing more accurate and less intrusive personalization, i.e., examining how familiarity with health topic influences health information searches. The identification of specific behaviours between unfamiliar and familiar consumers enables a non-intrusive familiarity prediction during a health information search session.

II. Research Design

A. Data Collection

In this study, the preliminary data collection was gathered from 60 participants. The criteria of the recruited participant are adult (age ≥ 18 years), non-medical professionals, the ability to write and read, basic skill in computer and Internet operation, and experiences in health information seeking on the Internet. The data was collected from transaction logs, history

978-1-5090-5756-6/17/$31.00 ©2017 IEEE
SERA 2017, June 7-9, 2017, London, UK

IEEE computer society

of Internet browsing, interviews, and screen and audio recording.

The 24 women and 36 men ranged in age between 21 to 45 participated in this study. The majority of the participants had a university-level education (52), mainly at the bachelor (30), or master (18) degree. The profile of the participants were students, faculties, researchers, administration staffs, and civil servants. Most of the students and faculties were from science and engineering, and information system department. All participants had been using the Internet for three or more years and had experience in health information seeking on the Internet. Most of the participants sought health information online as the need arise, mainly to look for information about a specific disease, certain medical treatment, and diet and nutrition.

TABLE I EXAMPLE OF HEALTH SEARCH TASK: TASK 1 (ENGLISH)

Health Topic: Dengue fever

Task 1
Your sister has been suffering a high fever and muscle pain for the last three days. You also noticed the appearance of rashes and purplish spots on her skin.

The doctor diagnosed her with *dengue fever* and advised hospitalization for a comprehensive treatment.

You want to know how a person can get a dengue fever, the likelihood of dengue fever can be life-threatening, and the treatment for dengue fever.

Pre-search familiarity self-assessment

Please rate your familiarity with the search task topic.
☐ 1 Never heard of it before.
☐ 2 Slightly familiar,
I am not familiar with this topic, but I have heard this topic somewhere.
☐ 3 Somewhat familiar,
I know a little about this topic.
☐ 4 Familiar,
I know this topic well.

Your Answer:
……………………………………………………………………
……………………………………………………………………
……………………………………………………………………
……………………………………………………………………

How would you rate the cognitive effort required to complete this task?
☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5
very large amount … very small amount

How would you rate the difficulty of this task?
☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5
very difficult … very easy

Please explain your general strategy to finish this task.
……………………………………………………………………
……………………………………………………………………
……………………………………………………………………

The instruments for data collection were written in English and Bahasa Indonesia. The instruments comprised of demographic profile survey, pre-search familiarity self-assessment, and health information search tasks. The search task consisted of three tasks, task 1 discussed dengue fever, one of the diseases that has grown dramatically around the world in recent decades [7]; task 2 was diabetes mellitus, one of the top 10 causes of death in 2015 [8]; and the topic for task 3 was gastroesophageal reflux disease (GERD). Table 1 presents an example of a complete health search task instrument in the data collection.

The participant completed a data collection in the following procedure:

1. Profile survey: the participant completed user profiles and health information search experiences.
2. Pre-search familiarity self-assessment: the participant was given time to read the search task scenarios carefully for 30 minutes. The participant rated the familiarity with the topic in each task based on his/her understanding on the scale of 1 (not familiar at all) to 4 (familiar).
3. Health information search task session: the participant was instructed to perform the search task one by one. The participant was free to use any search engine, website, or health information system at their own speed. All of the search session was screen recorded using CamStudio software [9]. After finished each search task, the participants provided comments about their perceptions on completing the task.

B. *Data Analysis*

The data collected from each participant consisted of profile data, a self-assessment of familiarity with health topic, and a video recording during the search session. Each participant produced three data instances, one for each task. The profile data captured the demographic characteristics of the participant. The self-assessment familiarity categorized participants into four familiarity groups, i.e., F1 (not familiar at all), F2 (slightly familiar), F3 (somewhat familiar), and F4 (familiar). The video data were used to examine how the participants completed the search task based on their familiarity with the health topic. Only the search session including the finding of the relevant answer that was transcribed further.

