

CAPTURING USER BROWSING BEHAVIOUR INDICATORS

Deepika¹, Dr Ashutosh Dixit²

^{1,2}Computer Engineering Department, YMCAUST, Faridabad, India

ABSTRACT

The information available on internet is in unsystematic manner. With the help of available browsers, user can get their data but that too are not relevant. To get relevant results, users' interest should be considered. But no available browser is considering users' interest in browsing the internet. There are some indicators that are used to indicate the users' browsing behaviour over the internet. These indicators are called as implicit indicators and explicit indicators. In this paper, a tool is proposed that will store the users' browsing behaviour indicators. These indicators may be used in future to analyze the web page visited by user.

KEYWORDS

Implicit indicators, explicit indicators, browsing behaviour, browser

1. INTRODUCTION

With the rapid development in technology in past years, use of internet is widely increased. It is becoming the fastest and best medium to access the information. The size of information is also increasing tremendously. But due to its large size, it is available in disorganized manner. User uses the search engines like Google, Mozilla etc to search their data over internet but the results available are not accurate according to users' desire. General working of a normal search engine is shown below in figure 1:

The architecture of general search engine constitutes of many parts as follows:

a. Search Engine

It acts as an interface between user and the repository where pages are stored. It will provide user an interface like Internet Explorer, Firefox, and Google Chrome etc where user can submit their query and get the results. In general search engine get data from ranker.

b. Crawler

It is a program which is responsible for downloading data from World Wide Web. It gets the URL from URL Queue and downloads the pages accordingly. The downloaded pages are stored in the repository. Many design issues are related with this module [20].

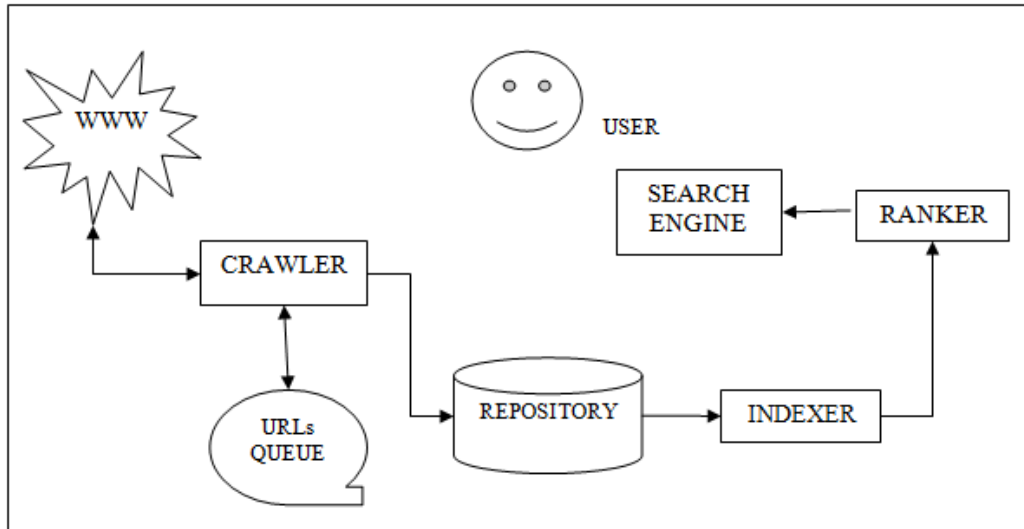


Fig 1 Architecture of Search Engine

c. Ranker

By using some ranking algorithm, it will rank the pages provided by indexer. There are many ranking algorithms like Page Rank, HITS, Link based and many more[21][22][23].

d. Indexer

It will get the pages from repository and index them. Indexing can be done using various techniques [24][25].

Although multiple user fires same query but their preferences are different from each other. For example if a search engine shows 10 web pages in response to user's query so it may be the case that some likes third and fourth while others first and second pages. To get relevant results both in the sense of query and users' interest, search engines has to modify their searching pattern by including users' interest. To get users' interest, their interaction with web pages should be recorded. Whenever user visits web page they perform some operations like clicking, scrolling, printing, saving, selecting etc. these operations are referred as users' behavior pattern. By using these patterns and user information, page can be evaluated and hence can be placed at more prefer location.

