Search Log Analysis Method to Uncover User Search Behaviour on Web Searching Environment

[1]Fadhilah Mat Yamin, [2]T. Ramayah, [3]Wan Hussain Wan Ishak,[4] Siti Norezam Othman
[1, 3,4] , Universiti Utara Malaysia, Sintok, Kedah, Malaysia
[2]Universiti Sains Malaysia, 11800 Penang, Malaysia
fmy@uum.edu.my

Abstract: User search behaviour was conceptualized as a strategy undertaken by the user in searching for information. Typically, searching activity on the web involved several steps; query formulation and re formulation, browsing the search results, and search results evaluation. The scope of this study has limited itself to query formulation that reflects the user search behaviour. The proposed method has been shown to successfully identify and classify user behaviour into two components namely; breadth search query and depth search query. The queries were initially recorded into search log through search interface. The search interface is one of the innovative tools that interface the Google search engine. Through this interface, user can enter the query and obtain the search results. In addition, the queries are also recorded for further analysis.

Keywords: *Search Interface, Search Behaviour, Search Log, Web Searching*

1. Introduction

Information environment can be categorized into conventional/traditional and web environment (Park et al., 2005). In these environments, even though the users share the same aim, but the behaviours exhibited by the users are different. A library is an example of conventional/traditional information environment. Searching for information in a library requires several activities such as user has to refer to the collection catalogue or the librarian to allocate the material, find the rack that situates the material, and scan all the materials to find the correct one and finally pick-up the wanted material. Searching in this environment is a tedious work and time consuming. Furthermore, there is a possibility that the material is not in place.

The web environment offers a flexibility that does not exist in conventional/traditional environment. In web environment, the information is vast, unlimited and the availability is guaranteed except when technical problem arises. The retrieval of the information can be made at any time without physical constraint. Locating the information is simply by a simple query through the search system. In this environment, the typical search behaviours are browsing, formulating and reformulating query, evaluating the search results, interface navigation, etc.

Search behaviour is a strategy undertaken by the user in searching for information. Nachmias and Gilad (2002) defined search behaviour or search strategy as a user plan that consists of a series of actions (steps), aimed at searching information. In order to successfully find useful information on the web, users need to consider the usage of their searching strategies to generate better outcome (Tu et al., 2008). Zhang et al., (2005) define search behaviour as the micro level of behaviour when a user interacts with a specific information retrieval system to search for relevant information.

The behaviours that are related to the Internet searching are search tactics (Thatcher, (2006); Wildemuth, (2004); Hong et al., (2002)), search strategy (White and Iivonen, (2001)), seeking behaviour (Asunka, Chae, Hughes and Natriello, (2009); Thatcher, (2006); Nesset, (2005); Seiden et al., (1997)) and search behaviour (Nachmias and Gilad, (2002)). These terms are used interchangeably in Internet and information searching by means to describe the behaviour of the users in order to achieve their goals.

The search behaviour on the web environment is mainly influenced by the search query. The search query is a set of words or phrases a user enters when looking for information on a specific topic or subject (Rieh and Xie,

2001). This paper focuses on user search behaviour that is query formulation and reformulation. Users frequently modify their queries in hope of retrieving better results (Huang and Efthimiadis, 2009). Formulating a search query is a challenging task for most of users because they are required to express their information need to search system.

Query formulation is the initial stage in which the search strategy is constructed and the following reformulation stage in which the initial stage is modified either manually or system assisted (Mastora, Monopoli and Kapidakis, 2008). Therefore, it is not easy for users to choose the right search query that represents the topic or subject that they are looking for. This behaviour is in line with Spink et al., (1998) which indicate that users tend to employ simple search strategies and conduct successive searches (changes or shifts in beliefs and cognitive, affective and situational states) over time to find information related to a particular topic.

2. Literature Review

User search behaviour is influenced by the knowledge possess by the web search user. The effects of user knowledge on user search behaviour have been shown by several studies such as a study on engineering and science students searching (Zhang et al., 2005), a study on the way children search for information (Hirsh, 1995) and travellers' information search behaviour (Gursoy and McCleary, 2004). In addition, users who utilize their knowledge during the searching process are overall most successful in their searching (Holsher and Strube, 2000). User knowledge will typically influence the query formulation and reformulation which apply as the search tactics which changed over time as the users' knowledge changed (Hong et al., 2002; Wildemuth, 2004).

Users that are knowledgeable on certain topics tend to focus on their searching. They will typically, begin with general search terms and then use more specific search terms to narrow their searching (Thatcher, 2006). Additionally, knowledge on the web search system such as the search features, tools, and the functionalities will also enhance the search process. Users with subject knowledge in the domain know how to select terms that make them effective information users. Bhavnani (2002) observes healthcare experts doing the searching within and outside their domains of expertise. When performing tasks within their area of expertise, the experts used declarative and procedural components of domain specific search knowledge. These enable them to perform effective searching. In contrast, when they performed tasks outside their domains of expertise, they used a range of general purpose search results. As prior studies (Duggan and Payne, 2008; White et al., 2008; Zhang et al., 2005; Wildemuth, 2004) have shown that user knowledge does have an influence on user search behaviour.