The video data of each search session was transcribed into a sequence of search activities. We developed the search activity coding scheme to transcribe the video based on the work in [10] and [11]. The coding scheme consists of three stages and fifteen activity types, as listed in Table II. An example portion of the transcription results is: “Q3-E1-E3-E3-E5-D1-E5-D3”. Based on the sequence, the participant started the search session by accessing a general search engine to construct a new query, examined the results retrieved, and selected two potential relevant results from health / medical websites. Subsequently, the participant evaluated the first selected articles and marked it as relevant. The last E5 and D3 showed that the participant evaluated the second article and marked it as not relevant.

TABLE II CODING SCHEME

Coding Scheme	Description
Stage 1: Query	
Q1	▪ Access and browse health/medical websites (e.g., consumer health informatics, health agency website, health specific search engine, medical dictionary), ▪ Submit a new query about a new/different health topic.
Q2	▪ Access and browse health/medical websites (e.g., consumer health informatics, health agency website, health specific search engine, medical dictionary), ▪ Modify previous query by making it more general or more specific.
Q3	▪ Access general search engine, ▪ Submit a new query about a new/different health topic.
Q4	▪ Access general search engine, ▪ Modify previous query by making it more general or more specific.
Stage 2: Evaluation	
E1	Evaluate the search result retrieved from the search engine.
E2	Discard the result retrieved without accessing any item.
E3	Select a retrieved item from a health/medical website.
E4	Select a retrieved item from a general website.
E5	Visit and evaluate the selected item to assess its relevance.
E6	Go to a specific item that has not been visited before by accessing link / media in the retrieved item selected.
E7	Access back to previously visited item.
Stage 3: Decision	
D1	Mark the selected result from a health/medical website as relevant and use it to answer the task.
D2	Mark the selected result from a general website as relevant and use it to answer the task.
D3	▪ Mark the selected result from a health/medical website as not relevant, or ▪ Discard the selected result from a health/medical website without visiting and evaluating it.
D4	▪ Mark the selected result from a general website as not relevant, or ▪ Discard the selected result from a general website without visiting and evaluating it.

C. *Identifying Search Activity Pattern*

To examine how the familiarity with health topic influences search behaviour, the next step in the data analysis was identifying common search activity using n-gram language models. The method to discover the pattern followed the method in [11] and [12]:

1. Transforming the sequence of search activities into n-gram language models. Each dataset was divided into 80% training data and 20% test data. The language models composed of 2-grams to 6-grams models. Given a user search session "Q1-E1-E3-E3-E5-D1-E5", the trigram represented three search activities consecutively, i.e., "Q1-E1-E3", "E1-E3-E3", "E3-E3-E5", and "E3-E5-D1". In the sequence "Q1-E1-E3", the likelihood of being in the state E3 depends on having been in state Q1 and E1 previously.
2. Evaluating the perplexity [13] of each language model. The perplexity $PP(p_M)$ of a language model p_M (next word w|history h) on a test set $T=\{w_1, \ldots, w_t\}$ was computed using the following formulation:

$$PP_T(p_M) = \frac{1}{\left(\prod_{i=1}^{t} p_M(w_i|w_1 \dots w_{i-1})\right)^{\frac{1}{t}}}$$

   The model with the lowest perplexity was selected as the number of search activities in a sequence that best represented the search pattern.
3. Applying the selected n-gram to the search activities data to identify common patterns in each familiarity group.

III. Results

A. *Familiarity with Health Topic*

The participants rated their familiarity with health topic for each task, as shown in Table III.

TABLE III FAMILIARITY SELF-ASSESSMENT RESULT

Task	Topic	Familiarity				Parti-cipants, n
		F1	F2	F3	F4	
1	Dengue fever	11	11	21	17	60
2	Diabetes mellitus	12	26	13	9	60
3	Gastroesophageal reflux disease (GERD)	27	15	11	7	60
	Total	50	52	45	33	**180**

The most recognizable health topic was dengue fever. All participants who rated their familiarity as F3 or F4 had developed an awareness of dengue fever since child age due to its prevalence in their surroundings. The most unrecognizable health topic in this study was GERD. The pre-search familiarity results in Table III also demonstrated that each participant was familiar with different health topics. A participant in this study was familiar with dengue fever and diabetes mellitus, but he/she was unknown to GERD. While another participant had experienced diabetes mellitus and GERD, but he/she had never of about dengue fever before.

