Multiuser Webpage Revisitation

Shenara Nazar , Neethu Babu , K T Mariya , Shyam Krishna K
Computer Science And Engineering Sahrdaya College Of Engineering And Technology Thrissur,India

Abstract

Web applications are providing vast informations to users in fraction of seconds. Web Revisitation is a web application that helps users to refind the previously viewed web pages using page URL. A web page can be localized by a fixed url, and displays the page content as time-varying snapshot. Web revisitation is the process of automatically getting back to the previous visited web pages based on user's behaviour and preferencesIt is an important matter in web development framework as it saves many cost and time of the user. Our proposed web revisitation exhibits some form of artificial intelligence by providing memorization techniques of context queries and content queries. There is a relevant feedback system which collects users feedback through thorough analysis of his usage and storing their preferences. We also include a cache memory mechanism in which each page is given a particular score. On each revisitation the score is incremented and updated. Once it reaches beyond the threshold score it is placed into the cache memory and hence can be retrieved easily. Thus our system increases the efficiency of web page revisitation

Keywords — *Revistation, cache, crawling, relevance feedback*

I. INTRODUCTION

Web revisitation is the process of visiting the webpages again that have been already visited. Web revisitation tools are used by almost all the web users. Thus web revisitation tools have much more importance in day today life. There are different types of web revisitation tools available such as history,favourites and so on. The web pages already viewed are stored in a database as they are visited by the users. When in need these web pages can be retrieved using the revisitation tools. Much care should be taken to retrieve the web pages as there are a lot of information available. In the technique described for web revisitation in this paper context and content keywords are used for web revisitation. According to the research conducted it was shown that 81% of the web pages are revisited by the users.

Web revisitation helps user to easily access the web pages within a short span of time. Web revisitation technique return not only the page URL but also the web page snapshot. Humans use both episodic memory and semantic memory to recall the information. Similarly this web page technique uses semantic memory to access contents and episodic memory to access context. This is a multiuser web revisitation technique which aims at returning the visited web pages within a short span of time and stores the visited web pages for a long period of time for future access.

A. ***Motivation***

As today, web is playing a significant role in everyone's day today life.web plays a role in delivering information to users finger tip. . A 6-week user study with 23 participants showed nearly 58% of web access belonged to web revisitation [2]. Another 1-year user study involving 114 participants revealed around 40% of queries were re-finding requests [3]. According to [4], on average, every second page loaded was already visited before by the same user, and the ratio of revisited pages among all visits ranges between 20% and 72%..

Different studies such as a psychological study have shown that people reply on both episodic memory and semantic memory to recall information or events from the past. The episodic memory receives and stores temporally dated episodes or events, whereas semantic memory, on the other hand, is a structured record of facts, meanings, concepts and skills that one has acquired from the external world. In other words, we could say that Semantic information is derived from accumulated episodic memory. These two human user's declarative memory work together in user's information recollecting activities. when a user's web revisitation behavior happens, she/he tends to utilize episodic memory, interweaved with semantic memory, to recall the previously focused pages. Inspired by these findings, this paper explores how to increase the efficiency of natural recall process to facilitate web revisitation. A relevance feedback mechanism is involved to enhance multiuser web revisitation performance.

B. ***Problem Statement***

Web revisitation is a process to re-find the previously viewed web pages, not only the page URL, but also the page snapshot at that access timestamp. Users often revisit pages on the Web, tool support for such re-visitation is still lacking. The existing tools only provide users with basic information such as the date of the last visit and title of the page visited as we see in browser histories. There are number of techniques and tools like,

metadata annotation and exploitation, search engines, history tools, and contextual recall systems have been developed to support personal web revisitation. The most related work of this study is the Memento system, which unifies context and content to aid web revisitation.