Spink and Jansen (2004) reveal patterns and trends in general web searching in term of human computer interaction and query level of web search behaviour namely; (1) most web queries are short, without query reformulation or modification, and have a simple structure, (2) few search sessions include advanced search techniques and when they are used, many include mistakes, (3) many web users retrieve a large number of web sites, but users' view few result pages and generally view about five web documents.

Various researches have been conducted by focusing on the search query such as; Tu et al., (2008), Zhang et al. (2005), Kim (2005), Park et al., (2005), Nachmias and Gilad (2002) and Rieh and Xie (2001). Rieh and Xie (2001), study on patterns and sequences of query reformulation based on query logs from a web search engine. From the study, there are six patterns of query reformulation; specified reformulation, parallel reformulation, generalized reformulation, dynamic reformulation, format reformulation and alternative reformulation. Nachmias and Gilad (2002) provide different classification of the queries that are keyword search, wide search definition, general knowledge, complex search, computer convention and Boolean search. The classification by these two studies is somewhat redundant. For example specified reformulation is identical to keyword search and generalized reformulation is identical to wide search definition.

Other studies tend to highlight the search behaviour based on query analysis as a part of the measurement such as the number of words in the query and the thesaurus terms used (Park et al., 2005; Tu et al., 2008). The number of attempts which includes number of searches and query formulation has also being considered as one of the measurements (Zhang et al., 2005; Kim, 2005). Additionally, the usage of advance features of web search engine is also another potential measurement (Nachmias and Gilad, 2002; Park et al., 2005).

3. Method for Capturing User Search Behaviour

This study involves the use of computer and the Internet technology to capture the actual user search behaviour. The use of computer technology in this research is time consuming and expensive (Hochstotter and Koch, 2009). As a consequence, this study has to deal with a number of problems such as the availability of computer laboratory, computer workability, Internet connection, space limitation, the availability of the respondents and time allocation. Therefore, the success of this study depends on the assistance of the technical support from the computer department. The respondents were instructed to perform the online searching related to search task given. During this session, the users performed searching through the search interface. The users' queries were stored in the transaction log.

Theoretically, the search task will influence the information need (Broder, 2002). Search tasks were designed to be subject neutral and broad in scope so that all the participants could find relevant information regardless of their subject areas. In this study, the search task chosen is related to Malaysian's homestay. Homestay is considered as general knowledge as homestay is a worldwide concept.

In this study in order to capture and record user's query, an interface called search interface has been developed (Yamin, et al., 2013). This interface act as a proxy by interfacing Google search engine. The interface will receive user's query, record in the search log and redirect the query to the Google search engine. The Google will process the query and return the results. Through this interface the user's query can retrieved and use in the analysis to determine the user search behaviour. Figure 1 ilustrate the procedure for using the search interface.

As shown in Figure 1, the first step of using the search interface is to run the web browser. Web browser is application software that installed in user's personal computer. Various web browsers are available and freely downloaded from the Internet. Example of the web browsers are Internet Explorer, Opera, and Mozila Firefox. Figure 2 shows example of web browser Internet Explorer. Internet Explorer (IE) is default web browser software that installed together with the Windows. To run the IE, user may click on the icon that display the image of an "e" or select shortcut from the [Start] menu.

When the IE is open, the user will has to enter the URL of the search interface. Figure 3 shows the location where the URL should be entered. Once the URL is entered, the reference number page will be loaded (Figure 4). The reference number page is the first part of the search interface. On this page user will has to enter his matrix number. The matrix number is use as a reference number in the search log to represent the user. During the searching session, all queries that the user entered will be associated with the reference number and store in the search log.

After the reference number is entered, the search page will be loaded. The page is shown in Figure 5. The slot with a circle is the location where user has to enter his query. The searching interface will records the query, the start time and sent the query to Google for processing and displaying results. This interface does not modify the query or delay the search process as it only records the query and then redirects the query to the Google search engine. This interface consists of two main parts. The upper part with the blue background is a section where students can enter their queries. The lower part is where the Google interface and results are displayed. When the students enter query in the blue area, the query will be time stamped and stored in the database. The query is submitted to Google which then returns a list of search result.

