EFFECTIVE PERSONAL WEB REVISITATION BY INCORPORATING SOCIAL CONTEXT FACTORS

[1]K . Rajitha, [2]Dr. A. Suresh Babu
[1]PG Scholar, [2]Professor
[1]Department of Computer Science, [2]Department of Computer science
Jawaharlal Nehru Technological University, Anantapur, India.

Abstract**:** *In daily life utilization of a web browser is a prevalent operation, but Personal web history revisitation is an un-facile task for web users due to the abundance of webpage accumulation. In this paper, a personal web revisitation tool is generated by considering human memory and its natural recall characteristics as an inspiration. WebPagePrev is a technique to facilitate web users to collect and revisit history/accessed pages through access context, content and frequency keywords. Context keywords depend on web user's time tree, location, activity context trees. Content keywords depend on visited web page data collection. Frequency depends on consideration of count/frequency of the web page visiting. In this paper, some discussion has done to explain how important to revisit, web pages of web browsers. Among the context, content, and frequency factors, the combination of context, content, and frequency based re-finding delivers the best performance.*

Keywords*: WebPagePrev, Browser History, Revisitation, Web Context, Web Content*

1. INTRODUCTION

The World Wide Web (WWW) has been dramatically increased due to the utilization of the cyber world. The World Wide Web is an immensely colossal hypertext system, provides a repository of information. The web acts as a medium where a substantial amount of knowledge can be obtained at a low cost, in this most frequented user interfaces is Web browsers. Generally, the web browser is utilized to distribute information to the utilizer. People can operate a browser through links and view web pages. The information in the web pages are obtainable in three forms namely, structured, unstructured, and semi-structured. People are searching a different type of information from the web for both reading and writing. The huge growth of the web, every second millions of data/information added to the web. In day to day life, users are finding and re-finding [1] the information on the web. People revisit the web pages that have ever been come across infrequently or intentionally. Web revisitation is to find the previously viewed web pages in history, not only the URL of the web page, but also the web page snaps at that time.

A study of the Psychological [2] shows that a human's memory can vary according to several descriptive features. Humans can use both episodic and semantic memories to recollect the past data and events. Human's memory of episodic [3] gather and stores the information of temporally dated episodes or events. For example, read a book, it the considers the location where a book is read, date like last month, and mood like the user is busy at that time. Semantic memory carries the content information like data, names, and headings. Consider previous example which book, book title, viewpoint of the book and author name. Inspired by the psychological memories, this paper explores how to elaborate our natural recall process of using episodic, semantic memory to perform web revisitation.

In the proposed system web revisitation can be performed by using a technique called WebPagePrev through context, content, and frequency keywords. By using this user can easily revisit the require web pages and read the information on that web page.

2. RELATED WORK

The literature, for the web revisitation [4] concept number of techniques is deployed in current market like bookmarks in browsers, history tools, web search engines and many more systems have been deployed to support the personal web re-visitation. In current search engines, all visited URLs of users have stored in browsers cache memory. And all the clicked links/visited links are fetched in the manner mostly used timestamp manner. All visited urls are stored in date and time manner to help the re-visit.

To bookmark [5] a web page people must mark a page at that time they are visiting a web page, the bookmark-page is maintained in a list. Number of web pages is visited and bookmark all the pages, the bookmark list becomes over messy, containing old and no longer useful pages. In bookmark, list pages are stored in alphabetical order, to find a required page is difficult.

In web browsers, history tools[6] store the accessed URL's according to date of the visited and time, and those URLs will represent to such manners like Today's, Yesterday, This Week and many more along with browsers store the data of accessed URL's Titles and URL. To find a web page from the history user search the entire list and also it is hard to remember the URL title. Search engines like google, bing also trying to facilities of web re-visitations by using some procedures. While displaying

results of the search, search engines provide some differentiate between new links and already visited like a difference in colors of the urls.

A. Cockburna& S. Greenberg[7] describe that The mouse-driven cursor, this will be the major input device for the navigation of the web page, when a user visits a link then the user point to the cursor for the link and revisit a page; then the user place the cursor to the back button. .Task one compared the effectiveness of the two interfaces in depth-first navigation. The way took after four connections on consequent pages at that point backtracked with four progressive Back charges, giving away a→b→c→d→e←d←c←b←a utilizing the documentation presented before. This errand speaks to a coordinated inquiry style of web-use: for example, scanning for an employee's site page beginning from a college's landing page.

Memento[8] defines the context of the web page as other pages in the browsing session that immediately follow the current page. In the web pages,the memento finds the topic phrases based on both content and context. Identifying page-context plays a key role in extracting topic phrases for a target page. Memento identifies the page.

Given the strong evidence from researchers in the cognitive sciences, a human's memory recollects the data based on social events, in what situation user is in. While personal data collection (e.g., MyLifeBits) and storage is not a modern problem, the creation of efficient and timely retrieval techniques and interfaces is central to leveraging this near flawless memory stores of computers [9]. To this end, YouPivot[10], a novel system that directly addresses and demonstrates how computers can implement the contextual search. At the heart of our system is the ability for a user to shift their point of view to that of a file, website, or some other activity in their own digital history.