So, the area where improvement is required is getting users' browsing behavior pattern. None of the available browser is considering users' interest in searching their information. They simply get the query from user and show the results from its database. They may be asking explicitly to rate the page for their future recommendation. But in such a busy life many of the times users' doesn't rate the page and just skip that part. So some other methods should be developed that measures users' interest and use it for future recommendation in displaying the result. The implicit and explicit indicators [1] are used to describe users' browsing behaviour.

Explicit indicators are those indicators in which there is direct involvement of user in deciding his interest like by rating a particular page. Many researchers consider explicit indicator alone as criteria for deciding users' interest. But researchers found that explicit indicators always not telling the right information about that particular page. It may be the case that due to lack of time or just

for fun user may not rate the page in a right manner. So, alone explicit indicators should not be the criteria. Implicit indicators should be involved. These are those indicators in which there is no direct involvement of user but in background his or her behaviour can be used for indicating his or her interest on web page. Much work has been done on this but none of them include all implicit indicators. In this proposed work, almost all are implicit indicators are trying to store. Here, a browser is proposed that will help in storing the users' browsing behaviour. A database is maintained at backend for storing these behaviour indicators.

2. RELATED WORK

Many researchers have done research on implicit and explicit indicators [1].

Claypool et al. 2000 [2], Morita et al., 1994 [3] and GroupLens System (Sarwar et al., 1998) [4] found that the explicit rating given by user was not accurate all the time. It was different from the normal patterns of users' browsing behaviour. User may feel difficulty in rating the page on the available criteria.

Palme 1997 [5] concluded that there may be case that the user who is evaluating the web page by rating is biased to do so. So, it will affect the actual result.

Malone et al. (1997) [6] used implicit indicators instead of explicit for analyzing users' browsing behaviour. The main concern is the reduction in cost for examining and evaluating the behaviour by using implicit indicators.

GroupLens developed by Resnick et al. 1994 [7] showed that the implicit indicator time spent on reading is nearly as accurate as explicit indicator numerical ratings. They also suggested further actions, such as printing, saving, forwarding, replying to, and posting a follow up message to an article, as sources for implicit ratings.

Stevens et al., 1992 [8] developed a system called InfoScope. It was used for automatic profile learning. InfoScope used three sources of implicit evidence about the user's interest in each message: whether the message was read or ignored, whether it was saved or deleted, and whether or not a follow up message was posted.

Powerize Server™ (Jinmook et al., 2001, Oard et al., 1998) [9] is Windows NT Web server-based text retrieval and filtering system. It was then modified to measure some implicit and explicit behaviour. It measures the reading time and printing behavior for each web page. It also considers explicit rating given by each user for each web page.

W.Y. Arms [10] suggests constructing user profile by analyzing user history by using user behavior forest in digital library environment: User behavior history has been employed to discover user's preferences and interests for personalized systems. Mostly a framework for extracting user interested items by analyzing user behavior history in digital library environment. Digital libraries (DL for short) have become indispensable tools for today's knowledge professional. User profile, also called user model, is the essential component of a personalization system to enable the adaptation effect

Granka et al[13] measured eye-tracking to determine how the displayed web pages are actually viewed. Their experimental environment was restricted to a search results.

Goecks and Shavlik [14] proposed an approach for an intelligent web browser that is able to learn a user's interest without the need for explicitly rating pages. They measured mouse movement and scrolling activity in addition to user browsing activity (e.g., navigation history).

In other research done by Xiaomin [15] obtained the following conclusions:

- The user's browsing behavior can be classified into three categories, namely physical behavior (eye rotation, heart rate changes etc.), significant behavior (save the page, print page and other acts) and indirect behaviors (browsing time, mouse keyboard operation etc.).
- Indirect behaviors are the main source to estimate user interest rate. Significant behaviors act in the event that the user is high degree of interest in the corresponding page, but fewer significantly behaviors happens, a large number of pages have no corresponding significant acts, as a result the significant behaviors only play a supporting role in the estimation of user interest.
- With the analysis of user's indirect behaviors, the smallest combination of browsing behaviors draws as follows: save a page, print a page, store a page in the Bookmark, the number of times to visit the same page, dwell time on a page.