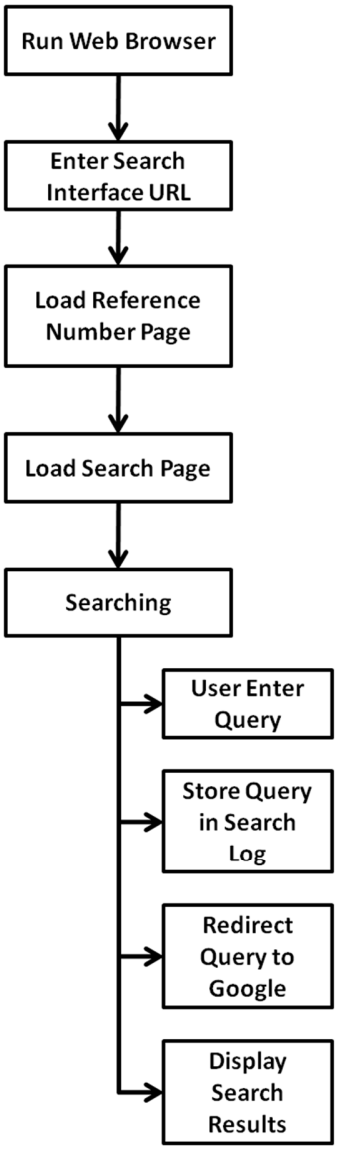


Figure 1: Procedure for Using the Search Interface

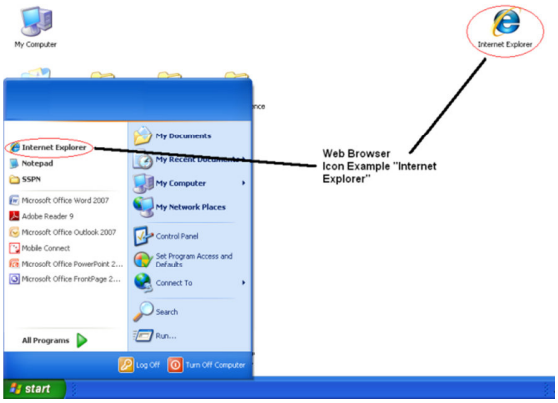


Figure 2: Running the Web Browser

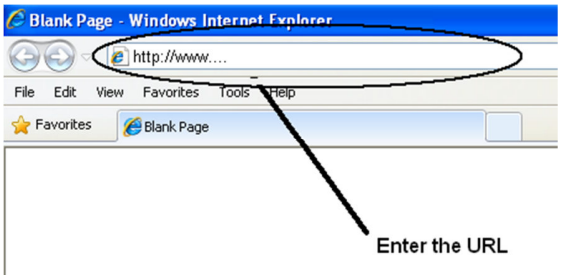


Figure 3: IE - Entering the URL

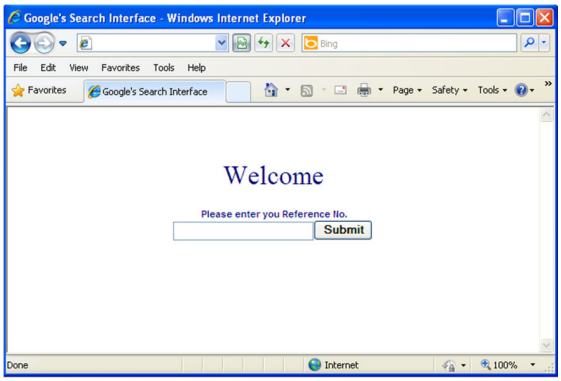


Figure 4: Search Interface - Reference Number Page

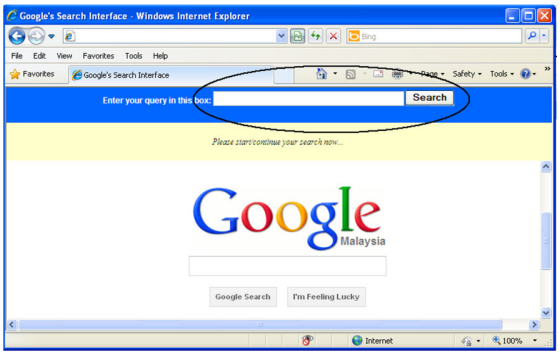


Figure 5: Searching Interface - Searching Section

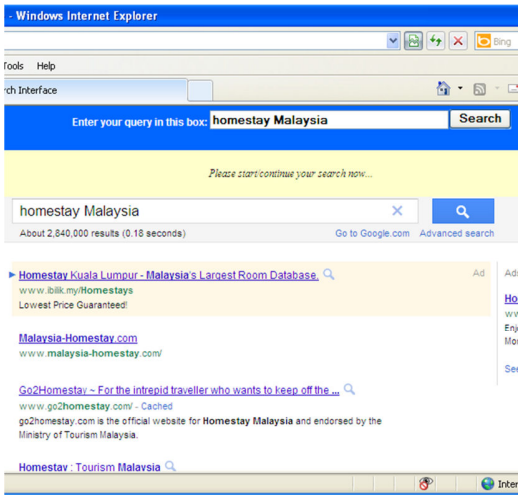
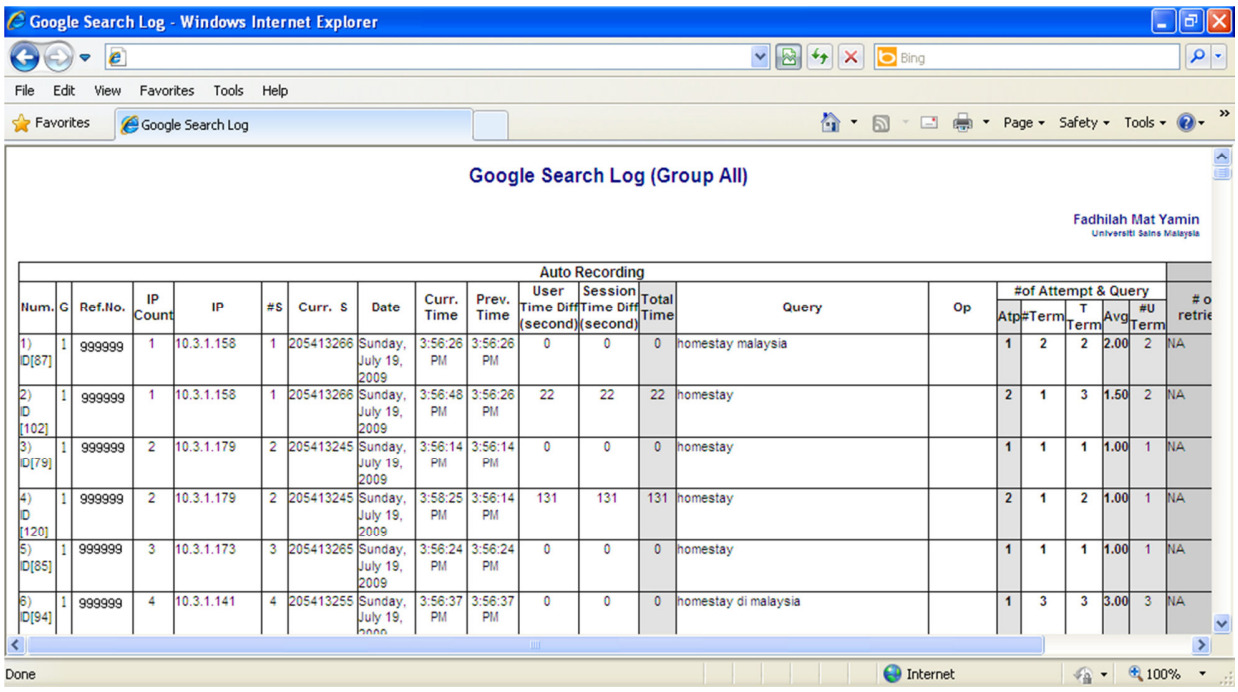