In this paper, we introduce a system that provides users with descriptive topic-phrases that aid re-finding.The existing problem states that the most of the revisitation techniques resides on storing user queries on the database and comparing them with a known dataset and previous knnowledge to classify it as either as valid revisitation query (important query) or an invalid query (non-important query).This requires a lot of manual work from humans behalf. And too there are chances of classifying an important query as non-important. By introducing our system we mainly focus on memorization capacity. This can be well explained by taking an example. In google histories we find all the data's are stored in its database but it can't be seen (ie,it can't be seen after a particular number of days). This is the problem we focus on. The hidden data should be revisited whenever required. And this problem can be solved to an extend by introducing our technique

II. EXISTING SOLUTION

YouPivot:Human memory is predicated on contextual cues. In recall tasks, we associate information with contextual cues. Computer systems do not directly borrow our natural process of using contextual cues to facilitate recall. Instead a new interaction technique called Pivoting, that allows users to search for contextually related activities and find a target information. Our interaction technique is originated in the cognitive science literature and is used in the system YouPivot. It describes how principles of human memory can be applied to enhance the search of digital information. A new personal annotation method called TimeMarks is also added to it, to support contextual recall and the pivoting process[5]. In the user study, participants were quicker to identify websites, and preferred using this tool when compared to current tools. It also describes how principles of human memory can be applied to enhance the search of digital information.Contextual cues are main component of human memory.We can define a contextual cue as any object that has a temporal relationship to a search target. They define search based on contextual cues as Contextual Search. Contextual search through contextual search does not have to match semantically to the search target. Research in cognitive science has shown that finding out context improves speed and accuracy in recall tasks. At present, personal file history search is limited to file metadata (e.g., file name, last date accessed, file size, file type and sometimes file content). If a user does not know this information, she cannot search for and find a file. Without a form of contextual search, users cannot leverage the natural process of recall. Search based on contextual cues will enhance recall and computer usability. YouPivot, a contextual history based search tool that bridges the gap between the literature on human memory recall and search[5]. It borrow our natural method of recall by allowing a user to search through their digital history for the context they do remember. The user can then Pivot, or see everything that was going on while that context was active. YouPivot introduces a new annotation method for contextual recall called TimeMarks, in which a user marks a moment in time as being important. This effectively leaves a temporal landmark for later contextual recall. Because YouPivot logs a user's personal activity, TimeMarks effectively bookmarks all the user's activity at that moment for easier recall as in Fig 1. Human memory functions by activating the contextual cues. The probability that the human mind recalls a piece of information is a function of the strength of the association between the context used for recall, and the target information.The richer the set of contextual cues available to an individual, the higher the probability of successful recall. As we increase the number of contexts in which information is acquired, performance on recall tasks also improve.To create a rich set of cues for future use, computer systems must collect a large sets of user data. TimeMarks, are the mechanism to allow a user to decide that the current moment is worth remembering. YouPivot then stores that TimeMark in the history, creating a custom contextual cue. The YouPivot interface is constructed using web technology and the Protivis visualization toolkit. It is platform independant. Users web activity is logged via the Chrome Extension and sent to the YouPivot server. YouPivot's default storage system uses Google's AppEngine. This provides secure authentication through a user's Google account, as well as secure cloud-based storage, and a dynamic and fast architecture. This infrastructure does not store the user's activity, rather it retains a record of the occurrence. This improves performance speed, security, and privacy.Whenever a user would like to look at their personal history, they log into their Google account. This provides a secure method for viewing and interacting with their stored data. All data is transferred over secure HTTP. To log data into YouPivot, applications send the data through secure web GET requests. When a user first signs up for YouPivot, they are given a unique alpha-numeric string. This string is used by any application willing to log data into YouPivot, by passing the alpha-numeric sequence with a submission request.

