

MULTIPLE COMPONENT BASED INFORMATION TRACKING SYSTEM

Mathi Seelan.V¹ and Dr.Sunitha Abburu²

¹Adhiyamaan College of Engineering, Department of Computer Application, Hosur
mathiseelan88@gmail.com

²Professor and Director, Adhiyamaan College of Engineering, Department of Computer Application, Hosur
ausunithaa@yahoo.co.in

ABSTRACT

Tracking Browsing history in internet is essential because web contents changes dynamically and web users would like to re-visit the web pages which they have visited in the past. The paper proposes and builds a System which integrates popular search engines and browsing history Information Tracking System (I-TS, Information Track System) in a single website. That tracks and maintains the browsing history for various components, as per users needs. I-TS consists of three main components, named as Search System, Keyword Summary and Item Viewed Summary. Search Area System as a Meta-search engine will direct to a commercial search engine, get the hits, do further analysis and derive a number of most relevant domain sites. Keyword Summary will extract the keyword, count, item (web, image, video, news) and date time. Item Name Summary will first extract the URL with item, count and item name. The Proposed System is implemented and Results are shown below.

KEYWORDS

Tracking, Browsing History, World Wide Web, Search engine

1. INTRODUCTION

The rapid growth of information technology, repositories, search engines and the extreme augmentation of the World Wide Web (WWW) over the past decade have significantly changed the Web browsers computing environment. Brisk development of information technology has resulted in dramatic changes in World Wide Web. The World Wide Web has become one of our primary means of information and communication. Web sites have raised the use of internet, people can use the World-Wide-Web significantly easier and faster. The World Wide Web is one among the most enveloping inventions of all time. Research, education, businesses, families and entertainment become increasingly intertwined with WWW. The volume of digital information has grown tremendously in recent years, due to low cost digital equipments, scanners, and storage and transmission devices. Due to the low cost of devices and the advancements with www, most of the information is present in online. This includes: text information, images, videos, music files etc. Therefore, the Web has become the most effective and powerful information source for individuals of any age and profession.

The WWW is playing a vital role in the information technology [1], unfortunately, it is difficult for a web user to locate the appropriate information that the user finds relevant, as different users need different data in different contexts. To regulate the effectiveness of the browsing and to decrease the time researchers have paid lot of attention on the web browsers actions. Cockburn et al.,[2] provides an empirical characterization of user actions at the web browser. Studies have suggested that approximately 81% of the pages viewed by Web users are ones they have previously visited. Maintaining and viewing Web browsing histories has been an

important part of browsers. While many of these page views are of well-known or bookmarked sites, there is a clear need for people to be able to easily re-find a page they have previously viewed. Orland Hoerber et al., [3] support this activity in a visual manner that introduces BrosweLine as a method for supporting users in the task of re-finding Web pages in their browsing histories. This increases the effectiveness of the browsing by quickly locating the previously visited sites.

Wide range of search engines available in the market does not support tracking of browsing history of web browsers in the internet. Web contents are changing dynamically, web user would like to Re-visit previously viewed Web pages or previously used key words according to their interest. However, due to the changes happen dynamically in web content, re-visiting previously viewed Web pages becomes time consuming. This raises the need for tracking the browser history of web users.

Tracking the browser history of web users is essential to maintain discipline. In order to improve the effectiveness of browsing and reducing the total time, it is essential to create logs and reports of all Web sites visited previously, individual Web pages visited by the computer user, individual key words used, images and videos viewed previously. This raises the need for a tracking system with a logging mechanism maintains a history of individual Web pages visited by the computer user, individual key words used, images and videos viewed previously with any Web browser of any version. And generates a final comprehensive report on users browsing actions according to users area of interests.

Companies need to monitor their employees browsing history. Having the ability to record visited web sites helps to improve employees' productivity. Parents needs to monitor their children's Web browsing activities to guide the children in the right path and colleges need to track the browsing history of student in internet and blocking the student to view unauthorized sites.