3. Web Revisitation Using WebPagePrev

Preparation for web re-visitation, Context keyword depends on web user's time tree, location, activity context trees. Content keywords depend on visited web pages and data collection. The Content search module searches keywords, and calculates the tf-idf score from web pages. And the frequency module calculates the frequency of web pages using visitation count.

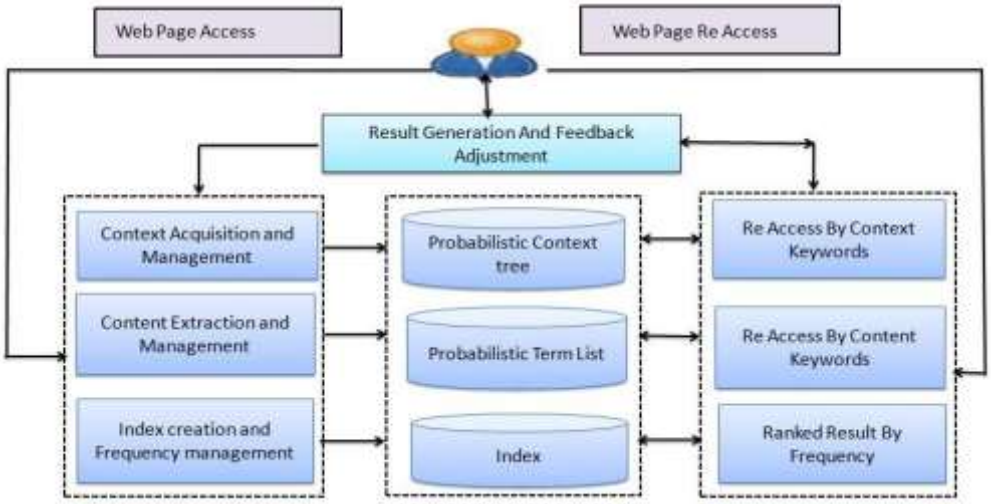

Fig 1: Architecture of the WebPagePrev

Fig 1 describes the architecture flow of the proposed technique. It contains six modules, namely context acquisition and management module, the content extraction management module, index creation and frequency management module re-access by context keywords, re-access by content keywords, ranked result to frequency, and result generation and feedback adjustment

When a user visits a web page, which is revisited later by the user, the context acquisition module capture present context (i. e time, location, and what activity user is performing in the computer system) into context tree. Simultaneously the content extraction module search-related data from the web pages and present them into a term list. And the frequency management module count visited web pages. When a user request to revisit a particular web page through content, context, and frequency keywords, the re-access context keyword module search the context tree, re-access content keyword module search term list and ranked result to frequency module count, the revisited web pages and the result generation and feedback adjustment module combine all three result and deliver the revisited web page to the user.

Context acquisition and management

When a user visiting a web page, context acquisition and management module collect all the context factor into the context tree. Here the context factors are time is determine as accessed time, location is considered as an IP address, and activity is a current running program on the computer system.

Content extraction and management

When a user access a web page content extraction and management module, extract the terms from the web page which is displayed to the user. This module captures the term which url's access time is more than 30 seconds, less than 30 seconds the url's terms are not considered.

Index creation and frequency management

The Frequency management module responsible when a user visiting web page, it will count how many times the user visiting that particular web page. This frequency module categorizes into three models like low, medium, and high.

Re-access by context, Re-access by content keywords and Ranked result by frequency

User requested to revisit a web page by giving inputs are context, content, and frequency. The re-access by context module search the context tree repository, re-access by content module search the content term list repository, and ranked result to frequency module count, the required web page in the index repository. By using this three-module user easily revisit the required web pages.

The Result generation and feedback adjustment

In this module, the ranking of the web page is calculated based on the context score, content score, and frequency index. All three results will be combined to sort the web pages according to ranking.

Algorithms of WebPagePrev

Input: A revisit query Q(W,e,d,t,f)
Output: W_m
Initialization:

i. User Revisitation request Q
ii. W Web pages and W_m relevant result web pages
iii. Tree= getMatchContextTrees (W; c; t); where c context keywords
iv. Lists=getMatchtermList (W; d; t); d content keywords
v. Let t is current time t_0 initial time
vi. W_c List of candidate matched page set based on c, d
vii. Frequency of web page in history

```
For each w∈ Wc
    Calculate dRank of w
        For each di∈ d
            K= calculate tf-idf score
            K=*k
        End for
    dRank=k;
    calculate cRank of w
        For each cTree
            k=calculate λ*e^(√t−t0)
            k=+k
        end for
    cRank=k;
    Web page ranking
        For each w∈ Wc
            Rank(w)=dRank(w)*cRank(w)+Freq
            Wm=sort (w)
        End for
    Return Wm
```

4. EXPERIMENTAL RESULTS

The Practical Result shows how the user revisits require web page by using context, content, and frequency keyword. The frequency is easy to remember (i. e how many times user visits a particular web page is quickly recollect) compare to content and context.

Fig:-2 Activity Context Tree Making Page

Fig 2 describe the context tree making a page, Admin can make the activity context tree by adding what activities the user is performing at that particular time when the user visiting a web page.

