Redefining Web Revisitation Process using Content and Context Keywords

Bhagyashree[#1], Assoc.Prof.Jyoti Patil[*2]
[1]*P.G.Student, Computer Science and Engineering, PDA College of Engineering, Kalaburgi, Karnataka (India).*
[2] *Assoc.Professor, Computer Science and Engineering, PDA College of Engineering, Kalaburgi, Karnataka (India).*

Abstract

In real word, use of the websites has been common in real life system. The retrieval and recall of the used web pages are often carried out by the users on World Wide Web. The viewed web pages are continuously stored in the web sites cache or the server. The colossal data of the huge number of the users is a uneasy task to maintain the data. Proposed work will concentrate on the recollecting the data which has been used by the user in order to re access the web pages. The method web page reviewing is a concept used to recall the information while accessing websites and revisitation process can be made by the user's keywords of like. Tracking the web pages has been made easy by the proposed work as it will make the users to find target piece of information. Contrasted to the current web revisitation technique history list Seeking strategy, and search engine technique, the proposed technique conveys the best finding rate by analysing web pages revisited with `consideration of user experience. Among time, place, and events in this technique, event is the best review factor, and context and content based refinding conveys the best rendering contrasted with content based refinding and content based refinding.

Keywords

Web Revisitation, Impression Score, Context Score.

I.INTRODUCTION

Web is acting as a vital part in transforming data to clients' hands. A web page can be identified by a URL, and presents the web page contents as time-changing preview. In real life the use of network is like essential for any users for collecting, sharing, distributing the information. The accessibility of the information is made the user to get attracted to it as it is easy to access and store.

The URLs searching by keywords has been supported by manual web browsers. Among the regular web activities,

Web revisitation is visiting the already seen pages. Revisitation occurs when a user visits web page that has been already visited. One study showed that 58% of the web pages are revisited and another work showed that 81% of pages are already visited by users on the internet. This shows that revisitation is common. There are four categories of web revisitation, short term (revisitation within 1 hour), medium term (revisitation within 1 day), a week (revisitation within 1 week), and long term (revisitation after 1 week).

The users activities in the web pages are generally are;

- Viewing the contents
- Saving the contents
- Browsing for the data
- Saving marking URLS
- Tracking history created etc.

There are many techniques like history tool, bookmarks, search engine have been developed to reduce the time taken to access the web sites .But only few people uses history tools to revisit because it is hidden tool, it is very difficult to organize the bookmarks because of large amount of information, Search engine method yields irrelevant pages along with the relevant pages. These techniques do not reduce the physical overhead in accessing the web sites. To overcome these issues we have developed a technique which helps users to easily find their earlier retrieved websites.

In the proposed technique we are making use of context and content information, using these users can easily re-access the web pages and this information helps users to find their required web sites on the Internet. As the data is re accessed many times based on revisitation the web links has to be ranked based on multiple criteria and shown to the users. Feedback system also included in order to increase performance of the application.

II.RELATED WORK

In [1] author describes three interface methods for enhancing website page revisitation are depicted. The initial, a signal based system for issuing the regular Back and Forward buttons, addresses low-level interface issues, and is appeared to be both prominent and powerful. The second, a chronological

behaviour for the Back and Forward buttons, aims to beat the issues related with poor comprehension of the present behaviour of Back. In spite of the fact that the outcomes don't indisputably demonstrate points of interest for the temporal behaviour of Back, they firmly recommend that revisitation can be enhanced by giving transiently requested arrangements of already accessed pages. The third interface plot explores how upcoming browsers could incorporate the present tools for revisitation into a solitary utility, and how basic representation techniques can be utilized to help clients in recognizing target pages showed in smaller than normal .

In [2] author provides experimental facts that validate the desideratum intended for appropriate history implementation in browser applications. Utilizing the techniques and guidelines in this, designers can exhibit and improve present history implementation and explore incipient methodologies. But there are many questions with no answers; they require estimating the somatic and mental effort for recalling a specific habituated set of history list estimation. They require evaluating design guidelines that they have suggested. They withal required to assess in what way utilization arrangements modify along with upcoming browser interfaces (such as redesigned history mechanisms) and HTML additions.

