International Journal of Computing Academic Research (IJCAR)
ISSN 2305-9184 Volume 2, Number 2 (April 2013), pp. 67-74
© MEACSE Publications
http://www.meacse.org/ijcar

Search Interface to Capture Searchers Behaviour

Fadhilah Mat Yamin[1], T. Ramayah[2] and Wan Hussain Wan Ishak[3]

[1]School of Technology Management & Logistic, College of Business, Universiti Utara Malaysia, Malaysia

[2]School of Management, UniversitiSains Malaysia, Malaysia

[3]School of Computing, College of Arts and Sciences, Universiti Utara Malaysia

Abstract

Study on searchers behaviour on the internet is one of the interests among internet researcher. Typically, many methods have been used to study the searcher behaviour such as questionnaire and search log analysis. Search log analysis provide in depth information on the keywords that have been used, change or manipulation of the keywords, the number of attempt, and the time taken to complete the searching. However, the search log is own by the search engine provider, thus obtaining the search log might be difficult. To overcome the problem, new method called search engine interfacing is introduced. This method is aim to interface the search engine using an interface that can accept keywords enter by searcher and send the keywords to the search engine for the results. Before sending the keywords to the search engine, the keywords and other related information will be stored in the database. The search interface has been implemented and used to study user search behaviour based on the given search task. The searching exercise shows that the search interface has successfully capture searchers keywords that reflect their search behaviour.

Keywords: Information Retrieval System, Search Engine, Search Log, Search Interface, Search Behaviour

Introduction

Studying searchers behaviour throughweb search log is vital especially to thesearch providers. The analysis providesin-depth understanding of the searchersrequirement. Furthermore, it is wellunderstood that the technology will failif it does not reflect the user's need [1].Thus, improvement can be made eitherimproving the web search facility orimproving the document database.Currently, there are many studiesincorporate search log to keep trackuser searching behaviour such as [1,2,3,4].

Search log is a timestamps electronic record of interactions between searcher and the web search engine [5]. The log contains searchers details such as IP and session ID and information searched which includes the search streams, terms, operators and etc. The information contains in the log can be used to study and understand the searchers searching behaviour [1,5].

According to Wang et al. [2], search log are accurate, unobtrusive, longitudinal, transactional, temporal and can be automatically collected and processed. In addition, [2] also highlighted some disadvantages of search log such as the data being open to interpretation (accurate or not), privacy concern, and the vast amount of data gathered can be difficult to manage. However, these advantages are not the main issue in search log research. It depends on the method and skill of the researcher to analyse the log and to ensure that it is as accurate as it should be.

However obtaining the log is "expensive" as it was not intended for public access. Therefore, alternative method is proposed to record the usage and creating researcher's own search log. In this paper a framework for interfacing search engine is proposed to capture user search activities. Google is used as the case study.

Information Searching and User's Behaviour

Information search on the web is a complex process. The components of information seeking and searching processes as proposed by [6, 7, 8] are similar but not identical. The major difference of those processes is the sequence of the execution of the components. Correspondingly, in every model the main components are the identification of the problem or search task analysis, information need articulation, formulation of the query, results evaluation, and decision to repeat or to stop the searching. Theoretically, users will stop searching when they have found what they are looking for or feel satisfied with what they have achieved.

Problem identification starts with task in hand that users have to search for. According to [9] the task will determine the information need which is verbalized and translated into a query posed to a search system. At this stage users need to understand the task. The complexity of the search tasks is also an important factor in users' ability to find relevant information and their satisfaction [10]. Complex task might be difficult to understand compared to less complex task.

Information need is the perceived need for information. This need leads to the use of information retrieval system to get the information [11]. Information need is also associated with the search task. The task particularly will state the kind of information that the user should acquire. Allen [12] raises a question "how can users express their information needs in their own terms and still obtain information that will meet their information need?" The Allen question is concerned with users' knowledge and strategy to address their need. In particular, different user might use different set of queries to achieve the same need.