Existing Proxy server providing tracking system that tracks the web user browsing history in internet and blocking the unauthorized sites, using internet protocol address, and it requires extra cost to install. This raises the need of an intermediate system which can track and generate report browsing history according to user's area of interests. In this paper we propose and build system that keeps track of the information of the users browsing history (I-TS, Information Track System). That maintains a history of individual Web pages visited by the computer user, news, individual key words used, images and videos viewed previously with any Web browser of any version. And generates a final comprehensive report on users browsing actions according to users area of interests.

Browser history Information Tracking System (I-TS) features are:

- Multiple component based information tracking system captures information about:
 - URL addresses, page titles, number of time accessed, time of last access.
 - List of keywords, number of time accessed, time of last access.
 - Images viewed and number of time accessed, time of last access.
 - Videos viewed and number of time accessed, time of last access.
 - Book marks.
 - News items.
- Generate a comprehensive report on multiple components usage.
- Calendar.

The rest of the paper is organized as follows. The literature survey report is in section 2. Section 3 discusses proposed method for Information Tracking System. In section 4, we present a practical implementation and experimental results. Finally, we conclude with a summary in section 5.

2. LITERATURE SURVEY

Due to low cost storage devices and the advancements in the World Wide Web gives the prospect to Digital Libraries. Digital Libraries have become essential in today's life. Different age groups and the people from various back grounds are highly relying on the World Wide Web and Digital Libraries. With the rapid growth of scientific literature, scientists, entrepreneurs and students are more and more incapable of efficiently finding relevant literature. To address this issue, personalization of the browsing history becomes a popular remedy. The browse history of the user helps them search their information need easily and revisit the sites they have visited in the past. Multiple commercial tracking tools available online. Web wide web is the primary means for getting the information from the digital library. The search engines retrieve the same set of results for a query irrespective of the user and the context. Generally, each user has different information needs for his query. Therefore, the search results should be adapted to users with different information needs. Kazunari et al., [4] propose several approaches to adapting search results according to each user's need for relevant information without any user effort. When a user submits a query to a search engine through a Web browser, the search engine returns search results corresponding to the query. Based on the search results, the user may select a Web page in an attempt to satisfy his/her information need. In addition, the user may access more Web pages by following the hyperlinks on his/her selected Web page and continue to browse. The system monitors the user's browsing history [5] and updates his/her profile whenever his/her browsing page changes. When the user submits a query the next time, the search results adapt based on his/her user profile. Development of personalized information retrieval service is subject of studies[6],[7]. Web page prediction is done by predicting the user's behaviour from previous web browsing history. Those predictions are afterwards used to simplify the user's future interactions [8]. A page interest estimation method based on analyzing the users' browsing behaviors is designed in [9] that mainly considers the reference length, the size, the visiting time and the visiting times of each accessed page is then used to determine the page interest. Chen Yu et al., [10] proposes a framework for extracting user interested items by analyzing user behavior history in digital library environment. The framework models every user's behaviors as a User Behavior Forest (UBF). Then user behavior log is formatted with several additional attributes. Prosunjit et al.,[11] proposes a new approach named Pagemap, which can guide a user to the most popular and relevant pages of that website based on user's current location within that domain. With the assistance of the access log file, Pagemap uses the previous browsing history to find the most popular pages in a website. Pagemap uses the vector space model to find out the pages with maximum matching in compare to the current page of the user. Tests proved that such personalized approach can improve search results effectively, and have good adaptability. when they switch from one computer device to another, their productivity will depend on the speed they recover personal browsing data, such as bookmarks, forms, history, open tabs and windows, passwords, custom entries to the spelling checker, etc. In order to provide ubiquitous access to those navigation data , Leandro et al.,[12] proposes a service that stores it in the web for future retrieval, regardless of computer device or operating system.