User behaviors can be divided into the following categories [16]:

- 1) Marking behavior: saving, printing, etc.
- 2) Operating behavior: copying, scrolling, clicking, etc.
- 3) Repetitive behavior: reading the same document repeatedly.

Many studies have found that the precision of user interests extracting without considering user behaviors is inaccurate. [17] Proposed that user behaviors including residence time, visiting frequency, saving, editing could reveal user interests. [18] presented that average reading speed play a key role in determining the grade of user interests. [19] Found that marking tags could reflect user interests.

3. PROBLEM IDENTIFICATION

From the above literature review it has been observed that existing systems has following drawbacks:

- By considering only explicit rating is alter the normal behaviour of user. They may be biased or not interested in rating the web page
- By analyzing users' behaviour from users' profile is not giving accurate results.
- By considering time spent on page as users' interest will not always be true. It will not always true that long time spend on page means his high interest on that page.
- Only few indicators are considering for indicating users' behaviour.

4. PROPOSED WORK

When user wants to search their information on the net, he enters the keyword in search area and gets the result in the form of urls. He will get all the urls that contains even a single word from users' search query. It will display the results on the basis of keyword matching. The existing browsers do not consider users' interest while displaying result. In this proposed work, a tool is generated that will capture users' actions on the net. The actions are categorized into implicit and explicit indicators. These indicators may be helpful in future in considering users' interest in particular web page.

In this work, a tracking tool is developed which is called as Web Browser. This Proposed Web Browser will monitor users' actions on particular web page for that particular search session. To implement browser, integrated development environment is required. JxBrowser is a Java library that allows web browser properties into Java SWING/AWT. It is highly helpful in creating user interface friendlier.

4.1 Web Browser Interface

The GUI of proposed Web Browser is simple. It has limited features that are required for tracking users' actions. Its interface is as similar to existing browser and also it will open the page as similar to any browser by typing any URL in address bar and click on go button. It is different from the existing ones when the page is closed by clicking on close button. It will ask to rate the page before closing by prompting a window. It also stores the mouse and keyboard activities in database. It will store in the database and will be used for further analysis. Figure 2 show the GUI of proposed web browser.

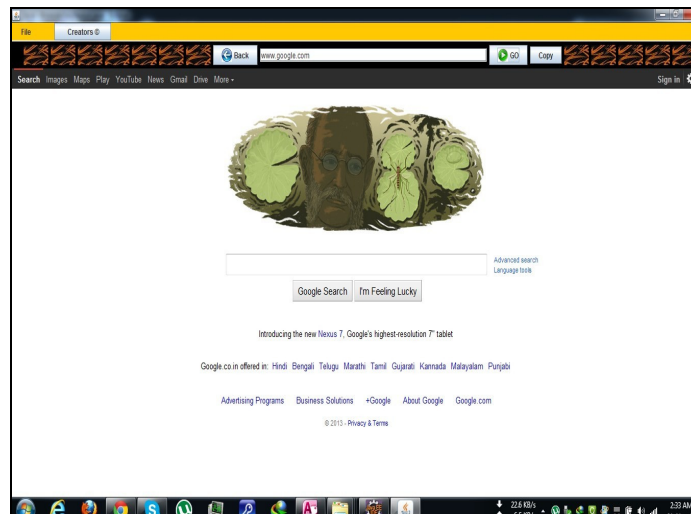


Fig. 2. Web Browser Interface.