Once the information need has been identified, the next step is how to represent the information need to suitable query. Queries are considered as formal statements of the information needs therefore, the quality of information retrieval depends on the user formulated query. The length of the query for example will influence the search results. Short queries are used to initiate the search when the users are not familiar with the subject [13]. This shows that the effect of user knowledge on query formulation. In contrast to short query, long queries can be used to address more specific need of the user. This query allows users to naturally and fully describe their information need [14]. As [14] have demonstrated, a long query in web environment is practical and can substantially improve the quality of information retrieval. Therefore, understanding and knowing how to formulate the query will benefit best the user.

Query reformulation is a modification to a search query that addresses the same information need [14, 15]. According to [15], examples of query reformulation are word reorder, white apace and punctuation, word removal, word addition, acronym formation and expansion, substring, abbreviation, word substitution and spelling correction. Users can also benefit from an improved search experience when performing reformulation [15]. Experience is a kind of knowledge that is produced from repeating process of searching. After the search session, user will typically update his or her knowledge about the query manipulation and how to use the search system.

Query reformulation also is a part of user's strategy to improve the search results [16;17]. This strategy is also called user's behaviour[18]. Nachmias and Gilad[18] define search behaviour as a user plan that consists of a series of actions (steps), aimed at searching information and satisfaction of the search result. The search results are considered relevant to the users when it matches the query entered during the search session [19].

Google Search Engine

Google is the general purpose search engine and one of the widely used search tools on the Internet [20, 21]. Google popularity is due to the number of reasons such as wide coverage and updated regularly, fast in access, provide user friendly interface, provide links to websites world over and separate interface for searching journals, images, news, audio and etc[20]. Figure 2 shows Google main interface.

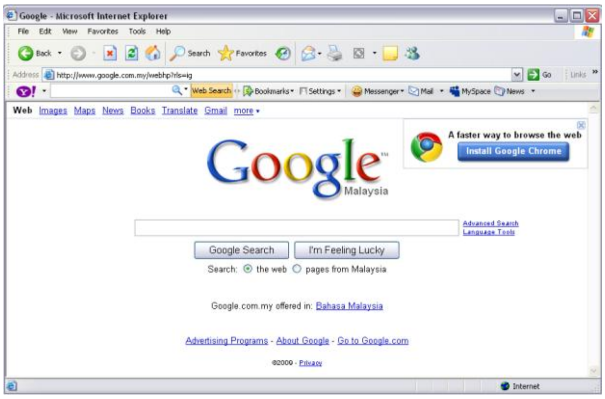

Google - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back
Search
Favorites
Address http://www.google.com.my/webhp?rls=ig
Go
Links
Web Search
Bookmarks
Settings
Messenger
Mail
MySpace
News
Web Images Maps News Books Translate Gmail more
A faster way to browse the web
Install Google Chrome
Google
Malaysia
Advanced Search
Language Tools
Google Search
I'm Feeling Lucky
Search: the web pages from Malaysia
Google.com.my offered in: Bahasa Malaysia
Advertising Programs - About Google - Go to Google.com
©2009 - Privacy
Internet

Figure 2. Google Search Engine

Framework of Interfacing Google Search Engine

Users' activities on Google are recorded in the Google's search log. However, this log was not available for public access. Furthermore, Google process millions of query every day to facilitate the searching activities. Therefore, getting the log from Google is not best decision. In this study in order to capture and record user's query, an interface called search interface has been developed. As defined by [2] interface is a layer between the user and the system that facilitates human computer communication. 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. Google will process the query and return the results. Through this interface the user's query can be retrieved and use in the analysis to determine the user search behaviour.

Figure 3 shows a model of the proposed search interface. Search interface consist of search interface engine and reporting module. Query entered by the searcher will be stored into a database and forwarded to Google. The query will not be modified. It will be forwarded as it is.