When web users need the system to track a particular html page or web content on the Internet for them, then web user need to register the URL of the particular html page in internet. And upon any changes happened to the html page, and then the web user will be acknowledged through e-mail. Usually, this kind of tracking tool can retrieve every detail of changes in web content, but unfortunately, because of new technology advancement, any changes happened to

the web page would trigger the system to push web user with acknowledgement e-mail, in order to rectify this type of problem, some systems, i.e. web beholder allows user to set a trigger level they are interested. Here, it and only if the total score changes is greater than the trigger level; the system will be triggered to send acknowledgement e-mail to the web user. But yet now there is no appropriate solution trigger level can be defined accurately since there are many changes possible in html page like title, header, content character, colour, text style and etc with different score. So, the users might be fed with e-mail although the changes are not interested by the web user and vice versa. Web users can register html pages using tracking system according to their interest.

Besides tracking the URL registered by the web user, some systems also featured in tracking the new pages containing the input provided by web users. Informant is one of the "best search monitoring tools" in the market. At it will allow the web user to input keyword of interests and user can select any one of the commercial search engines for tracking purpose. Then, after certain time interval, with the aids of the particularly selected search engines, it will detect the new html pages related to the keyword and acknowledgement send to the web user. However, author found that the relevancy and the status of "new" of the results are not giving appropriate solution after trying it.

Thus far, we have gone through a multiple tracking tools and some of their drawbacks. In general, in specific html page tracking, web user will be acknowledged when the html page is modified. While tracking keyword, web user will be displayed with a chain of new html pages containing the keyword. However, the web user needs to verify through every html page in order to find out appropriate html page what are the main topics behind the changes. To conclude, correct page tracker only tells us some html pages have been changed or some html pages are new. At this point, we still lack of a tool that can track a particular area of user's interests, collect the changes in html page at a certain time interval, process and generate a summary of the most discussed Issue in the changes in web content to the web user from time to time.

3. INFORMATION TRACKING SYSTEM

Information Tracking System (I-TS) consists of three main components:

- Search System
- Keyword Summary
- Item Viewed Summary.

Search Area System is a meta-search engine. That will direct the user keyword to a commercial search engine, get the hits, do further analysis and derive a number of most relevant domain sites. Keyword Summary will track the entire list of keywords used in multiple components like web, images, news, videos, the number of times the each keyword is used by the user in various components, date and time of the last search of the keywords. Item Viewed Summary will first extract the entire newly added URL with item name or URL of the html pages and then count the number of times same URL or item name is viewed by the user, the mechanism is illustrated in the flow chart in Figure 1.