In [3] author investigates client's revisitation configurations to web pages on the web, and exploits these to form an exact framework for the plan of history components in browser applications. Through history, a client can access rapidly to earlier seen web page, conceivably decreasing the subjective and somatic effort desired to explore to it without any preparation. They investigated a month and a half of use information gathered from 23 clients of a web browser. They found that 58% of a person's pages are returns to, and that clients persistently include new site pages into their collection of already seen web pages. Most of the web browsers support history feature, only few people use this history to find their previously visited web pages.

In [4] author incorporates a seek toolbar and a toolbar symbol to the web program chrome to build perceivablity of contextual web history(CWH). The seek toolbar gives the client a chance to get to CWH by giving an inquiry question. Clients could enter diverse kinds of data. All information was handled inside the single textbox with the goal that the client could rapidly get to the outcomes. This information demonstrated that including the hunt bar succeeded in expanding perceivability, with 75% of members utilizing this component to get to history. Afterward, in second model, they incorporated the inquiry seek tool with the Firefox seek bar situated on the upper right of the browser. Another probability is coordinate it with the URL locator bar, however they thought that it was hard to actualize and hence did not continue this alternative.

[5] this will additionally advantage for author to estimate in what way Memento accomplishes in combination with further revisitation techniques like particular match for browsing history While clients frequently return to web pages on the web, implement provision aimed at such re-appearance is quiet destitute. Present implements for histories of browsers simply afford clients with rudimental data, for example, the day of the last seen and titles of the web page accessed. The authors portray a framework that furnishes clients by means of terms description that avail rediscovery. Compared with earlier study, this framework considered both the context and content information to facilitate web revisitation.

In [6] author finds that as indicated by intellectual science study, human's recall nature is predicated on social events (e.g., room, music) around us. In the course of review tasks, they relate data/events/entities with contextual signs. On the other hand, PC frameworks don't use user characteristic procedure of utilizing logical prompts to encourage review. Authors came to know that contextual cue will not help users to enable recall for web revisitation. So they introduce a new technique called pivoting which helps users to find contextually interrelated actions and find a target portion of data. They also present another method called TimeMarks, which helps users to mark instant as important. Users can rapidly recognize websites by using this strategy as compared with history tools.

In [7] investigates rehash seek conduct through the examination of a one year web inquiry records of 114 unknown clients and a different controlled study of an extra 119 volunteers. Their investigation exhibits that upwards of 40% of all queries are revisitation inquiries. Refinding has all the earmarks of being a critical conduct for web crawlers to expressly support, and they investigate how this should be possible. They exhibit that progressions to internet searcher results can obstruct re-finding, and give an approach to naturally identify rehash searches and anticipate rehash clicks. They showed that re-finding occurs more at the beginning of session as compared to re-finding at end of session. They also showed Queries occurs across session is more common and queries within session are less common.

In [8] author finds that lot of clients exercises on the web involves returns to pages or destinations. Causes behind returning to sites include dynamic observing of content, verification of information, normal utilization of available facilities, and reappearing tasks. Client browsers bolster for reaccessing is mostly centred on repeatedly and recently visited pages. Author proposed two methods.

One is ranking method and another is propagation method. Ranking method ranks the web pages based on frequency and recency of visits. While propagation method finds more commonly visited web pages along with current web pages. They found that sites can be easily estimated as compared to page estimation and that the enactment of a recommended work predominantly relies upon change in the clients' online behaviour.

III.SYSTEM ARCHITECTURE

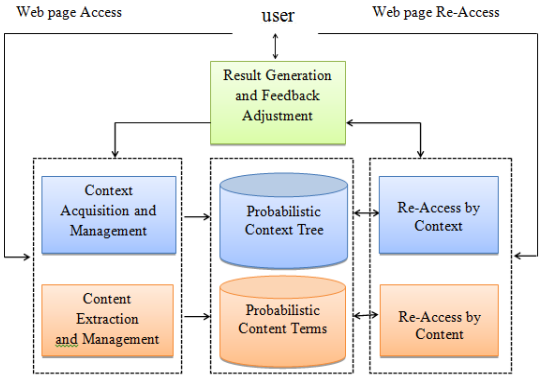

Fig 1: The architecture of proposed system

Fig 1 shows architecture diagram of the proposed system. This will show the control flow of execution of our proposed technique. It mainly contains four components: context acquisition and management module, content extraction module, reaccesses by context module, reaccess by content module and result generation and feedback adjustment module.