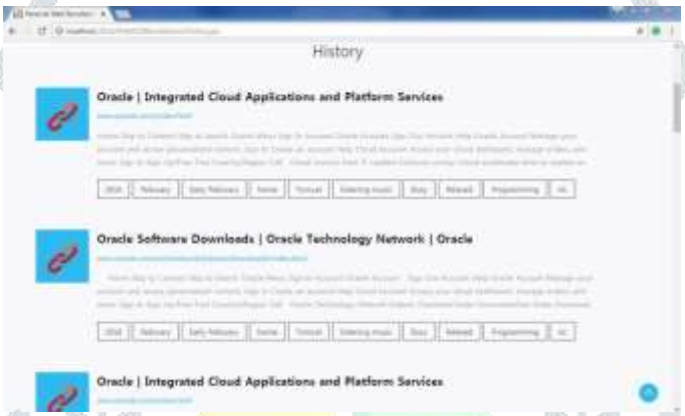

Fig:-3 Search Results

Fig 3 depicts the history list of the user visited web pages. In this list not only shows the URL'S but also display the time, location, and activity of the computer system.

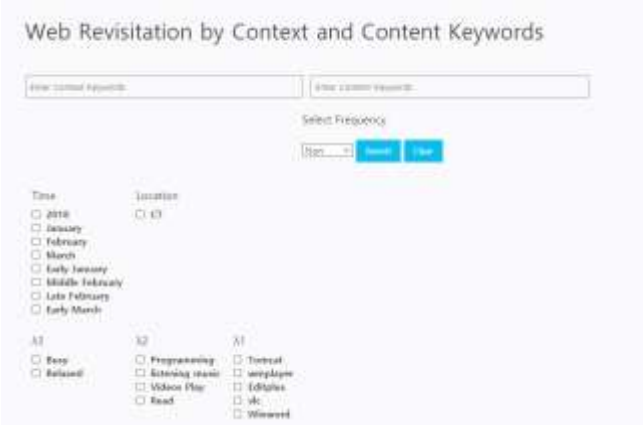

Fig:-4 Context, Content and Frequency Keywords

Fig 4 shows the how content, context, and frequency keywords are used to revisit require web pages. Here the content, context is input given by the user and the frequency is selecting options (how many times the user visited the web page).

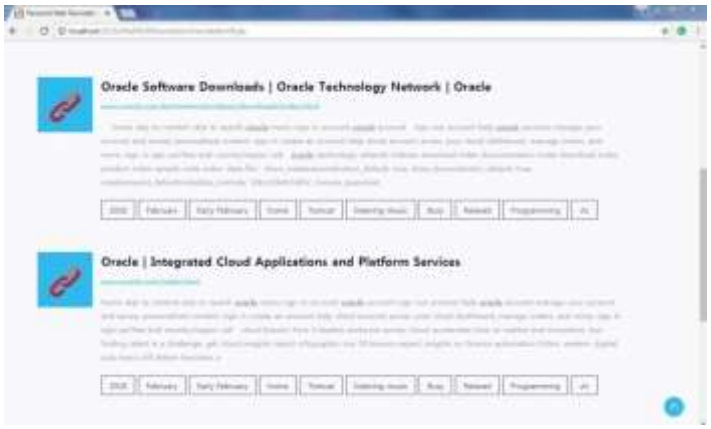

Fig:-5 Final revisitation pages

Fig 5 describes the final web page list which matched to the user content, context, and frequency will be displayed at the top. The content is considered as how many terms are matched with required content.

5. CONCLUSION

This paper present individual web revisitation based on context, content, and frequency keywords. The context, content, and frequency are organized by a probabilistic context tree, a probabilistic content list, and frequency indexing. The experimental results show the effectiveness of the proposed technique. The future work extends this topic to including the social context factors, declaring revisitation before user behavior.

REFERENCES

[1] T. Deng, L. Zhao, H. Wang, Q. Liu, and L. Feng. Refinder: a context-based information re-finding system. *IEEE TKDE*, 25(9): 2119–2132, 2013.

[2]Benjamin, A. S. Memory is more than just remembering: Strategic control of encoding, accessing memory, and making decisions. Psychology of Learning and Motivation, 48, 2007, 175-223.

[3] E. Tulving. What is episodic memory? *Current Directions in Psychological Science*, 2(3):67–70, 1993.

[4] 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.

[5] S. Kaasten and S. Greenberg. Integrating back, history and bookmarks in web browsers. In *HCI*, pages 379–380, 2001.

[6] M. Mayer. Web history tools and revisitation support: a survey of existing approaches and directions. *Foundations and Trends in HCI*, 2(3):173–278, 2009.

[7] A. 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.

[8] C. E. Kulkarni, S. Raju, and R. Udupa. Memento: unifying content and context to aid webpage re-visitation. In UIST, pages 435–436, 2010.

[9] L. C. Wiggs, J. Weisberg, and A. Martin. Neural correlates of semantic and episodic memory retrieval. Neuropsychologia, pages 103–118, 1999.

[10] 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.