4.2 User Actions on Web Page

Many actions are done by user while browsing a web page. These actions are called as implicit indicators. In this paper all possible actions are trying to capture and stored in database. Following are the actions which are used by user and are shown in browser window in Figure 3:

- i) Mouse Click:- count the number of clicks done by mouse on a particular page
- ii) Hyperlink Click:- note down whether hyperlinks are clicked on a particular web page
- iii) Time spent:- how much time user spend on a particular web page
- iv) Print:- note down whether user take the print of that particular web page
- v) Save:- note down whether user save that particular web page
- vi) URL:- note down the URL entered by corresponding user for analyzing
- vii) Date of visit
- viii) Scroll Bar activities:- count the number of times scroll move up & down
- ix) Number of key up and down:- count the number of keys up n down
- x) Number of scrollbar clicks:- count scroll movement
- xi) Back:- number of times back button click

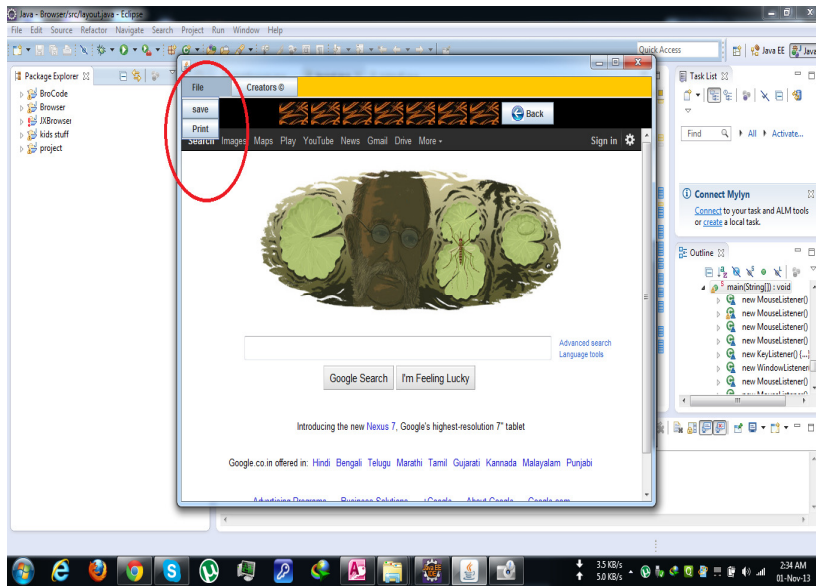


Fig 3 Browsing Indicators

All these details then store in MS-Access database. For data collection, this proposed browser is installed on multiple computer systems. Then different users will use this browser and their actions will be collected automatically. To store all above stated activities MS- Access is used as database. All the indicators are column headings and id numbers corresponding to each URL are row headings. Given below in figure 4 is snapshot of database file in MS- Access.

URL Visited	Clicks	Copy Clicks	Session Time	Date of Visit	Save Button	Print Button	Data Fetched(in KB)
[www.google.com]	1	4	24	Fri Nov 01 02:10:46 IST 2013	1	1	14229
[www.gmail.com]	1	4	24	Fri Nov 01 02:10:46 IST 2013	1	1	14229
[www.google.com]	1	4	24	Fri Nov 01 02:10:46 IST 2013	1	1	14229
[www.yahoo.com]	1	4	24	Fri Nov 01 02:10:46 IST 2013	1	1	14229
[www.google.com]	1	4	24	Fri Nov 01 02:10:46 IST 2013	1	1	14229
[www.gmail.com]	1	4	24	Fri Nov 01 02:10:46 IST 2013	1	1	14229
[www.facebook.com]	0	0	46	Fri Nov 01 02:12:08 IST 2013	0	0	-1
[www.google.com]	1	4	24	Fri Nov 01 02:10:46 IST 2013	1	1	14229
[www.facebook.com, www.google.com]	0	0	20	Fri Nov 01 02:13:07 IST 2013	0	0	14228
[www.facebook.com, www.ntel.litm.ac.in]	3	5	60	Fri Nov 01 02:13:07 IST 2013	1	1	2355
[www.facebook.com, www.google.com]	5	8	70	Fri Nov 01 02:13:07 IST 2013	5	5	156998
[www.facebook.com, www.google.com]	6	9	20	Fri Nov 01 02:13:07 IST 2013	2	0	35450
[www.facebook.com, www.google.com]	9	4	10	Fri Nov 01 02:13:07 IST 2013	0	0	1540
[www.gmail.com]	6	4	25	Fri Nov 01 02:13:07 IST 2013	1	2	5610
[www.rediffmail.com]	8	2	10	Fri Nov 01 02:13:07 IST 2013	0	2	4712
[www.gmail.com]	2	0	60	Fri Nov 01 02:13:07 IST 2013	5	4	2450
[www.facebook.com]	5	0	90	Fri Nov 01 02:13:07 IST 2013	7	9	452
[www.gmail.com]	7	1	112	Fri Nov 01 02:13:07 IST 2013	5	8	7450
[www.ntel.com]	4	2	7	Fri Nov 01 02:13:07 IST 2013	2	7	245
[www.gmail.com]	3	0	2	Fri Nov 01 02:13:07 IST 2013	5	2	4520
[www.facebook.com, www.google.com]	0	0	20	Fri Nov 01 02:13:07 IST 2013	0	0	14228
[www.ntel.com]	4	2	7	Fri Nov 01 02:13:07 IST 2013	2	7	245
[www.gmail.com]	1	4	24	Fri Nov 01 02:10:46 IST 2013	1	1	14229
[www.facebook.com]	0	0	46	Fri Nov 01 02:12:08 IST 2013	0	0	-1
[www.google.com]	1	4	24	Fri Nov 01 02:10:46 IST 2013	1	1	14229