When a utilizer accesses a web page and when that web page to be reaccessed later by the utilizer, the context acquisition and management component captures present ongoing activities along with time and location into a probabilistic context tree. Meantime, the content extraction and management component extracts the terms one by one from page segments displayed to the users. Obtained context facts and content terms reflect in what manner the utilizer will utilize them as recollection clues to return to the earlier accessed web page.

In future whenever a utilizer desires to return to an earlier accessed web page through content and context keywords, the reaccess by context component search the corresponding probabilistic context tree and reaccess by content component search the corresponding probabilistic content terms. Both these results are unified by result generation and feedback adjustment and finally yield the utilizer a ranked list of all accessed web page URLs.

IV.METHODOLOGY

We are making use equations to find web page ranking. Equations used in the proposed technique are:

A. Context Acquisition and Management

When the users are accessing the web pages, this module acquires the current context factors into the context tree. Context factors in this module are time, location and associated computer activities. Access time is determinate and location can be determined by using IP address of the computer system. Activities may include reading or writing a document, doing programming in eclipse, listening to songs, playing any games. These activities can be automatically inferred from the computer while accessing web page. Only those activities that are added by the admin are considered.

B. Re-access by Context and Re-access by content

Context and content are the input to our application and they should not be empty. So user should enter at least one context and content term. When the user enters context keywords and content terms, Re-access by context will search context keywords and Re-access by content module search the content terms and display the web pages according to the user entered context and content keywords. This will show how context and content terms helps users in recalling earlier visited web pages.

C. Content Extraction and Management

When the user re-accesses the web page that has visited earlier through context and content keywords, this module extracts the terms one by one from the page segments displayed to the users. This module display URL's having access time more than 30 seconds, on what date user accessed that web page and how much time users spent their time on that web page. The web pages having access time less than 30 seconds will not be considered even though it is accessed by the user.

D. Result Generation and Feedback Adjustment

In this module, web pages will be re-arranged according to the ranking. Ranking can be calculated based on content score and context score. Content score will be calculated using ratio of time length and total frequency of the terms in the web page using equation 3. Counting the number of terms occurred in web page and ratio of time length of the web page will reflect user impression on that web page. Context score will be determined by calculating age of the link and decay rate using equation 4. Context score reflects how likely the users use the context keywords in recalling the earlier viewed web pages. Web page ranking will be calculated using equation 5, which is product of content and context score. Finally, web

pages are sorted according to this ranking. Feedback system collects the feedback from the users and this helps in improving the performance of the application.

V. RESULTS AND DISCUSSIONS

Our practical results show that how content and activities of users while accessing web pages along with time and location helps in finding earlier seen pages. Here activity is best recall indication as compare to time and location. Here context and content are the input and they should not be empty so user should enter at least one context and content term.

The enhancement of user experience is one of the key concerns while we consider for innovative features for web browsers. This motivation drives to consider the user behaviour with respect to usage of activities, location and time during the surfing of web. With these factors we can re consider the revisition concept to next higher level of enhancing and making rich user experience.

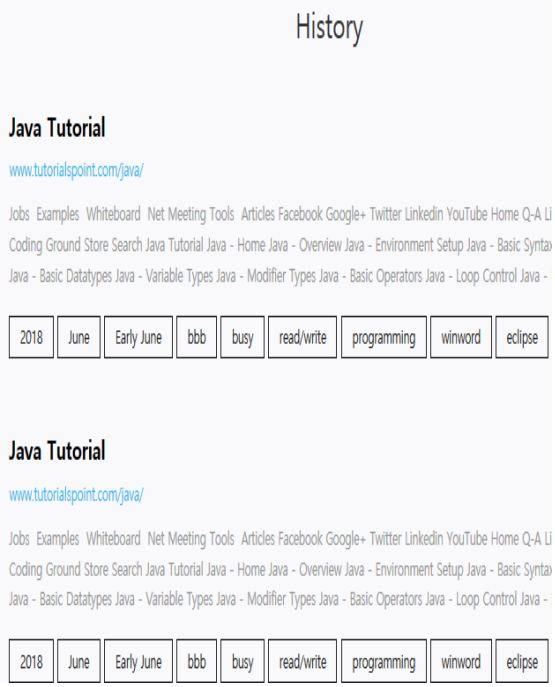

Fig 2: Users history display page

Fig.2. shows the history list of user's accessed web pages. It includes page title, URL along with this information it also contains time, location and associated computer programs.