B. *Search Activity*

All participants managed to find the correct answer to the questions in health search task. Thus, all search sessions were transcribed using the coding scheme in Table II. The transcription produced 180 sequences of search activities (1 sequence of search activities for each search session) and 4605 search activities. The longest and the shortest sequence in a search session contained 212 and 5 activities respectively. On average, a health information search session contained 25.1 search activities (SD 20.76). The composition of each search activity in each familiarity group is shown in Table IV.

The most frequent search activity performed by all participants was visiting and evaluating the selected item from the search result to assess its relevance (E5). On average, each familiarity group produced 284.25 (24.69%) E1 activity. The second and the third most frequent search activities in group F1 and F2 were evaluating the result retrieved from the search engine (E1) and selecting an article from a health website (E3). On the contrary, the second and the third most frequent search activities in group F3 and F4 were selecting an article from a health website (E3) and evaluating the result retrieved from the search engine (E1). In terms of assessing the selected item from a health website, group F1 encountered irrelevant article (D3) more frequent than relevant article (D1). In contrast, the other three groups discovered relevant article more often than irrelevant article. In the query stage, participants in group F1 and F2 modified the query more frequent than those in group F3 and F4.

TABLE IV PROPORTION OF SEARCH ACTIVITY IN EACH FAMILIARITY GROUP

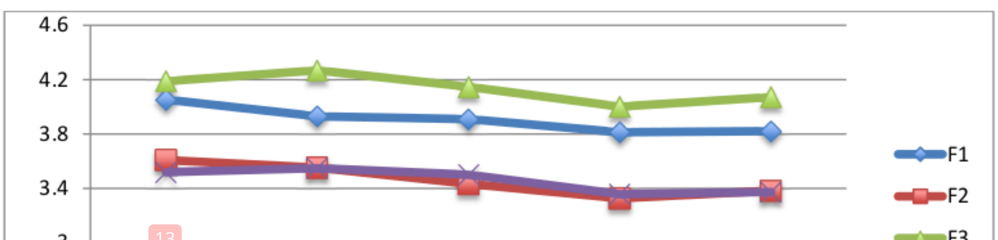
F1		F2		F3		F4	
Type	**%**	**Type**	**%**	**Type**	**%**	**Type**	**%**
E5	25.12	E5	24.53	E5	24.45	E5	23.62
E1	13.50	E1	15.67	E3	18.28	E3	19.87
E3	12.83	E3	15.26	E1	15.02	E1	15.67
D3	9.84	D1	9.68	D1	10.94	D1	13.25
Q4	7.57	Q4	7.14	D3	7.22	D3	7.28
E6	5.93	D3	5.74	Q3	5.24	Q3	5.30
D1	5.74	Q3	5.41	Q4	5.24	Q4	3.31
E4	5.35	E4	4.92	E4	3.26	E4	2.65
D4	3.76	D4	3.12	E6	3.26	Q1	2.43
E7	3.57	E6	3.12	D4	2.21	E6	1.99
Q3	3.18	E7	2.30	D2	1.16	D4	1.77
D2	1.74	E2	1.72	E2	1.16	D2	0.88
E2	1.74	D2	1.39	E7	1.16	Q2	0.88
Q2	0.10	Q1	0.00	Q1	1.05	E7	0.66
Q1	0.05	Q2	0.00	Q2	0.35	E2	0.44
Σ	100.00	Σ	100.00	Σ	100.00	Σ	100.00

C. Health Information Search Pattern

A search activity pattern captures the behavioural tendency of health information search in each familiarity group. The first step in identifying the pattern in this study was determining the pattern size. We used n-gram language model and perplexity evaluation to determine the number of search activities in a sequence that best represented the behavioural pattern.

Figure 1 shows the perplexity evaluations of all language models. According to the result, 5-gram language model had the lowest perplexity in all datasets, thus we used 5-gram sequences to identify the prevalent pattern in each familiarity group. There were 978, 548, 428, and 521 5-gram sequences types in group F1, F2, F3, and F4 respectively. In the analysis, only the 20 most frequent 5-gram sequences were examined further because the numbers of 5-gram sequences above top 20 were too small to be considered as common occurrences.