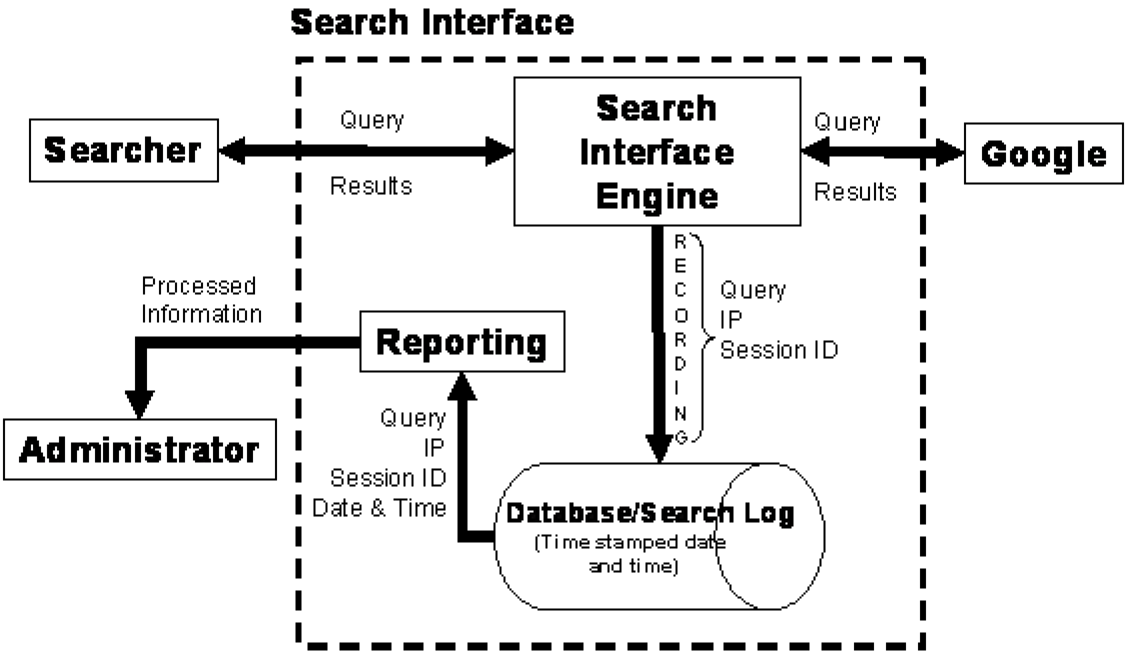
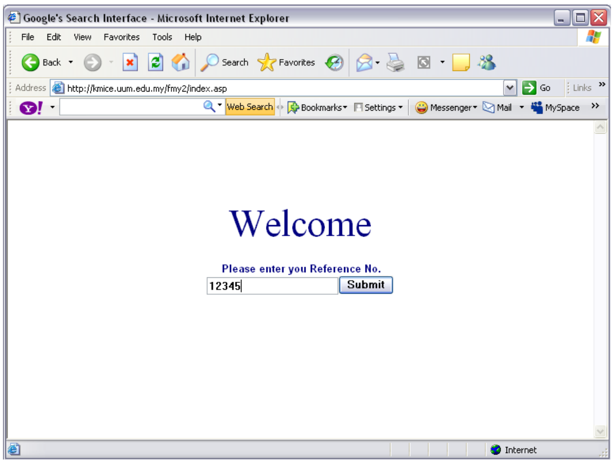


Figure 3. A Model of Search Interface

The search interface consists of two parts namely; reference number page (Figure 4) and the searching interface (Figure 5). Reference number page is an interface that accepts user's reference number. In this study, the matrix number was used as the unique reference number. The unique reference was used to group and index the user's queries information.