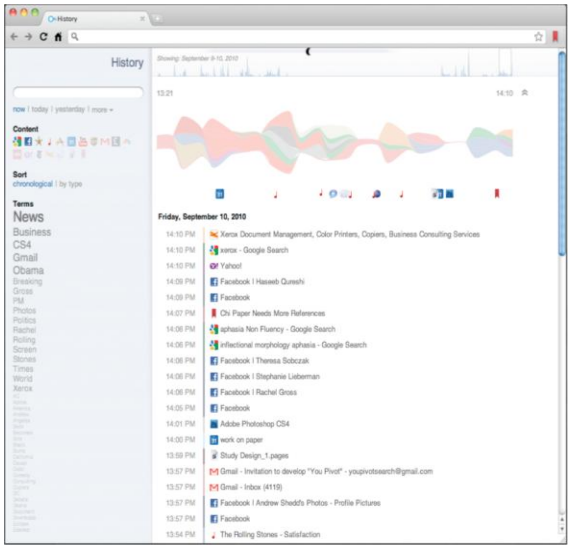

Fig 1. Youpivot System

Memento System: Revisitation is the process in which revisiting the pages that we have already visited and the current tools (such as browser histories) for various webpage revisitation only provide users with basic information such as the date of the last visit and title of the page visited. The basis of webpage revisitation technique starts with Memento: Unifying Content and Context to Aid Webpage Re-visitation. In' Memento system', it considers both the content of a webpage and the context in which the page was visited. Unfortunately, during 2010 where this system was established, those time "re-finding" was often difficult using web-browsers. Users have reported "Not being able to return to a page I once visited," as one of "the biggest problems in using the Web" [1]. Although earlier studies of directed search focused on keyword search, most of the search behavior we observed did not involve keyword search [2].In the memento system, there are various phases such as identifying page context, topic mining etc. In the first phase ie,Identifying page-context plays a key role in extracting various topic phrases for the required target page. Memento identifies the page context by determining the session boundary. Session is defined as the number of pages considered in the page's context. The length of a session varies based on the user's information need.In the second phase, topic-phrases are mined using content from the page and a pool of pages identified as its context.In this phase it ensures that the phrases are well-formed and meaningful.This system was not found effective as it doesn't differentiates between an important query and non-important query.

Context Web History(CWH): Most of the modern web browsers are having history functionality but few people are awared and use it to revisit previously viewed web pages. Contextual web history is the history tracking technique which aims to improve the usability and utility of the history feature in web browsers[6]. CWH is developed to be more visible to users and supported search along with it helps browsing, thumbnails, and data about data.These relatively simple features are combined and performed on Mozilla Firefox 3's built-in browser history function which greatly reduced the time and effort needed to find and revisit a web page. Contextual web history seeks to improve the browser history functionality in general. A person can use contextual web history directly if they know that they are looking for a specific page they have already visited, rather than also seeing search results for pages that may not be relevant for their current needs as the normal search results in the browsers. It is depending more on a combination of existing searching and browsing style interfaces that people are more familiar with and are easier to build on top of the existing web browser APIs. Visual elements were found useful and were incorporated into contextual web history in the form of thumbnails to helps recognition.Two interactive prototypes built as Firefox extensions. They focus on the interactive prototypes[6].

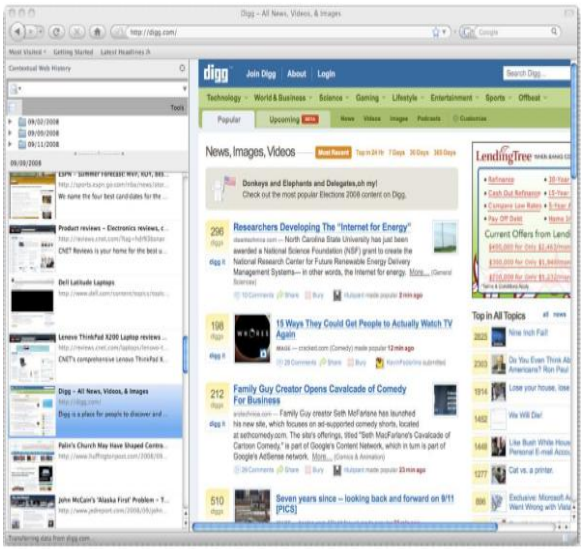

Fig 2.Contextual Web History

In the first prototype, added a search toolbar and a toolbar icon to the web browser chrome to increase visibility of contextual web history. The search toolbar lets the user access contextual web history by providing a search query. Users could input different types of information such as keywords, dates, and times. All input was processed within the single textbox so that the user could quickly get to the results as in the Fig 3.