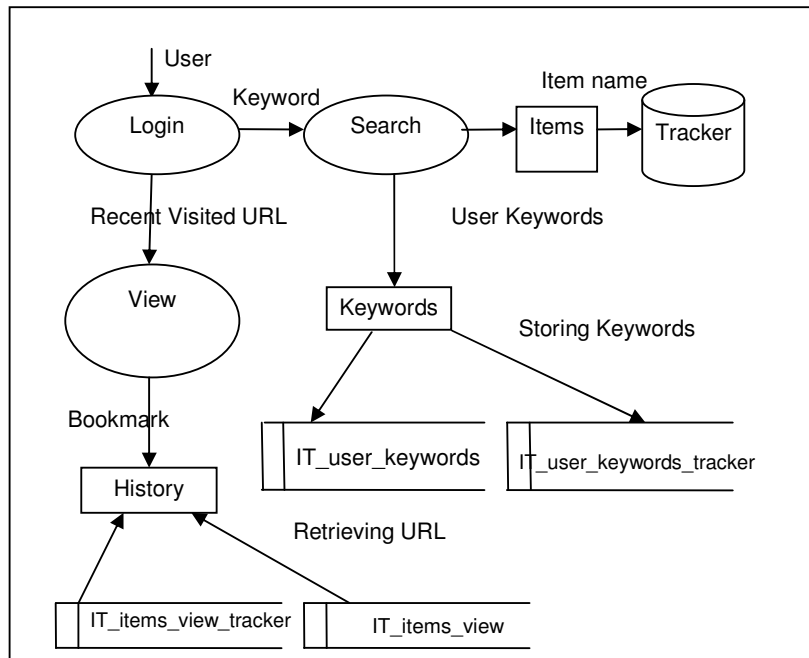


Figure 1. I-TS

3.1. SEARCH SYSTEM

Search Area System is a meta-search engine. Search Area System integrates popular search engines in a single website. The user given keyword will be sent as a parameter to a popular commercial search engine. The search engines draw the hits on the Internet that represent the particular input keyword from the web user, do further analysis and derive a number of most relevant domain sites. The retrieved sites are basically a group of domain sites most related to the keyword.

3.1.1. HTML PAGES RETRIEVAL APPROACH

The Search Area Meta search engine identifies the sites that are HTML pages in representing a particular keyword. At first, Search Area System will direct the keyword to any one of the popular commercial search engines and collect up to 100 hits. Each page of hits has a unique URL, name, path, and file name together. For example, the page <http://www.iyristech.com/KTS/index.html> as a domain URL of <http://www.iyristech.com/>, path of [KTS/](#) and a file name of [index.html](#). Now, from the 100 hits, Search Area System will further collect 10 HTML pages with their domain site URLs and most frequently occurred names. HTML page is the top page of a domain site if the domain has its all contents relevant to the keyword. But some of the domain sites have only the sub-directory according to the keyword. In this case, the salient page is the most visited page of the subdirectory.

3.2. KEYWORD SUMMARY

Keyword Summary is used to collect the keywords and item name searched by Search Area System. Keyword Summary stores keywords and item name in I-TS database. Keyword Summary displays the list of top visited keywords and item name to the web user. Keyword Summary consists of two components, they are Search System and View System. Search System is used to search web content according to components (web, images, videos, news,

etc..), keywords and item name searched by the web user are stored in database, web user can view the keywords and item name

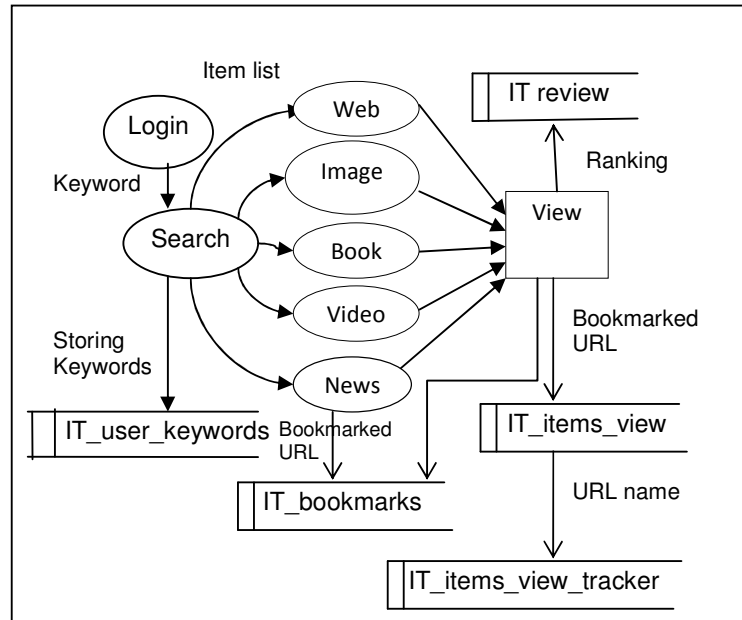


Figure 2. View System

that are retrieved from history table, from the View System, the mechanism is illustrated in the flow chart in Figure 2.