Figure 6: Example of Search Session

Figure 6 shows example of the search result for the query "homestay malaysia". The students were advised to browse the results to find answers for their search task. They were also advised to repeat the searching to find better answer. During this session, students would perform query reformulation that was the adjustment of the query in order to get better results.

Users' queries and other information from the searching session are recorded in the transaction log or search log. Figure 7 shows the example of the search log that was captured through the search interface. The search log contains information about the user and the computer used such as user ID and computer IP and information about the search session which includes the session ID, date and time. Other item in the log such as time different, IP and session counter, number of attempts and queries, number of terms, terms average and number of unique terms were calculated by the system. Table 1 shows the list of items in the log and its description. In this study, only queries were taken for analysis. Other information was used as a reference.



Auto Recording

Num.	G	Ref.No.	IP Count	IP	#S	Curr. S	Date	Curr. Time	Prev. Time	User Time Diff (second)	Session Time Diff (second)	Total Time	Query	Op	#of Attempt & Query Atp	#Term	T Term	Avg	#U Term	# o retrie
1) ID[87]	1	999999	1	10.3.1.158	1	205413266	Sunday, July 19, 2009	3:56:26 PM	3:56:26 PM	0	0	0	homestay malaysia		1	2	2	2.00	2	NA
2) ID [102]	1	999999	1	10.3.1.158	1	205413266	Sunday, July 19, 2009	3:56:48 PM	3:56:26 PM	22	22	22	homestay		2	1	3	1.50	2	NA
3) ID[79]	1	999999	2	10.3.1.179	2	205413245	Sunday, July 19, 2009	3:56:14 PM	3:56:14 PM	0	0	0	homestay		1	1	1	1.00	1	NA
4) ID [120]	1	999999	2	10.3.1.179	2	205413245	Sunday, July 19, 2009	3:58:25 PM	3:56:14 PM	131	131	131	homestay		2	1	2	1.00	1	NA
5) ID[85]	1	999999	3	10.3.1.173	3	205413265	Sunday, July 19, 2009	3:56:24 PM	3:56:24 PM	0	0	0	homestay		1	1	1	1.00	1	NA
6) ID[94]	1	999999	4	10.3.1.141	4	205413255	Sunday, July 19, 2009	3:56:37 PM	3:56:37 PM	0	0	0	homestay di malaysia		1	3	3	3.00	3	NA

Figure 7: Example of Search Log

Table 1: Log Item and Description

Column	Item	Description
1	Num (and record ID)	Num is a continuous line number and the record ID is a reference number of the record in database
2	G (Group number)	Indicate the group number
3	Ref. No.	Ref. No. is the user ID that is used as a reference for the particular user.
4	IP Count	Counting the number of IP –the counter increase when new IP found
5	# S (Session)	Counting the number of session - the counter increase when new session found
6	Curr. S	Shows the current session
7	Date	Shows the date
8	Curr. Time	Shows the time of the current search session
9	Prev. Time	Shows the time of the previous search session
10	User Time Diff (second)	Shows the time different (in second) for each user based on current and previous search session
11	Session Time Diff (second)	Shows the time different (in second) for each session based on current and previous search session
12	Total Time	Total time taken by each user to complete the search task
13	Query	Query entered by user
14	Op (Operator)	Boolean operator used
15	# of Attempt & Query	Summarize the query used by each user
15 (a)	Atp (Attempt)	Shows the number of attempt made by user
15 (b)	# Term	Shows the number of term used
15 (c)	T Term	Shows the total number of the terms
15 (d)	Avg (Average)	The query average.
15 (e)	# U Term	Number of unique terms in the query

Data Processing and Coding: The search log is undergoing some processes to clean the data. Users who did not perform query formulation were removed. These users were recognized based on the query entered. Typically, users are expected to enter more than one query which shows that they formulate and reformulate the queries. Otherwise, the users are suspected browsing that is visiting each link exhaustively. Browsing activity is not recorded in the transaction log. The irrelevant queries were also identified and removed. Then query was classified either as breadth or depth search query. The classification was based on the criteria in Table 2 and Table 3.