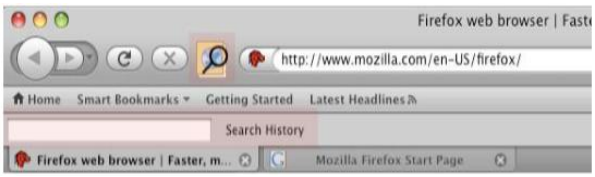

Fig 3. Search tab and search icon

In second prototype,integrated the search toolbar with the Firefox 3's search bar located on the top of the web browser as in the Fig 2. Hereimproved the usability and overall design based on findings from the previous prototype, making contextual web history more visible, easy to use and coherent with the rest of the browser. The types of information remained the same and made several variations to the front-end. They integrated contextual web history more with the browser, moving all the core features browsing, searching, and displaying results into a sidebar. Users can now browse the entire history and search results with thumbnails, titles, urls, and descriptions. From the first prototype evaluation, we found that people still tended to use their favorite search engines to revisit websites, because they are more familiar with that .To accommodate this feature, our second prototype alerts the user to relevant history results when searches are made in web-based search engines.contextual web history was implemented as a Firefox 3 extension using JavaScript, XUL, and HTML.The latest prototype extends Scrapbook, a Firefox 3 extension. They chose Scrapbook because it saves copies of visited web pages and allows full-text search. The features were added to Scrapbook include storing and displaying thumbnails for each visited page, displaying contextual web history results whenever a user searches on a search engine and having direct contextual web history search input in the toolbar. A thumbnail is saved by taking an automatic low-resolution screenshot of the page using the Canvas property in Mozilla API.They embedded CWH results into Google's search result page by modifying the DOM with JavaScript. Keywords are extracted and obtained from the title of the webpage, URL, meta tag, and content of the web page by eliminating HTML markup and eliminating words less than three letters.

HiBO A system for automatically organizing bookmarks:HiBO is an extension to the bookmarks system that organizes the bookmarks according to the topics[8]. It allows user to customize and organize the bookmarks system. It organize the bookmarks into a hierarchical structure. HiBO uses a built-in mechanism for automatically organizing the bookmarks within the local repositories. HiBO works by downloading the pages that are bookmarked by the users. Then it process these bookmarks one by one and then identify the important terms in it. Finally it organize them into a hierarchy. HiBO perform its functionalities through a series of steps such as downloading the URL and parsing the HTML pages, segment the contents of the pages, mapping the words into a hierarchical structure, computing a relevance score and then retrieving the web pages according to the greater relevance score.Advantages of HiBO system include that it access the web pages using queries. It helps users to customize and organize the web pages. Thus the web pages can be accessed easily and within a short span of time.

III. PROPOSED WORK

Based on user action logs, we distinguished revisitation process as short-term revisits, medium-term and long-term revisits. We analyzed the current problems and provide suggestions for improving support for different revisitation types. For say, in browser histories the visited webpages are stored and on revisitation ,the revisited webpages can only be seen for past one month. This limitation(ie,pages can be retrieved only if it lies within one month)can be resolved to an extend by our proposal of new work. Fig 4.depicts the system architecture of our system.