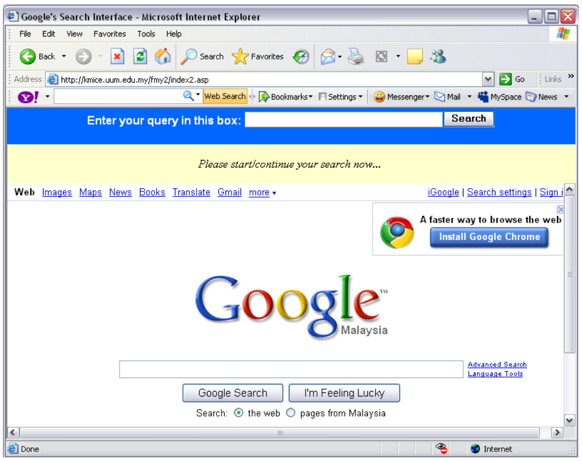
Welcome

Please enter you Reference No.

12345 Submit

Figure 4. Search Interface - Reference Number Page

The searching interface (Figure 5) will receives user's query, 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 6 shows example of the search session.

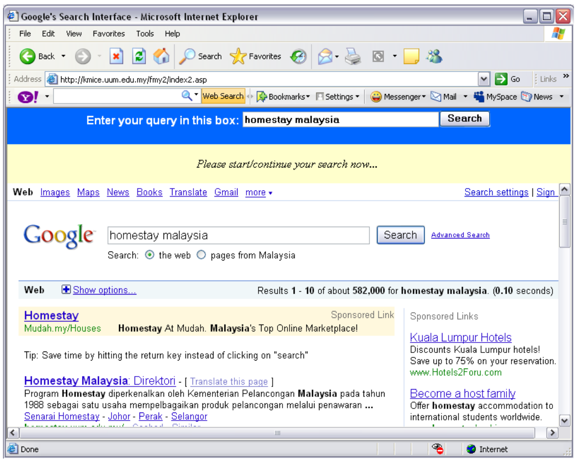

Google's Search Interface - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Search Favorites
Address http://kmice.uum.edu.my/fmy2/index2.asp Go Links
Web Search Bookmarks Settings Messenger Mail MySpace News
Search
Web Images Maps News Books Translate Gmail more
iGoogle | Search settings | Sign i
A faster way to browse the web
Install Google Chrome
Google Malaysia
Advanced Search
Language Tools
Google Search
I'm Feeling Lucky
Done
Internet

Enter your query in this box:

Please start/continue your search now...

Search: ● the web ○ pages from Malaysia

Figure 5. Search Interface - Searching Section


Google's Search Interface - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Search Favorites
Address http://kmice.uum.edu.my/fmy2/index2.asp Go Links
Web Search Bookmarks Settings Messenger Mail MySpace News
Search
Web Images Maps News Books Translate Gmail more
Search settings | Sign
Google
Search
Advanced Search
Web
Show options...
Done
Internet

Enter your query in this box: homestay malaysia

Please start/continue your search now...

homestay malaysia

Search: ● the web ○ pages from Malaysia

Results **1** - **10** of about **582,000** for **homestay malaysia**. (**0.10** seconds)

Homestay — Sponsored Link
Mudah.my/Houses **Homestay** At Mudah. **Malaysia**'s Top Online Marketplace!

Tip: Save time by hitting the return key instead of clicking on "search"

Homestay Malaysia: Direktori - [Translate this page]
Program **Homestay** diperkenalkan oleh Kementerian Pelancongan **Malaysia** pada tahun 1988 sebagai satu usaha mempelbagaikan produk pelancongan melalui penawaran **...**
Senarai Homestay - Johor - Perak - Selangor