Table 2: Criteria for Breadth Search Query

Coding Symbol	Strategy	Description	Example
B1	Keyword search	Direct typing the query subject	Typing the words Homestay
B2	Wide search definition	Searching using a broad query	Searching for Ministry of Tourism to find the Homestay
B3	General knowledge	Using information that is not mentioned in the search task	Searching for the Homestay mentioning Guest House.

Table 3: Criteria for Depth Search Query

Coding Symbol	Strategy	Description	Example
D1	Boolean search	Using Boolean syntax	Homestay AND Pahang
D2	Computer convention	Using a computer convention	Homestay.gif, homestay. jpeg
D3	Complex search	Cross searching with more than one query	Homestay, jungle tracking, fishing, etc.

The following are the steps of the coding process.

Step 1: The transaction log is retrieved from the database. Examples of the log have been shown in Figure 7.

Step 2: The transaction log is sorted according to the group, date/time, computer IP and session number. Example in Figure 7 has been sorted based on these criteria.

Step 3: Data cleaning is performed to remove single query user and irrelevant queries. Figure 8 shows example user with at line 2 (ID 102) only entered one query. This user was suspected to do browsing which was not recorded in the transaction log. This user was discarded from the list.

1) ID[87]	1	999999	1	10.3.1.158	1	205413266	Sunday, July 19, 2009	3:56:26 PM	3:56:26 PM	0	0	0	homestay malaysia
2) ID [102]	1	999999	1	10.3.1.158	1	205413266	Sunday, July 19, 2009	3:56:48 PM	3:56:26 PM	22	22	22	homestay
3) ID[79]	1	999999	2	10.3.1.179	2	205413245	Sunday, July 19, 2009	3:56:14 PM	3:56:14 PM	0	0	0	homestay

Figure 8: Example of Single Query User

During the search process, users were expected to formulate and used queries that were related to the homestay. Users may enter many queries. Each query will is time stamped so that the sequence of the query can be traced. Users may also enter the same query. This shows that users are returning to the previous results list. Figure 9 shows example of user that has entered more than one query. Queries in line 8 to 12 were entered by the same user at different time.

Further cleaning process is to find and remove the irrelevant queries. This procedure is done manually. Figure 10 shows example of irrelevant query. At line 1019 used the word "library uum" as the query. This query was not relevant to homestay, therefore, this particular record was removed from the list.

Step 4: Query classification or marking is performed. In this step, each query was examined and classified either as breadth or depth query. Table 4 shows example of queries for user A and B. During the coding, each query strategy was denoted with a symbol as shown in Table 2 and Table 3. The breadth search queries are represented by symbols B1, B2 and B3, where each symbol represents keyword search, wide search and general knowledge respectively. Depth search queries are represented by symbols D1, D2, and D3. D1 represents the Boolean search, while D2 and D3 represent computer convention and complex search respectively.

Step 5: The classified queries are then transferred into the table and the frequency of each query type was calculated. For the ease of the analyses in SPSS, the search logs were transferred into a table (Figure 11). For example, user A and user B were among the respondents in this study. Queries by both users were transferred into the table. Based on the query occurrences, user A formulated eight queries, showing that user A had eight attempts. Out of these attempts, five queries fell under breadth search strategy, while the other three were depth search strategy. Further classification has shown that five queries identified as breadth search strategy

can be divided into direct keyword (B1), wide search (B2), and general knowledge (B3) by which each strategy represent 2, 2, and 1 queries respectively. Three queries under the depth search query are classified as complex search strategy (D3). Other types of queries that are Boolean operator (D1) and computer convention (D2) were not used by this user. User B on the other hand has formulated nine queries. Seven of the queries are falls under breadth search query and the other two are depth search query.

7) ID [171]	1	999999	4	10.3.1.141	4	205413255	Sunday, July 19, 2009	4:04:29 PM	3:56:37 PM	472	472	472	homestay di malaysia
8) ID[91]	1	999999	5	10.3.1.175	5	205413257	Sunday, July 19, 2009	3:56:30 PM	3:56:30 PM	0	0	0	homestay
9) ID [107]	1	999999	5	10.3.1.175	5	205413257	Sunday, July 19, 2009	3:57:00 PM	3:56:30 PM	30	30	30	homestay pahang
10) ID [160]	1	999999	5	10.3.1.175	5	205413257	Sunday, July 19, 2009	4:03:34 PM	3:57:00 PM	394	394	424	homestay kehidupan kampung
11) ID [168]	1	999999	5	10.3.1.175	5	205413257	Sunday, July 19, 2009	4:04:08 PM	4:03:34 PM	34	34	458	homestay kedah
12) ID [187]	1	999999	5	10.3.1.175	5	205413257	Sunday, July 19, 2009	4:05:49 PM	4:04:08 PM	101	101	559	homestay homestay selangor
13) ID [101]	1	999999	6	10.3.1.147	6	205413261	Sunday, July 19, 2009	3:56:48 PM	3:56:48 PM	0	0	0	homestay