3.2.1. VIEW SYSTEM

View System is used to view top visited keywords and item name searched by the web user are stored in database. View System will retrieve keywords, item name from history table from the I-TS database. View System consist of number of times a particular key word is been used in the search, also retrieves the top visited keywords and top visited item name. Last In First Out (stack) technique is used for history of keywords list.

3.3. ITEM NAME SUMMARY

Item Name Summary is designed to generate a summary of URL, item name and count (number of times a particular URL is viewed by web user). It will display the recent browsing history of web user. Item Name Summary consists of two components named as Bookmark Folder, Calendar.

3.3.1. BOOKMARK FOLDER

Book Mark Folder is used to add and delete bookmarked links. User can add his own bookmark folder or user can use existing folder to store the bookmark link. Once the bookmark link is established, it can be viewed from anywhere in any system by the user. This provides valuable information the user as per his preferences and as per his requirements.

3.3.2. CALENDAR CONTROL

Calendar control is used to view browsing information of web user based on date, month, year and Calendar control is used to view recent browsing information of web user. The mechanism is illustrated in the flow chart in Figure 3.

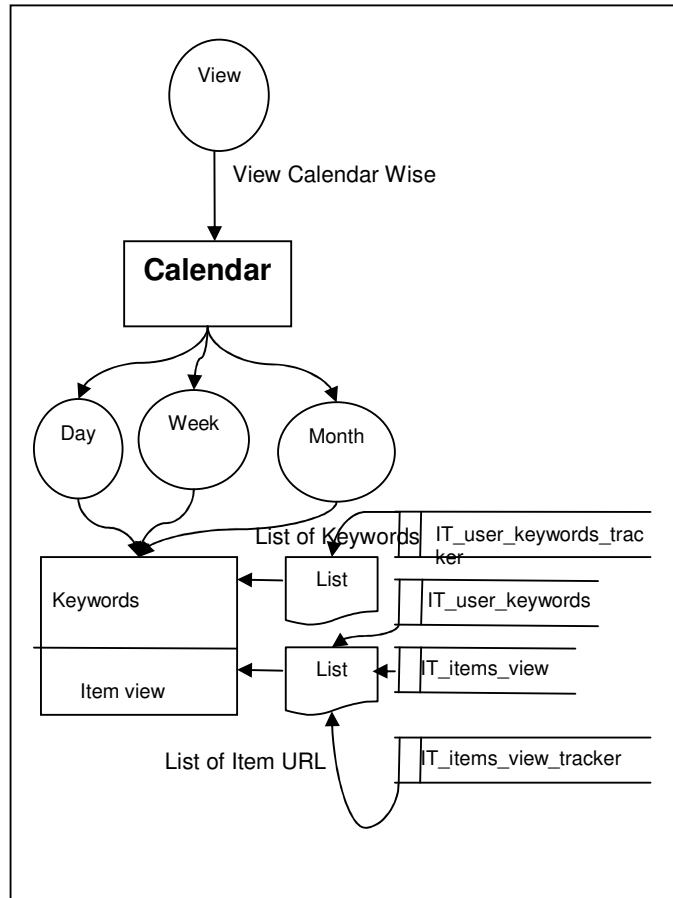


Figure 3. Calendar Control

used to get domain sites based on web user input (keyword) by directing keyword to search engine(Yahoo, Google, etc..) and track the browsing information of web user in internet. MY-SQL is used to store browsing history of web users in internet.

4.2. FIRST EXPERIMENTAL MODEL-KEYWORD BASED SEARCHING

The proposed Information Tracking System is implemented and tested for the features available. The first experimental model is based on the keyword based search, a keyword of "iris" was searched, see figure 4. There were 10 html pages retrieved by Search Area System with the help of one of the popular search engines (Google, Yahoo, etc..). As a result of the search the I-T System generates a comprehensive report on the browsing history. The report consists of various columns recording various browsing history details. First column in the report, is used to record the URLs of the html pages and the names of the respective domain sites are record in the second column, third column shows the count of how much time particular salient pages are viewed by particular user, while the forth column shows the date and time when the salient pages are viewed by the user. A comprehensive report is produced by

the I-TS is shown in the figure 5. That lists the information about various components like keywords, bookmarks, recent sites visited etc.