To identify the common pattern in each familiarity group, we categorized 5-gram sequences data into six clusters. The categorization is as follows: 1) Cluster 1 comprised issuing a new or modified query, examining the retrieved results, and selecting an article from a health website or general website, and evaluating its relevance; 2) Cluster 2 comprised examining the retrieved results, selecting multiple articles from the search results, and evaluating the articles' relevance one by one; 3) Cluster 3 included finding a relevant article in the results retrieved from the first query; 4) Cluster 4 included finding a relevant article in the results retrieved from the modified query; 5) Cluster 5 consisted of continuing the search process after finding a relevant article by examining the previous search results; 6) Cluster 6 consisted of continuing the search process after finding a relevant article by issuing a new or a modified query. Table V shows the frequent search patterns in each familiarity group based on the cluster categorization.



	2-gram	3-gram	4-gram	5-gram	6-gram
F1	4.05253	3.93121	3.90943	3.81414	3.82099
F2	3.60957	3.55311	3.43361	3.32802	3.38267
F3	4.18965	4.26801	4.14734	4.00154	4.07288
F4	3.52004	3.55017	3.50314	3.35959	3.37452

FIGURE 1 PERPLEXITY VALUES FOR EACH FAMILIARITY GROUP TEST DATA

TABLE V THE FREQUENT SEARCH PATTERNS IN EACH FAMILIARITY GROUP BASED ON THE CLUSTER CATEGORIZATION

F1		F2		F3		F4	
5-gram sequences[1]	**Σ (%)**[2]	**5-gram sequences**[1]	**Σ (%)**[2]	**5-gram sequences**[1]	**Σ (%)**[2]	**5-gram sequences**[1]	**Σ (%)**[2]
Cluster 1: Issuing a new or modified query, examining the retrieved results, and selecting an article from a health website or general website, and evaluating its relevance							
1. Q4 E1 E3 E5 D1 2. Q4 E1 E3 E5 D3 3. E2 Q4 E1 E3 E5	18.11%	1. Q4 E1 E3 E5 D1 2. D1 Q4 E1 E3 E5 3. Q3 E1 E3 E5 D1 4. D1 Q3 E1 E3 E5 5. Q4 E1 E3 E5 D3 6. Q4 E1 E4 E5 D4	30.63%	1. Q4 E1 E3 E5 D1 2. Q3 E1 E3 E5 D1 3. D1 Q3 E1 E3 E5 4. D1 Q4 E1 E3 E5	20.67%	1. D1 Q3 E1 E3 E5 2. Q3 E1 E3 E5 D1 3. D1 Q1 E1 E3 E5 4. Q1 E1 E3 E5 D1 5. Q4 E1 E3 E5 D1	27.07%
Cluster 2: Examining the retrieved result, selecting multiple articles from the search results, and evaluating the articles' relevance one by one							
-	-	1. E1 E3 E3 E5 E5 2. E1 E3 E3 E3 E5	7.01%	1. E1 E3 E3 E3 E5 2. E3 E3 E5 D3 E5 3. D1 Q3 E1 E3 E3	12.85%	1. E3 E3 E5 D3 E5 2. E1 E3 E3 E5 D3 3. E1 E3 E3 E5 E5	9.77%
Cluster 3: Finding a relevant article in the results retrieved from the first query							
-	-	Q3 E1 E3 E5 D1	4.07%	Q3 E1 E3 E5 D1	5.59%	1. Q3 E1 E3 E5 D1 2. Q1 E1 E3 E5 D1	11.28%
Cluster 4: Finding a relevant article in the results retrieved from the modified query							
Q4 E1 E3 E5 D1	8.65%	Q4 E1 E3 E5 D1	9.59%	Q4 E1 E3 E5 D1	6.15%	Q4 E1 E3 E5 D1	3.76%
Cluster 5: Continuing the search process after finding a relevant article by re-examining the previous search results							
-	-	E3 E3 E5 D1 E5	3.69%	1. E1 E3 E5 D1 E1 2. E3 E5 D1 E1 E3 3. E5 D1 E1 E3 E5	17.32%	1. E1 E3 E5 D1 E1 2. E3 E5 D1 E1 E3 3. E5 D1 E1 E3 E5 4. D1 E1 E3 E5 D1 5. D1 E1 E3 E3 E5	28.57%
Cluster 6: Continuing the search process after finding a relevant article by issuing a new or modified query							
1. E1 E3 E5 D1 Q4 2. E3 E5 D1 Q4 E1	8.65%	1. E5 D1 Q4 E1 E3 2. E5 D1 Q3 E1 E3 3. E1 E3 E5 D1 Q3 4. E1 E3 E5 D1 Q4 5. E3 E5 D1 Q3 E1 6. D1 Q4 E1 E3 E5 7. D1 Q3 E1 E3 E5	47.23%	1. E5 D1 Q3 E1 E3 2. E3 E5 D1 Q3 E1 3. E5 D1 Q4 E1 E3 4. D1 Q3 E1 E3 E3	20.11%	1. E5 D1 Q3 E1 E3 2. D1 Q3 E1 E3 E5 3. D1 Q1 E1 E3 E5 4. E3 E5 D1 Q3 E1 5. E5 D1 Q1 E1 E3 6. E1 E3 E5 D1 Q1 7. E1 E3 E5 D1 Q3 8. E3 E5 D1 Q1 E1 9. E1 E3 E5 D1 Q4	46.62%