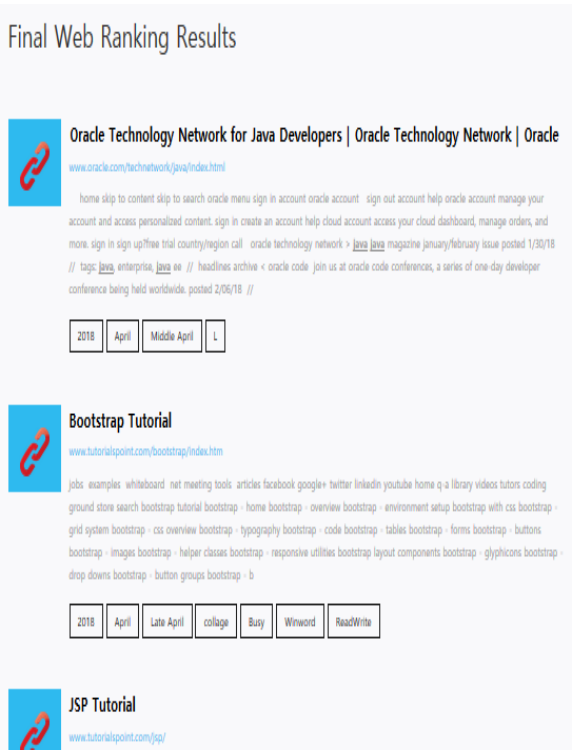

Fig 3: Final web page results

Fig.3. shows final web page ranking results. Ranking for content can be determined by calculating frequency of terms from web pages displayed on user's screen and impression score using equation 2 and 3. Ranking for context can be determined by context score using equation 4. Web page ranking is calculated by multiplying content and context ranking using equation 5.

VI. CONCLUSION AND FUTURE WORK

The work presents results and applicability of the objectives presented. Proposed technique presents an Individual web re-visitation technique based on context and content keywords. The context and page content are separately sorted out as probabilistic context trees and probabilistic content terms, which progressively advances by degradation and support with pertinence feedback.

Upcoming work will be on presage of end users revisitation, elongating the technique to fortify to end-users equivocal refinding requirements, and including convivial social elements in information refinding.

REFERENCES

[1] Cockburn, S. Greenberg, S. Jones, B. Mckenzie, and M. Moyle. Improving web page revisitation: analysis, design and evaluation. IT & Society, 1(3):159–183, 2003.
[2] L. Tauscher and S. Greenberg. How people revisit web pages: empirical findings and implications for the design of history

systems. International Journal of Human Computer Studies, 47(1):97– 137, 1997.
[3] L. Tauscher and S. Greenberg. Revisitation patterns in World Wide Web navigation. In CHI, pages 399–406, 1997.
[4] S. S. Won, J. Jin, and J. I. Hong. Contextual web history: using visual and Contextual cues to improve web browser history. In CHI, pages 1457–1466, 2009.
[5] E. Kulkarni, S. Raju, and R. Udupa. Memento: unifying content and context to aid web page re-visitation. In UIST, pages 435–436, 2010.
[6] J. Hailpern, N. Jitkoff, A. Warr, K. Karahalios, R. Sesek, and N. Shkrob. Youpivot*:* improving recall with contextual search. In CHI, pages 1521–1530, 2011.
[7] S. Tyler and J. Teevan. Large scale query log analysis of re-finding. In WSDM, pages 191–200, 2010.
[8] R. Kawase, G. Papadakis, E. Herder, and W. Nejdl. Beyond the usual suspects: context-aware revisitation support. In HT, pages 27–36, 2011.