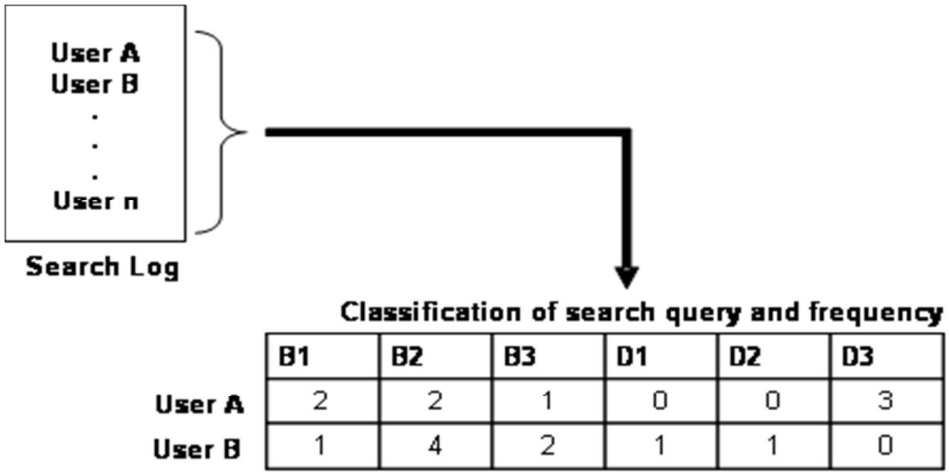
Figure 9: Example of Multiples Query

1018) ID [1007]	5	999999	165	10.3.1.178	191	910941422	Saturday, July 25, 2009	2:06:31 PM	2:01:36 PM	295	295	615	homestay in malaysia
1019) ID [918]	5	999999	166	10.3.1.148	192	910941098	Saturday, July 25, 2009	12:31:53 PM	12:31:53 PM	0	0	0	library uum
1020) ID [954]	5	999999	167	10.3.1.85	193	910941410	Saturday, July 25, 2009	1:57:24 PM	1:57:24 PM	0	0	0	village jungle tracking fishing visit peta

Figure 10: Example of Irrelevant Query

Table 4: Example of Queries and Classification

User	Query	Classification
User A	Homestay	B1
	Homestay Malaysia	B2
	Homestay	B1
	Homestay in Kedah	B2
	Rumah tumpangan	B3
	Homestay, kedah, jungle tracking, makanan tradisional	D3
	Homestay, Malaysia, tradisional	D3
	Sarawak, jungle traking, package	D3
User B	Website Homestay	B1
	Homestay in Malaysia	B2
	Homestay Selangor	B2
	Perak homestay	B2
	Guest house	B3
	Kelantan Guest House	B3
	Malaysia Homestay Aktiviti	B2
	Homestay AND Terengganu	D1
	Peta ke homestay Kuala Medang	D2



	B1	B2	B3	D1	D2	D3
User A	2	2	1	0	0	3
User B	1	4	2	1	1	0

Figure 11: Example of the classification of search query and frequency

Step 6: Calculate mean for both breadth and depth search query. Example is shown in Table 5. These mean values are then used in the statistical analysis.

Table 5: Mean Value for Breadth and Depth Search Query

User	Mean Breadth	Mean Depth
User A	1.667	1
User B	2.333	0.667

The processes explained in Step 1 – 6 are summarised in Figure 12.