[1] a 5-gram sequences could be in more than one cluster.
[2] Σ (%) is the ratio (in percent) of all related 5-gram sequences in the cluster to the total number of top 20 frequent 5-gram sequences in each familiarity group.

IV. Discussion

A. Search Activity in Each Familiarity Group

All participants allocated significant efforts in the evaluation stage by visiting and evaluating the selected item to assess its relevance, evaluating the search result retrieved from the search engine, and selecting a retrieved item from a health/medical website. Most participants also used general search engines (such as Google and Bing) to start a health information search session. These findings show that most non-medical professionals are not accustomed to health information search.

In the query stage, participants in the less familiar groups issued more modified queries than the participants in the higher familiarity groups. A participant in group L1 submitted six queries to answer the question in Task 3, i.e., "GERD", "what is GERD", "heartburn", "burning sensation in your chest", "GERD burning sensation", and "how to treat GERD". Another participant in Group F2 issued five queries in Task 2, i.e., "what is diabetes mellitus", "diabetes mellitus symptoms", "sugar diet", "diabetes medicine", and "differences between type 1 and type 2 diabetes". The query submitted by participants in group F1 and F2 contained more keywords from the task description than the query submitted in group F3 and F4. These indicate that less familiar participants encountered more difficulties in the query formulation. In terms of the selection of search engines, most participants in all groups opted general search engine to start the search session. Some participants in group F4 directly accessed consumer health websites to seek health information. Knowledgeable users targeted more appropriate resources to increase the search efficacy.

In the decision stage, participants in the higher familiarity groups were likely to achieve higher search efficacy than less familiar participants. Search efficacy measured the number of relevant assessment (D1 and D2) against the number of items visited (E3, E4, E6, and E7) in a search session. Group F4 accomplished the highest efficacy of 56.14% compared with 27.00%, 43.27%, and 46.64% for group F1, F2, and F3 respectively. Based on the results in Table III, participants in group F1 found more irrelevant articles than the relevant articles. In the other three groups, the proportion of D1 (relevant item from a health/medical website) was higher in the more familiar groups.

B. *The Impact of Familiarity with Health Topic on Search Pattern*

The participants in different familiarity groups exhibited specific behaviours in health information search. In cluster 1, all groups issued new or modified queries, examined the retrieved results, accessed items from health website, and evaluated the item's relevance. A frequent 5-gram sequence in group F1 was modifying the query after discarding the previously retrieved result without accessing any article. In term of the query type, all 5-gram sequences in group F1 composed only modified query submissions to a general search engine. Cluster 1 in group F2 and F3 included new and modified query submission to a general search engine. On the other hand, cluster 1 in group F4 consisted of new and modified query submissions to a general search engine and new query submissions to a consumer health informatics. The search engine selection in group F4 demonstrated that knowledgeable users targeted more appropriate resources to increase the search efficacy. This finding is in line with the previous study that reported more familiar participants achieved better search efficiency than the less familiar participants [14].

A search strategy of finding the potential relevant items is selecting multiple items from the results retrieved and evaluating the items one by one. This strategy was performed by participants in group F2, F3, and F4. The participants in groups F1, who had never heard of the health topic before, tended to select an item and evaluate the item immediately. These unfamiliar searchers need to construct their understanding with health topic definition first before they can select another potential relevant article.