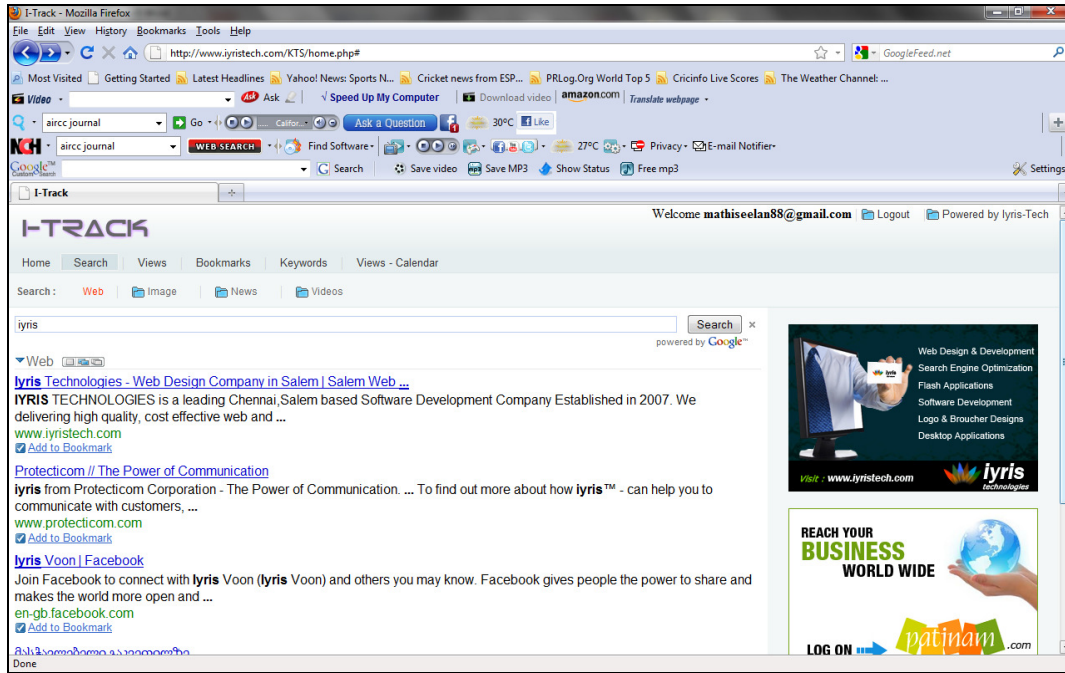


Figure 4. Showing Hits According To User Input

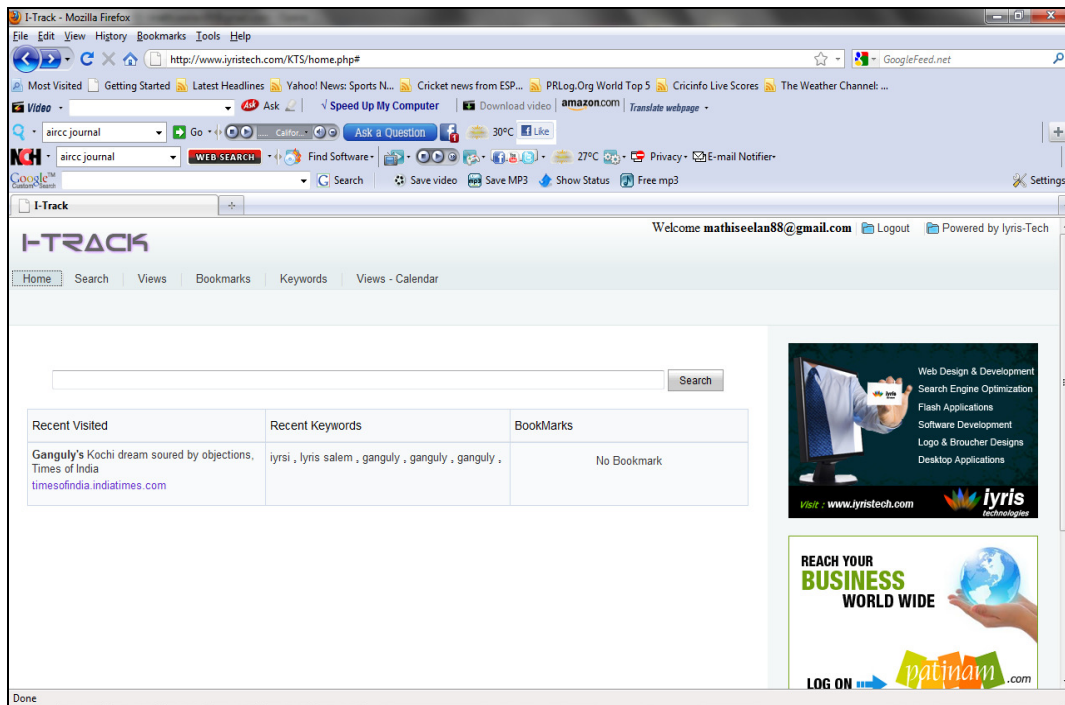


Figure 5. Displaying Recent Browsing Information of Web User

4.3. SECOND EXPERIMENTAL MODEL-ITEM URL BASED SEARCHING

The second experimental model is based on the Item URL based search, A keyword of "iyris Salem" was used. There were 20 html pages derived by Search Area System with the assist of the commercial search engine Google. As a result of the search the I-T System generates a comprehensive report on the browsing history. The report consists of various columns recording various browsing history details. In the first column shows the list of URLs of the html pages, and the names of the respective domains are recorded in the Second column. Html pages and the names of the respective domains are recorded in the second column. Third column shows the count of how much time particular html pages are viewed by particular user, while the forth column shows the date and time when the html pages are viewed by the web user, Figure 6 shows the Websites URL viewed by a web user. Figure 7 shows the keywords searched by web user. Figure 8 shows the calendar listing the sites visited with respect to day, month, and year.

itemname	item	count	datetime
<input type="text"/>			
Tamil language - Wikipedia, the free encyclopedia	web	1	2011-03-
SALEM Free Music, Tour Dates, Photos, Videos	web	1	2011-03-
Destination Salem	web	1	2011-03-
Salem Massachusetts - The Comprehensive Salem Guide	web	1	2011-03-
Salem, Massachusetts - Wikipedia, the free encyclopedia	web	1	2011-03-

Figure 6. List of Item URLs visited by the user in the past

keyword	item	count	datetime
<input type="text"/>			
trees	web	6	2011-03-10 11:55:43
tamil today	web	3	2011-03-10 11:06:01
cricket score	web	1	2011-03-09 17:13:00
gyyyu	web	1	2011-03-09 16:52:26
news india	web	1	2011-03-10 11:54:38
sports	web	1	2011-03-09 16:45:15
sports today	web	1	2011-03-10 11:53:52
Tamil nadu	web	1	2011-03-10 11:38:54
tamil nadu today	web	1	2011-03-10 11:54:15