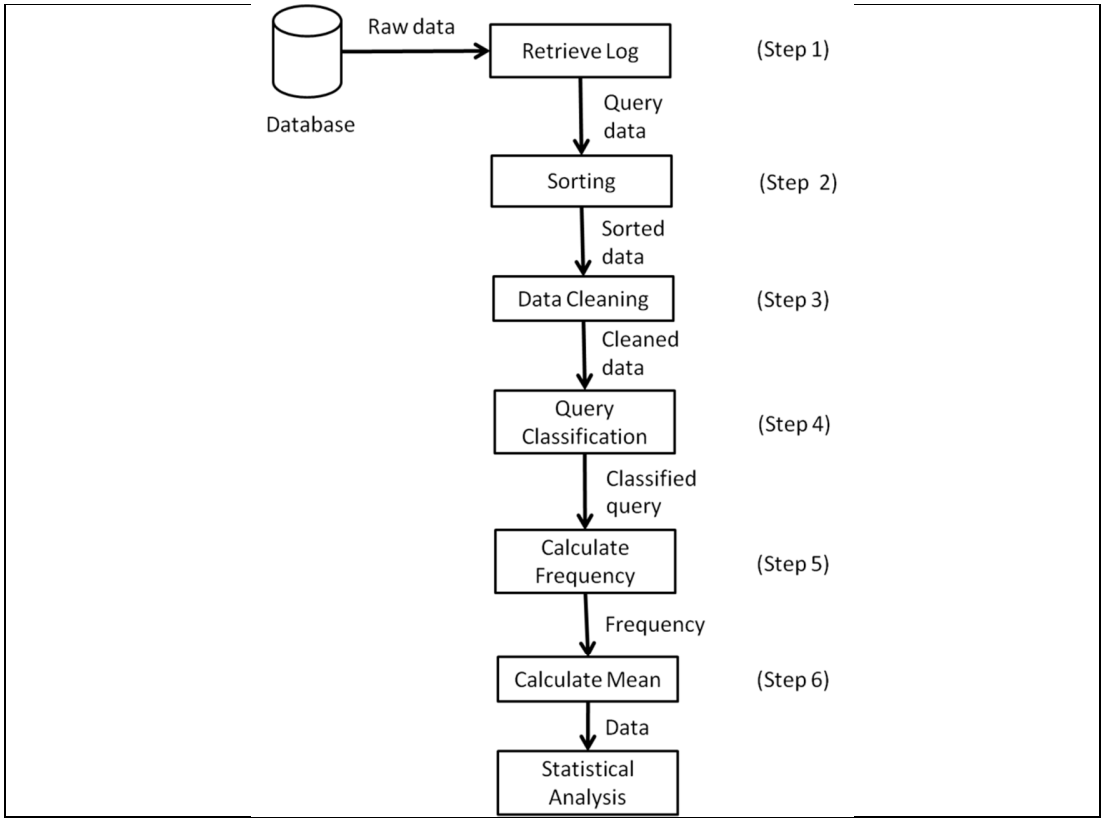


Figure 12: Coding Process of Query

4. Example of Data Collection and Descriptive Analysis

Due to the time and space constraints, only 335 students were randomly selected as the sample. From the 335 students, only 239 students took part in this study. Out of 239 students, 30 students were selected as a pilot test group. The rest were allocated for actual data collection.

The respondents were divided into six groups. Each group was limited to 35 respondents as it is the maximum capacity of the computer lab. The time average for each group is approximately one hour. Overall, 195 respondents attended the allocated session, while 14 were absent. Out of 195 questionnaires received, 64 questionnaires were discarded because of invalid answers, technical problem with the server that hosted the interface system, and no query formulation (single query entered). The affected number of respondents was 1, 33, and 30 respondents respectively. The percentage of the total response rate is 93.3 %; however the usable response rate is only 67.2 %.

In this study, a total 1,072 queries are used for the analysis. Out of these totals, 596 queries are breadth search query and 476 are depth search query. The mean scores of depth search query and breadth search query are 1.21 to 1.52 respectively with the same standard deviation value which is 1.03. Based on these findings, it might be obvious to say that the breadth search query is moderately higher compared to depth search query.

Table 6 shows the total query for each category. Table 7 shows specific mean values and standard deviation for each user search query categories. For the breadth search query, the mean values are ranged from .97 to 2.36. Among the query categories in the breadth search query, general knowledge achieved the highest mean value 2.36. The lowest mean value is the wide search category with the mean value .97.

The second group of search query that is depth search query, the mean values are ranged from .62 to 1.63 and the standard deviation score is from .77 to 1.84. In this category, the complex search shows the highest mean value followed by computer convention and Boolean operator.

Table 6: Total Query for Each Category

Query Categories	Query	Total
Breadth search query	Direct search	162
	Wide search	127
	General knowledge	307
Depth search query	Boolean operator	81
	Computer convention	181
	Complex search	214

Table 7: Mean values and standard deviation for Query search formulation

Query categories	Query	Mean	Std. deviation
Breadth search query	Direct search	1.24	1.39
	Wide search	0.97	1.21
	General knowledge	2.36	2.86
Depth search query	Boolean operator	0.62	0.77
	Computer convention	1.38	1.40
	Complex search	1.63	1.84

5. Conclusion

This method proposed in this paper provides in-depth understanding about the users and their behaviour when searching on the Internet. The search task contains instruction or information about the searching activity. The users were instructed to read and understand the task. The understanding will form the information need which will be expressed in the verbal from. As highlighted by Broder (2002), the verbal form of information need will be entered into the search system as the search query. The searching was made through the search interface that will capture the query and store into the transaction log. The series of query stored in the transaction log represent the users' behaviour (selecting, changing and modifying the query) during the searching.

The analysis of this log is based on the steps or procedure as outline in this paper. This procedure is novel as it was not previously discussed in the literature. Additionally, the deployment of search interface as a tool to capture users' behaviour by interfacing the Google search engine is also a novel approach. This approach has made the data collection easier as the transaction log was stored in the local server.

Acknowledgment

The authors wish to thank the Research Management & Innovation Center (RIMC), Universiti Utara Malaysia for funding this study under University Research Grant Scheme.

References

Asunka, S., Chae, H. S., Hughes, B., & Natriello, G. (2009). Understanding Academic Information Seeking Habits through Analysis of Web Server Log Files: The Case of the Teachers College Library Website. *The Journal of Academic Librarianship, 35(1)*, 33-45.

Bhavnani, S.K. (2002). Important Cognitive Components of Domain Specific Search Knowledge, Proceedings of TREC'2001, 571-578.

Broder, A. (2002) A Taxonomy of Web Search. *SIGIR Forum*, *36(2),* 3–10.http://scholar.google.com/scholar?hl=en&lr=&q=A. Broder.+A+taxonomy+of+web+search.&btnG=Search

Duggan, G. B., & Payne, S. (2008). Knowledge in the Head and on the Web: Using topic Expertise to Aid Search. Conference on Human Factors in Computing Systems. *Proceeding of the twenty-sixth annual SIGCHI*

conference on Human factors in computing systems, Florence, Italy, Pages: 39-48, ISBN:978-1-60558-011-1.