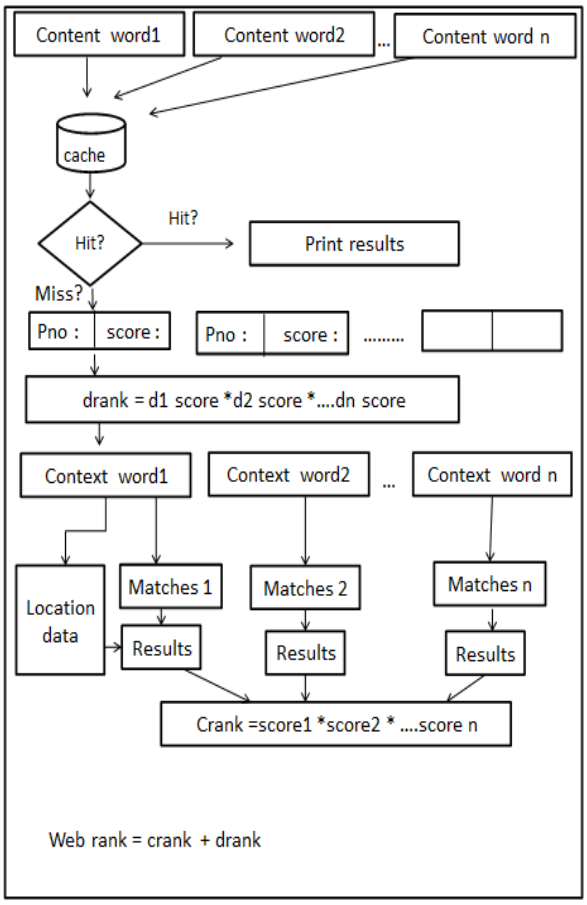

Fig 4. System Architecture

The main contributions of our paper thus lie in the following aspects:

- The end product of our project is a revistiation search engine.
- We present a multiuser web revisitation technique, called smart search, that allows users to get back to their previously focused pages through access context and page content keywords.
- Dynamic tuning strategies to enhance the individual's memorization strength and recall habits

based on relevance feedback are developed for performance improvement.

• We evaluate the effectiveness of the proposed technique smart search and report the findings in web revisitation through a 3-month user study with 17 participants.

The proposed system categories into following modules

1. Web User Module
2. Admin Module
3. Chart Module
4. Image Upload Module
5. Feedback module

Web User Module: The web user (User) is assigned a global user identity Uid .User should register page before going toto login the page. While login user should give location because based on location we can search the products and search keys after login user can search anything like a search engine ,they can view the similar product, similar word recommendation help for user.

Admin module: Admin module, admin is super user,admin can view all the details,here admin can view the chart based on user product revisitation,user can add the shop details andproduct details based on location,semantic memory cues to facilitate recall, and presents a personal web revisitation technique called smart search through context and content keywords.

Chart module: chart module based on number times product revisit by someone,so user can easily find out which product move very quickly it's all based on user searching, User relations and temporal property are used simultaneously in this work. Prediction strategy is used to predict the subsequent mobile behavior.

Image upload module: Upload image module, admin can upload any image , after uploading the image, user can provide correct authentication details they can view file and presents a multiuser web revisitation technique called smart search through context and content keywords.

Feedback module: A relevance feedback module is also integrated to collect users search history and count the total query usages and display the revisitation queries based on the incremental order and likeliness of the queries.

Relevance feedback mechanism

Users of various online search engines often find it difficult to express their need for information in the form of a query[8].If the user can identify examples of the kind of documents they require then they can employ a technique known as relevance feedback. Relevance feedback covers a range of techniques intended to improve a user's query and facilitate retrieval of information relevant to a user's information requirement.