Figure 7. List of Keyword used by the user in the past



Figure 8. Calendar

5. CONCLUSIONS

The objective of this work is to propose, design and implement browse history information tracking system I-TS. A unique interface application for accessing world's popular commercial search engines with user friendly environment where the user can record the browse history without any effort and time. The I-TS system allows the user to decide which information is to be maintained, information about various components like web, image, news, videos, and keywords. Due to the www (Internet) is open and dynamic. Web contents or any information is changing dynamically. If the user is willing to revisit the sites which he visited in the past, the browsing history along with the user preferences needs to be recorded. The I-TS propose to track the information area of user interest, and further summarize a recent browsing actions and preferences of user. For each user's keyword search, Search Area System will track and analysis's a group of web domains, in order to gain a set of web domains that can represent that information area in perfect. Search Area System calculates the count of domain sites user visited mostly. From the experiments done, we found that the approach adapted by Search Area System is appealing. Search Area System is excellent in tracking or retrieving information about browsing history on World Wide Web. By building I-TS and evaluating the performance issues, we conclude that this system is satisfying in fulfilling our research objectives.

REFERENCES

- [1] Venkat N Gudivada, Vijay V, "Raghavan. Information retrieval on the world wide web", IEEE Internet Computing, 1977(1) 5-68.
- [2] Cockburn and B. McKenzie, " What do Web users do? an empirical analysis of Web use" , International Journal of Human-Computer Studies, 54(6), 903-922, 2001.
- [3] Orland Hoeber, Joshua Gerner, "BrowseLine: 2D Timeline Visualization of Web Browsing Histories", Proceedings of the IV '09, 13th International Conference Information Visualisation, IEEE Computer Society Washington, DC, USA ,2009.

- [4] Kazunari Sugiyama, Kenji Hatano, Masatoshi Yoshikawa “ Adaptive Web Search Based on User Profile Constructed without Any Effort from Users” WWW2004, May 17–22, 2004, New York, USA, pp 675-684.
- [5] Andrea Bacic , Andrina Grani, “The Design of a Web Page Prediction Tool” , Proceedings of the ITI 2010 32nd Int. Conf. on Information Technology Interfaces, June 21-24, 2010, Cavtat, Croatia, pp 263-268.
- [6] Xiangli Huang, “Study on personalization of the search engine and its design “2nd International Conference on Information Science and Engineering (ICISE), 2010 , pp 1 – 4.
- [7] Micarelli, F. Gasparetti, F. Sciarrone, S. Gauch, “User profiles for personalized information access”, In P. Brusilovsky, A. Kobsa, W. Nejdl (eds.): The Adaptive Web: Methods and Strategies of Web Personalization, Lecture Notes in Computer Science, Vol. 4321. Springer-Verlag, Berlin Heidelberg New York, 2007, pp.54-89.
- [8] Andrea Bacic , Andrina Grani, “Intelligent Interaction: A Case Study of Web Page Prediction” , Proceedings of the ITI 2009 31st Int. Conf. on Information Technology Interfaces, June 22-25, 2009, Cavtat, Croatia, pp 287-292.
- [9] Yan Li, Bo-qin Feng, Feng Wang, "Page Interest Estimation Based on the User's Browsing Behavior," ICIC, vol. 1, pp.258-261, 2009 Second International Conference on Information and Computing Science, 2009, pp 258-261.
- [10] Chen Yu ,Yu Yang, Zhang Wei, Shen Junyi, “Analyzing User Behavior History for Constructing User Profile” , Proceedings of 2008 IEEE International Symposium on IT in Medicine and Education, ITME 2008, 12-14 Dec. 2008, pp 343 – 348.
- [11] Prosunjit Biswas, Sifatur Rahim, “Pagemap: A Dynamic User Guiding Approach to the Most Relevant and Popular Pages in a Website” , Proceedings of 13th International Conference on Computer and Information Technology (ICCIT 2010), 23-25 December, 2010, Dhaka, Bangladesh, pp 159-164.
- [12] Leandro G. de Carvalho, Raquel F. do Valle, Alexandre Passito, Edjard S. Mota, Edjair S. Mota, Raoni Novellino, Adriana G. Penaranda, “Synchronizing Web Browsing Data with Browser” , Proceedings of the the IEEE symposium on Computers and Communications, ISCC '10, IEEE Computer Society Washington, DC, USA, 2010, pp 738-743.