When locating the potential relevant search result, participants in the higher familiarity groups tend to be more successful than the participants in the lower familiarity groups. Previous work in [11] also reported the similar finding. Cluster 3 and cluster 4 demonstrate this tendency. Participants in group F4 had the highest proportion of finding the relevant article in the results retrieved from the first query than the other groups. Group F1 discovered the relevant article in the results retrieved only from the modified query. While group F2 and F3 located relevant articles in the retrieved results from the modified query more frequent than from the first query.

Most participants in all groups continued the search session after they discovered a relevant article. Participants in group F1 continued the search process by issuing another query, while participants in other groups proceeded the search session by reexamining the last retrieved results and issuing another query. They continued the search process to verify the correctness of the information they discovered and to resolve their incertitude about an unfamiliar health topic. These activities demonstrate that participants are likely to be more cautious in health information search.

V. Conclusion

This study investigates an emerging important issue in health information search, i.e., individual health topic familiarity. The consumer's behaviour in health information search can be examined using language models approach. The results of this study affirm the impacts of familiarity with health topic on search behaviour in health information searches. Each familiarity group in this study exhibits unique search patterns.

Delivering health information that matches the user's familiarity is crucial since misunderstanding in health information may lead to fatal consequences. By analysing the consumer's search behaviour, a heath information search system could predict the consumer's familiarity to present more relevant and understandable results and to improve the overall health information search experience

References

[1] G. Eysenbach, "Recent advances: Consumer health informatics", *BMJ*, vol. 320, no. 7251, pp. 1713-1716, 2000.

[2] P. F. Brennan and C. Safran, "Empowered Consumers," *Health Informatics Consumer Health Informatics*, pp. 8–21..

[3] Q. Zeng, J. Crowell, R. Plovnick, E. Kim, L. Ngo and E. Dibble, "Assisting Consumer Health Information Retrieval with Query Recommendations", *Journal of the American Medical Informatics Association*, vol. 13, no. 1, pp. 80-90, 2006.

[4] Q. Zeng and T. Tse, "Exploring and Developing Consumer Health Vocabularies", *Journal of the American Medical Informatics Association*, vol. 13, no. 1, pp. 24-29, 2006.

[5] X. Mu, K. Lu and H. Ryu, "Explicitly integrating MeSH thesaurus help into health information retrieval systems: An empirical user study", *Information Processing & Management*, vol. 50, no. 1, pp. 24-40, 2014.

[6] G. Luo, C. Tang, H. Yang and X. Wei, "MedSearch: A Specialized Search Engine for Medical Information Retrieval", in *17th ACM conference on Information and knowledge management*, California, 2008, pp. 143-152.

[7] "Dengue and severe dengue", *World Health Organization*, 2017. [Online]. Available:

http://www.who.int/mediacentre/factsheets/fs117/en/. [Accessed: 22-Jun- 2016]

[8] "The top 10 causes of death", *World Health Organization*, 2017. [Online]. Available: http://www.who.int/mediacentre/factsheets/fs310/en/. [Accessed: 15-Jun- 2016].

[9] "Camstudio Open Source Free Streaming Video Software". *Camstudio*, 2013. Available: http://camstudio.org/. [Accessed: 12-Mar-2016].

[10] I. Xie and S. Joo, "Transitions in search tactics during the Web-based search process", *Journal of the American Society for Information Science and Technology*, vol. 61, no. 11, pp. 2188-2205, 2010.

[11] I. Puspitasari, K. Moriyama, K. Fukui and M. Numao, "Effects of Individual Health Topic Familiarity on Activity Patterns During Health Information Searches", *JMIR Medical Informatics*, vol. 3, no. 1, p. e16, 2015.

[12] H. Chen and M. Cooper, "Stochastic modeling of usage patterns in a web-based information system", *Journal of the American Society for Information Science and Technology*, vol. 53, no. 7, pp. 536-548, 2002.

[13] S.F. Chen, D. Beeferman, and R. Rosenfield, *"Evaluation metrics for language models*", in DARPA Broadcast News Transcription and Understanding Workshop, pp.275-280, 2008.