Fig.4. MS-Access Database

5. CONCLUSION AND FUTURE WORK

The goal of this paper is to gather user browsing behaviour actions. For this a tool in the form of web browser is proposed that will do this job. It will store all the user action which he/she is doing while browsing a web page. In MS-Access this information is stored and may be used for further analysis.

The data captured by this proposed tool may be used to get more relevant results for the user on his submitted query. By observing the behavior with the help of this proposed tool will contribute in analyzing the user behavior on web. By analyzing it can be possible to predict its interest on particular web page and will use in future for showing its interest pages only Another benefit can be obtained from this stored information is to analyze whether is there an association between explicit ratings of user satisfaction and implicit measures of user interest. Which implicit indicators are strongly associated with user interest can also be calculated.

REFERENCES

- [1] H. Kim, K. Chan, Implicit Indicators for Interesting Web Pages, Web Information System and Technologies(WEBIST2005), pp. 270-277,2005.
- [2] M. Claypool., and L P Waseda, M., et al, "Implicit interest indicators," In: Campbell, M., ed. Proceedings of the ACM Intelligent User Interfaces Conference (IUI). New York: ACM Press, 2001, pp. 14-17.
- [3] Morita, M. and Shinoda, Y. (1994). Information Filtering Based on User Behavior Analysis and Best Match Text Retrieval. *Proceedings of the seventeenth annual international ACM-SIGIR conference on Research and development in information retrieval*. Dublin, Ireland: pp. 272-281.
- [4] B. Sarwar, J.Konstan, A. Borchers, J. Herlocker, B. Miller, and J.Riedl. Using Filtering Agents to Improve Prediction Quality in the GroupLens Research Collaborative Filtering System. In *Proceeding of the ACM Conference on Computer Supported Cooperative Work (CSCW)*, 1998
- [5] Palme, J. (1997), Choices in the Implementation of Rating, in Alton-Scheidl, R., Schumutzer, R., Sint, P.P. and Tscherteu, G. (Eds.), *Voting, Rating, Annotation: Web4Groups and other projects; approaches and first experiences*, Vienna, Austria: Oldenbourg, 147-62
- [6] Malone, T.W., Grant, K.R., Turbak, F.A., Brobst, S.A. and Cohen, L.R. and Riedl, J. (1997), Applying collaborative filtering to usernet news, *Communications of the ACM* 30(5), 390-402.
- [7] Paul Resnick, Neophytos Iacovou, Mitesh Suchak, Peter Bergstrom, and John Riedl. GroupLens: *An open architecture for collaborative filtering of netnews*. In Richard K. Faruta and Christine M. Neuwirth, editors, Proceedings of the Conference on Computer Supported Cooperative Work, pages 175-186. ACM, October 1994. <http://www.cs.umn.edu/Research/GroupLens/cscwpaper/paper.html>.
- [8] Curt Stevens. *Knowledge-Based Assistance for Accessing Large, Poorly Structured Information Spaces*. PhD thesis, University of Colorado, Department of Computer Science, Boulder, 1992.
- [9] Jinmook Kim, Douglas W. Oard, and Kathleen Romanik. *Using Implicit Feedback for User Modeling in Internet and Intranet Searching*. College of Library and Information Services University of Maryland May, 2001.
- [10] W.Y. Arms, *Digital Libraries*. MIT Press, Cambridge,2001
- [11] IFLA(1998)Metadata Resources, <http://www.nlc-bnc.ca/ifla/II/metadata.htm>
- [12] Konstan, J., Miller, B., Maltz, D., Herlocker, J., Gordon, L., and Riedl, J. (1997).GroupLens: Applying Collaborative Filtering to Usenet News. *Communications of the ACM*. Vol. 40, No. 3, pp. 77-87. Internet Site.
- [13] Granka, L. A., Joachims, T., Gay, G., 2004. Eye-tracking analysis of user behavior in WWW search. In Proc. 27th annual international conference on Research and development in information retrieval.
- [14] Goecks, J. and Shavlik, J., 2000. Learning users' interests by unobtrusively observing their normal behavior. In Proc. 5th international conference on Intelligent user interfaces, 129-132.
- [15] Ying X. The Research on User Modeling for Internet Personalized Services. National University of Defense Technology,2003: 55-57.