Sponsored Links

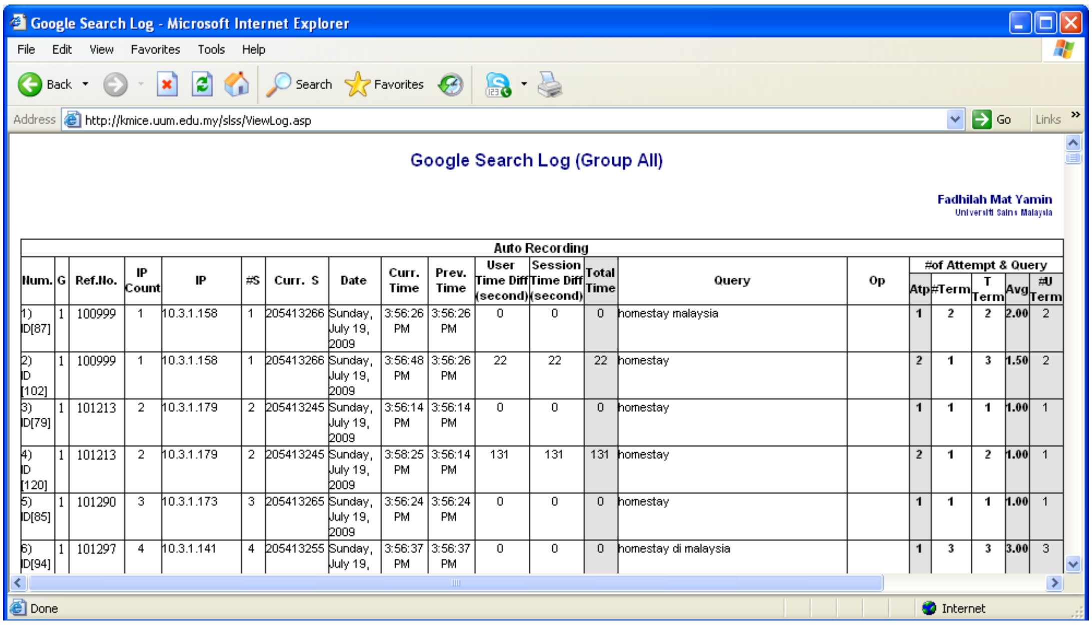
Kuala Lumpur Hotels
Discounts Kuala Lumpur hotels! Save up to 75% on your reservation.
www.Hotels2Foru.com

Become a host family
Offer **homestay** accommodation to international students worldwide.

Figure 6. Example of Search Session

Findings

Users' queries and other information from the searching session have been recorded in the searchlog. The log is a file where all activities on the web are recorded.Figure 7 shows the example of the search log that has been recorded. 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.



http://kmice.uum.edu.my/slsl/ViewLog.asp

Google Search Log (Group All)

Fadhilah Mat Yamin
Universiti Sains Malaysia

Auto Recording																			
															#of Attempt & Query				
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	Atp	#Term	T Term	Avg	#U Term
1) ID[87]	1	100999	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
2) ID [102]	1	100999	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
3) ID[79]	1	101213	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
4) ID [120]	1	101213	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
5) ID[85]	1	101290	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
6) ID[94]	1	101297	4	10.3.1.141	4	205413255	Sunday, July 19,	3:56:37 PM	3:56:37 PM	0	0	0	homestay di malaysia		1	3	3	3.00	3

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

Conclusion

The search interface proposed in this paper is to be used to study the searcher behaviour. In our previous study [22] this interface has been used to study users' behaviour on query formulation. The searcherswere asked to perform a search for a given topic. As the search log provides rich information on the search activity, the data provide the insight on how the searchers performing the search and the strategy used in order to achieve the search goal.

The use of search interface to study the search behaviour is an alternative approach as the search log is very expensive to be obtained. There might exist some delay between the actual search time and the time recorded at the server which might reflect the speed. However, this limitation can be ignored as the concern is not the speed of the search task but the overall time spends on searching.

References

[1] Z. Jin, and S. Fine, "The Effect of Human Behavior on the Design of an Information Retrieval System Interface", International Information & Library Review, vol. 28, pp: 249-260, 1996.