Fig 5.Crawl page

IV. ANALYSIS OF EXISTING WORK

The conducted formative studies to understand reason why people didn't use the history function, and what cues people found useful for recall and recognition of infrequently visited web pages. To avoid the difficulties in using existing web histories, contextual web history provides a set of cues about the content of a page, time of visit, visual appearance, and text search. Contextual web history seeks to improve the browser history functionality in general. A person can use contextual web history directly if he know that they are looking for a specific page they have already visited, rather than also seeing search results for pages that may not be relevant for their current needs as the normal search results in the browsers. From user studies on history feature, we tried to obtain a qualitative understanding of how people found previously visited web pages and what aspects of those web pages were most useful and memorable for retrieval. The user studies consisted of an interview and several recall and recognition tasks. It is depending more on a combination of existing searching and browsing style interfaces that people are more familiar with and are easier to build on top of the existing web browser APIs. Visual elements were found useful and were incorporated into contextual web history in the form ofthumbnails to helps recognition. From the interview to understand participants revisitation patterns and their understanding of the default history features in browsers, even the recalled incidents were obtained but revisit a web page were not successful due to too many unhelpful results. In the recall part of the study, conducted user studies to see what aspects of web pages people could easily remember[6]. From the studies,cues such as Colors on a site, excluding images, Visual structure & layout of page, Time user visited the site, Logos, prominent images, Presence of animated content, Title of the page, Domain name of the URL, Path and Filename of URL. The goal was to gain a qualitative understanding of what aspects of web pages people could remember. In recognition part of study, the goal was to evaluate how well thumbnails would aid in revisiting web pages, and what size of thumbnails was best for participants and also understand the best way of integrating

thumbnails and using them for contextual web history prototypes. The entire results from the user studies done, half of the participants were not aware of the browser's history functionality and other users stopped using the history feature because it provided too many results and the information displayed was not helpful. To address the former issue, the eventual solution was to integrate contextual web history with the browser search functionality, and to added relevant historical results within web-based search engine results. Initially, we only supported searching on history, but later found that browsing also useful for finding recent web pages. From user studies, we found that people were best at recalling text and visual characteristics of web pages. Other characteristics, such as time of visit and URL were somewhat useful, so they include these features, giving users multiple ways to search and filter results. The evaluation compared the effectiveness of the prototype against both the users preferred method of revisitation and our proposed sysytem. The effectiveness of each method was evaluated by counting the number of actions the user performed. For total number of actions, we summed up the number of intentional mouse hovers and clicks, words typed, and directional keystrokes[6]. Overall, users will take less steps looking for a web page using our proposed system compared to other methods.

YouPivot:The purpose was to demonstrate that YouPivot could support and improve recall given some contextual cues, compared to existing methods. Also our proposed system is above the Youpivot system in the performance level. The browser's history which is of two levels, The traditional browsing history user interface condition and the contextual browsing history user interface condition. From the user study, for each user, the total time to identify a website for all tasks was calculated. Tasks with the contextual search had a mean time lesser than our proposed system. Statistical significant differences were between the traditional history user interface condition and the contextual history user interface condition. YouPivot performs faster than traditional history. Our proposed system performs faster than the Youpivot, since it is having both context and content keywords. Statistical significant differences were found between participants satisfaction ratings for the traditional history user interface condition and the contextual history user interface condition. Therefore, participant were more satisfied with the contextual history user interface condition, compared to the traditional history user interface condition.

V. CONCLUSION

Drawing on the characteristics of human memory in organizing and exploiting episodic and semantic words in information recall, this paper presents a multiuser web revisitation technique based on context and content keywords.The current system of us can easily assign different weights to the page-content and page-context. We focus on the process of crawling too. Crawling is the process which is used up by the search engine crawlers to visit and download the page and extract its links in order to discover additional pages.Fig 5 indicates the interface of crawling page.

REFERENCES

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

[2] Mayer, M. Web History Tools and Revisitation Support: A Survey of Existing Approaches and Directions. In Foundations and Trends in Human-Computer Interaction, Mar 2009.

[3] Rainie, L. and Shermak, J. Search engine use, Pew Internet & American Life Project, Washington DC, 2005.

[4] Rizzo, F. et al. Evaluating the semantic memory of web interactions in the xMem project, In Proc. AVI 2006 .

[5] Personal Web Revisitation by Context and Content Keywords with Relevance Feedback Li Jin, Gangli Liu, Chaokun Wang and Ling Feng, Senior Member, IEEE,2017.

[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. 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.

[8] I. Ruthven and M. Lalmas. A survey on the use of relevance feedback for information access systems. Knowledge Engineering Review, 18(2):95–145, 2003.

[9] Pavlos Kokosis Vlassis Krikos Sofia Stamou Dimitris Christodoulakis, HiBO: A System for Automatically Organizing Bookmarks, IEEE.