- [16] S. Tieli, and Y. Fengqin, "An approach of building and updating user interest profile according to the implicit feedback," Journal of Northeast Normal University (Natural Science Edition), 2003, 35 (3) pp. 99-104.
- [17] T. qiong, and L. Xiaoli, "A Method of Providing Individual Service in Search Engine," Computer Science 2002, 29-1.
- [18] T. P Liang, and H J Lai, "Discovering User Interests from Web Browsing Behavior : An Application to Internet News Services," Proc of the 35th Hawaii Int Conf on System Sciences 2002.
- [19] H. Lieberman, "Letizia, an agent that assists web browsing," In:Burke, R., ed. Proceedings of the International Joint Conference on Artificial Intelligence. Menlo Park, CA: AAAI Press, 1995, pp. 924-929.
- [20] Deepika, Dixit Ashutosh, "Web Crawler Design Issues: A Review", published in International Journals of Multidisciplinary research Academy (IJMRA) in August 2012.
- [21] L. Page, S. Brin, R. Motwani, and T. Wino grad, "The PageRank Citation Ranking: Bringing order to the Web". Technical report, Stanford Digital Libraries SIDL-WP- 1999-0120, 1999.
- [22] C. Ding, X. He, P. Husbands, H. Zha, and H. Simon, "Link Analysis: Hubs and Authorities on the World". Technical Report: 47847, 2001.
- [23] Abou-Assaleh T., Das T., Weizheng G., Yingbo M., O'Brien P., Zhen Z., "A Link -Based Ranking Scheme For Focused Search".In:WWW2003, ACM Press.2007.
- [24] Changshang Zhou, Wei Ding, NaYang, "Double Indexing Mechanism of Search Engine based on Campus Net", Proceedings 2006 IEEE Asia-Pacific Conference on Services Computing (APSCC'06).
- [25] Murat Kalender, Jiangbo Dang, Suzan Uskudarli "Semantic TagPrint - Tagging and Indexing Content for Semantic Search and Content Management", 4th IEEE International Conference on Semantic Computing, 2010.

AUTHORS

DEEPIKA PUNJ

Deepika Punj is working as assistant professor in Department of computer engineering at YMCA University of Science and Technology, Faridabad, India. She is having total 10 years of experience in teaching. Currently, she is doing research in the area of Internet technologies. She has authored papers in national and international journals



ASHUTOSH DIXIT

Ashutosh Dixit received his PhD and M. Tech. in Computer Engineering from MD University Rohtak, India in the years 2010 and 2004 respectively. He is presently serving as Associate Professor in the department of computer engineering at YMCA University of Science & Technology, Faridabad India. He has published around 70 research papers in various International journals and conferences. His research interests include Internet Technologies, Data Structures and Mobile and Wireless networks.