[2] P. Wang, W. B. Hawk, and C. Tenopir, "Users' Interaction with World Wide Web Resources: An Exploratory Study Using a Holistic Approach", Information Processing and Management, pp. 229-251, 2000

[3] C.W. Choo, B. Detlor, and D. Turnbull, "Information Seeking on the Web: An Integrated Model of Browsing and Searching", First Monday, vol. 5(2), 2000

[4] A.D. Madden, B. Eaglestone, N.J. Ford, and M. Whittle, "Search Engines: A First Step to Finding Information: Preliminary Findings from a Study of Observed Searched", Information Research, vol.11(2), 2007

[5] B.J. Jansen, "Search Log Analysis: What is, what's been done, how to do it", Library & Information Science Research, vol. 28, pp. 407-432, 2006.

[6] G. Marchionini, Information Seeking in Electronic Environments, Cambridge University Press, 1995

[7] A. Sutcliffe, and M. Ennis, "Towards a Cognitive Theory of Information Retrieval", Interacting with Computers, vol. 10, pp. 321-351, 1998.

[8] M. Mat Hassan, and M.Levene, "Associating Search and Navigation Behavior Through Log Analysis", Journal of the American Society for Information Science and Technology, vol .59(9), pp. 913-934, 2005

[9] A. Broder, "A Taxonomy of Web Search", SIGIR Forum, vol. 36(2), pp. 3–10, 2002

[10] D. J. Bell, and I. Ruthven, I., "Searcher's Assessments of Task Complexity for Web Searching", Advances in Information Retrieval, Springer Berlin/Heidelberg, pp. 57-71, 2004

[11] B. Schneiderman, D. Byrd, and W. B. Croft, "Clarifying Search: A User Interface Framework for Text Searchers", D-Lib Megazine, 1997. Available at http://www.dlib.org/dlib/january97/retrieval/01shneiderman.html

[12] B. Allen, "Expressing Information Needs", Book Series: Library and Information Science, 96, ISSN: 1876-0562, Emerald Group Publishing Limited, pp: 126-151, 1996

[13] E. Barsky, and J. Bar-Ilan, "From the Search Problem through Query Formulation to Results on the Web", Online Information Review, vol. 29 (1), p.75, 2005

[14] J. Shapiro, and I. Taksa, "Constructing Web Search Queries from the User's Information Need Expressed in a Natural Language", Proceedings of the 2003 ACM symposium on Applied computing (SAC'03), 2003

[15] J. Huang, and E. N. Efthimiadis, "Analyzing and Evaluating Query Reformulation Strategies in Web Search Logs", Proceeding of the 18th ACM Conference on Information and Knowledge Management, Hong Kong, China, pp. 77-86, 2009.

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

[17] F. Mat-Yamin, and T. Ramayah, "Searching for Information on the Web: A Guideline for Effective Searching", In Alias, N.A, & Hashim, S., Instructional Technology Research, Design and Development: Lesson from the field, pp: 184-201, 2012

[18] R. Nachmias, and A. Gilad, "Needle in a Hyper stack: Searching Information on the World Wide Web", Journal of Research on Technology in Education, vol. 34(4), pp. 475-486, 2002

[19] S. Y. Rieh, "Judgement of Information Quality and Cognitive Authority in the Web", Journal of the American Society for Information Science and Technology, vol. 53(2), pp. 145-161, 2002.

[20] M. Nazim, "Information Searching Behaviour in the Internet Age: A Users' Study of Aligarh Muslim University", The International Information & Library Review, vol. 40(1), pp. 73-81, 2008

[21] S. Brin, and L. Page, "The Anatomy of a Large Scale Hyper Textual Web Search Engine", WWW7 / Computer Networks, vol. 30(1-7), pp. 107-117, 1998.

[22] F. Mat-Yamin, and T. Ramayah, "User Web Search Behaviour on Query Formulation", Proceedings of 2011 International Conference on Semantic Technology and Information Retrieval, pp. 182-188, 2011