Gursoy, D., & McCleary, K.W. (2004). Traverlers' Prior Knowledge and its Impact on Their Information Search Behavior. *Journal of Hospitality and Tourism Research, 28*, 66-94.

Hirsh, S.G. (1995). The Effect of Domain Knowledge on Elementary School Children's Search Behaviour on an IR System: The Science Library Catalogue. *Proceedings of Conference on Companion on Human Factors in Computing Systems, Denver, Colorado, US,* 55-56.

Hochstotter, N., & Koch, M. (2009). Standard Parameters for Searching Behaviour in Search Engine and Their Empirical Evaluation. *Journal of Information Science, 35(1),* 45-65.

Holscher, C., & Strube, G. (2000). Web Search Behavior of Internet Experts and Newbies. *Computer Networks, 33*, 337-346.

Hong, W., Thong, J.Y. L., Wong, W., & Tam, K. (2002). Determinants of User Acceptance of Digital Libraries: An Empirical Examination of Individual Differences and System Characteristics. *Journal of Management Information Systems*, *18(3),* 97-124.

Huang, J., & Efthimiadis, E. N. (2009). Analyzing and Evaluating Query Reformulation Strategies in Web Search Logs. Conference on Information and Knowledge Management (CIKM'09). *Proceeding of the 18th ACM Conference on Information and Knowledge Management,* Hong Kong, China, 77-86.

Kim, J. (2005). Task Difficulty in Information Searching Behavior: Expected Difficulty and Experienced Difficulty. *Proceedings of the 5th ACM/IEEE-CS Joint Conference on Digital libraries*, 383.

Mastora, A., Monopoli, M., & Kapidakis, S. (2008). *Exploring Query Formulation and Reformulation: A Preliminary Study to Map Users' Search Behavior.* Research and Advanced Technology for Digital Libraries, Springer Berlin / Heidelberg, 427-430.

Nachmias, R., & Gilad, A. (2002). Needle in a Hyper stack: Searching Information on the World Wide Web. *Journal of Research on Technology in Education, 34(4)*, 475-486.

Nesset, V. (2005). An Exploratory Study Into The Information-Seeking Behaviour Of Grade-Three Students. In L. Vaughn, (Ed.), *Data, Information, and Knowledge in a Networked World. Proceedings of the Canadian Society for Information Science and Technology.* June 2-4, 2005, London

Park, S., Lee, J. H., & Bae, H. J. (2005). End User Searching: A Web Log Analysis of NAVER, a Korean Web Search Engine. *Library of Information Science Research, 27*, 203-221.

Rieh, S.Y., & Xie, H. (2001). Patterns and Sequences of Multiple Query Reformulations in Web Searching: A Preliminary Study. *Proceedings of the 64th ASIST Annual Meeting, 38,* 246-225.

Seiden, P., Szymborski, K., & Norelli, B. (1997). Undergraduate Students in the Digital Library: Information Seeking Behavior in a Heterogeneous Environment. *http://news.ala.org/ala/acrlbucket/nashville1997pap/ seidenszymborski.htm, download 2007-04-21.*

Spink, A., & Jansen, B. J. (2004). A Study of Web Search Trends. *Webology,* 1(2), Available online at http://www.webology.ir/2004/v1n2/a4.html, retrieved on April 24, 2007.

Spink, A., Bateman, J., & Jasen, B. J. (1998). Searching Heterogeneous Collections on the Web: Behavior of Excite Users. *Information Research,* 4(2), Retrieved from http://informationr.net/ir/4-2/paper53.html on April 24, 2007.

Thatcher, A. (2006). Information-Seeking Behaviours and Cognitive Search Strategies in Different Search Tasks on the WWW. *International Journal of Industrial Ergonomics,* 36, 1055-1068.

Tu, Y., Shih, M., & Tsai, C. (2008). Eight Graders' Web Searching Strategies and Outcomes: The Role of Task Types, Web Experiences and Epistemological Beliefs. *Computers & Education,* 51(3), 1142-1153.

White, M. D., & Iivonen, M. (2001). Questions as a Factor in Web Search Strategy. *Information Processing & Management*, *37(5),* 721-740.

White, R. W., Dumais, S., & Teevan, J. (2008). How Medical Expertise Influences Web Search Interaction. In *Proceeding of the 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval,* Singapore, 791-792.

Wildemuth, B.M. (2004). The Effects of Domain Knowledge on Search Tactic Formulation. *Journal of the American Society for Information Science and Technology,* 55(3), 246-258.

Yamin, F.M., Ramayah, T., Ishak, W.H.W. (2013). Search Interface to Capture Searchers Behaviour. International Journal of Computing Academic Research, 2(2), pp: 67-74

Zhang, X., Anghelescu, Hermina, G.B., & Yuan, X. (2005). Domain Knowledge, Search Behaviour, and Search Effectiveness of Engineering and Science Students: An Exploratory Study. *Information Research,* 10(2), 1-17. Available online at http://infrmationr.net/ir/10-2/paper217.html retrieved on June 3, 2007.
Zhang, X., Anghelescu, Hermina, G.B., & Yuan, X. (2005). Domain Knowledge, Search Behaviour, and Search Effectiveness of Engineering and Science Students: An Exploratory Study. *Information Research,* 10(2), 1-17. Available online at http://infrmationr.net/ir/10-2/paper217.html retrieved on June 3, 2007.