[14] I. Puspitasari, K. Fukui, K. Moriyama and M. Numao, "Predicting Consumer Familiarity with Health Topics by Query Formulation and Search Result Interaction", in *Pacific Rim International Conference on Artificial Intelligence*, Gold Coast, Australia, 2014, pp. 1016-1022.

The Impacts of Consumer's Health Topic Familiarity in Seeking Health Information Online

ORIGINALITY REPORT

15%	10%	7%	1%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	www.scilit.net Internet Source	8%
2	Ira Puspitasari, Ken-ichi Fukui, Koichi Moriyama, Masayuki Numao. "Chapter 88 Predicting Consumer Familiarity with Health Topics by Query Formulation and Search Result Interaction", Springer Science and Business Media LLC, 2014 Publication	2%
3	Stillman, M.. "Grobner bases, monomial group actions, and the Cox rings of Del Pezzo surfaces", Journal of Algebra, 20071015 Publication	1%
4	linknovate.com Internet Source	<1%
5	B. Ruščić, G. L. Goodman, J. Berkowitz. "Photoelectron spectra of the lanthanide trihalides and their interpretation", The Journal of Chemical Physics, 1983 Publication	<1%

6	Ira Puspitasari. "Stakeholder's expected value of Enterprise Architecture: An Enterprise Architecture solution based on stakeholder perspective", 2016 IEEE 14th International Conference on Software Engineering Research, Management and Applications (SERA), 2016 Publication	<1%
7	www.science.gov Internet Source	<1%
8	Submitted to Visvesvaraya Technological University Student Paper	<1%
9	Kun Lu, Soohyung Joo, Taehun Lee, Rong Hu. "Factors that influence query reformulations and search performance in health information retrieval: A multilevel modeling approach", Journal of the Association for Information Science and Technology, 2017 Publication	<1%
10	beachmat.berkeley.edu Internet Source	<1%
11	"Detection of Acted Happy Emotion using Facial Patches", International Journal of Recent Technology and Engineering, 2020 Publication	<1%

12	Parth Shah, Vishvajit Bakarola, Supriya Pati. "Chapter 62 Neural Machine Translation System for Indic Languages Using Deep Neural Architecture", Springer Science and Business Media LLC, 2018 Publication	<1%
13	mafiadoc.com Internet Source	<1%
14	rucore.libraries.rutgers.edu Internet Source	<1%
15	"Proceedings of the 1st International Conference on Electronics, Biomedical Engineering, and Health Informatics", Springer Science and Business Media LLC, 2021 Publication	<1%
16	Adan Ortiz - Cordova, Bernard J. Jansen. "Classifying web search queries to identify high revenue generating customers", Journal of the American Society for Information Science and Technology, 2012 Publication	<1%
17	Jin Zhang, Dietmar Wolfram, Peiling Wang, Yi Hong, Rick Gillis. "Visualization of health-subject analysis based on query term co-occurrences", Journal of the American Society for Information Science and Technology, 2008 Publication	<1%

18	Remm, E.. "Affine structures on abelian Lie groups", Linear Algebra and Its Applications, 20030201 Publication	<1%
19	theory.stanford.edu Internet Source	<1%
20	www.justice.gov Internet Source	<1%
21	Xu-Shen Zhou. "Can Fuzzy Logic Make Technical Analysis 20/20?", Financial Analysts Journal, 07/2004 Publication	<1%
22	Carlos Eduardo Durange de C. Infante, Fabricio Molica de Mendonça, Paula Michele Purcidonio, Rogerio Valle. "Triple bottom line analysis of oil and gas industry with multicriteria decision making", Journal of Cleaner Production, 2013 Publication	<1%
23	PUSPITASARI, IRA, KEN-ICHI FUKUI, KOICHI MORIYAMA, and MASAYUKI NUMAO. "HEALTH INFORMATION SEARCH PERSONALIZATION WITH SEMANTIC NETWORK USER MODEL", Theory and Practice of Computation, 2014. Publication	<1%

Exclude quotes	Off	Exclude matches	Off
Exclude bibliography	On		

The Impacts of Consumer's Health Topic Familiarity in Seeking Health Information Online

GRADEMARK REPORT

FINAL GRADE

/0

GENERAL COMMENTS

Instructor

PAGE 1

PAGE 2

PAGE 3

PAGE 4

PAGE 5

PAGE 6

PAGE 7