

## A HIGH SECURED REVISITATION IN WEB BASED ON CONTEXT AND CONTENT KEYWORDS WITH RELIABLE FEEDBACK

<sup>1</sup>Y Vijetha, <sup>2</sup>J Hymavathi, <sup>3</sup>Rammohanreddy Dondeti, <sup>4</sup>Chella Sivaprasad  
<sup>1,2,3</sup>Assistant Professor, <sup>4</sup>Student, Dept. of Computer Science Engineering, Newton's Institute of Engineering, Macherla, Andhra Pradesh, India.

### ABSTRACT

Nowadays, the web is playing a significant role in delivering information to user's fingertips. A web page can be localized by a fixed URL, and displays the page content as time-varying snapshot. Among the common web behaviors, web re-visitation is to re-find the previously viewed web pages, not only the page URL, but also the page snapshot at that access timestamp. Getting back to previously viewed web pages is a common yet uneasy task for users due to the large volume of personally accessed information on the web. This project "Personal Web re-visitation by Context and Content Keywords with Relevance Feedback" leverages human's natural recall process of using episodic and semantic memory to facilitate recall, and presents a personal web re-visitation technique called WebPagePrev through context and content keywords. Among time, location and activity context factors WebPagePrev technique is the best as it delivers the best performance compared to context based re-finding and content based re-finding. This supports fast and effective finding and re-finding with Meta data Annotation and Exploitation. This could be implemented through Probabilistic content tree algorithm and Probabilistic Term list algorithm which were developed based upon the human brain functioning in reference to semantic and episodic memory.

### 1. INTRODUCTION

Nowadays, the web is playing a significant role in delivering information to users' fingertips. A web page can be localized by a fixed URL, and displays the page content as time-varying snapshot. Among the common web behaviors, web re-visitation is to re-find the previously viewed web pages, not only the page URL, but also the page Snapshot at that access timestamp.

Psychological studies show that humans rely on both episodic memory and semantic memory to recall information or events from the past. Human's episodic memory receives and stores temporally dated episodes or events, together with their spatial-temporal relations, while human's semantic memory, on the other hand, is a structured record of facts, meanings, concepts and skills that one has acquired from the external world. Semantic information is derived from accumulated episodic memory. Episodic memory can be thought of as a "map" that ties together items in semantic memory. The two memories make up the category of human user's declarative memory, and work together in user's information recollecting activities. Thus, when a user's web re-visitation behavior happens, s/he tends to utilize, Episodic memory interweaved with semantic memory, to recall the previously focused pages. Here, semantic memory accommodates content information of previously focused pages, and episodic memory keeps these pages' access context (e.g., time, location, concurrent activities, etc.) The most closely related work of this study is Memento system, which unifies context and content to aid web re-visitation. It defined the context of a web pages other pages in the browsing session that immediately precede or follow the current page, and then extracted topic-phrases from these browsed pages based on the Wikipedia topic list. content items from the full web page as done in Memento system we extract them from page segments displayed on the screen in the user's view, and assign a probabilistic value to each extracted term based on user's

page browsing behaviors (i.e., dwell time and highlighting), as well as page's subject headings and term frequency-inverse document frequency reflecting user's impression and likeliness of using the keyword as recall content cues. Other closely related works enabled users to search for contextually related activities (e.g., time, location, concurrent activities, meetings, music playing, interrupting phone call, or even other files or web sites that were open at the same time), and find a target piece of information (often not semantically related) when that context was on.

This body of research emphasizes episodic context cues in page recall. How to grasp possibly impressive semantic content cues from user's page access behaviors, and utilize them to facilitate recall are not discussed. To tailor to individual's web revisitation characteristics, as well as human user's context and content memory degradation nature, this study presents methods to dynamically tune influential parameters in building and maintaining probabilistic context and content memories for recall.

Inspired by the psychological findings, this project explores how to leverage our natural recall process of using episodic and semantic memory cues to facilitate personal web revisitation. Considering the differences of users in memorizing previous access context and page content cues, a relevance feedback mechanism is involved to enhance personal web revisitation performance. To support personal web revisitation, a number of techniques and tools are developed, including bookmarks, history tools, search engines, metadata annotation and exploitation, and contextual recall systems.

**Bookmarks:** Apart from back=forward buttons, manually/automatically bookmarking favorite web pages in web browsers enables users to get back to the previously accessed pages. According to user's every visited web page and browsing preferences built bookmarks automatically and organized them into a recency list or layered structure, respectively. Gamezet al. further used classifiers to forecast a few of the bookmarks that are more probably to be visited later and showed them in the browser bookmarks personal toolbar, so that the user can access the desired webpage through a single mouse click..

**History Tools:** History tools of web browsers maintain user's accessed URLs chronologically according to visit time (e.g., today, yesterday, last week, etc.), and accessed page titles and contents. Tauscher and Greenberg analyzed 6 weeks of usage data collected from 23 participants when using a commercial browser Mosaic, and discovered that people tend to revisit pages just visited, access only a few pages frequently, browse in very small clusters of related pages, and generate only short sequences of repeated URL paths, which can be used to develop guidelines for the design of history mechanism.

Google Web History<sup>1</sup> keeps user's search keywords and clicked pages, and categorizes them into image, news, ordinary page, etc. Users can navigate or search the history by page title/content keywords. Contextual Web History improved the visual appearance of the web browser history by combining web site thumbnails and content snippets to assist users to easily browse or search their histories by time. Visual History Tool encoded four features of a visited web page, which consist of user's page interests measured by dwell time, the frequency and recency of the visit, and navigational associations between pages. List- and graph-based forms are then adopted to provide navigation histories, improved history mechanisms by intermixing semantic aspects with the temporal dimension of the accessed pages. It organized the pages into groups and presented a navigational history instead of simply exploiting time sort history. Search Panel combined web page and process metadata into an interactive representation of the retrieved documents that can be used for sense-making, navigation, and re-finding documents.

**Search Engines:** Tyler and Teevan studied how search engines are used for re-finding previously found search results. It explored the differences between queries that had substantial/minimal changes between the previous query and the revisit query. Through observing the differences between re-finding behavior occurring within the same session and across multiple sessions, the results showed that cross-session re-finding may be a way to bridge a task between two different sessions. Research supported simultaneous finding and re-finding on the

web. Past queries were indexed to identify repeated searches, and the most recently viewed results were stored in a result cache

**Metadata Annotation and Exploitation:** Haystack stored arbitrary objects of interest to a user, and recorded arbitrary (predefined or user-defined) properties of and relationships between the stored information. It coined a uniform resource identifier (URI) to name anything of interest, including a document, a person, a task, a command/menu operation, or an idea. Once named, the object can be annotated, related to other objects, viewed, and retrieved through arbitrary properties, which served as useful query arguments, as facets for metadata-based browsing, or as relational links to support the associative web browsing. Bearing the similarity to Haystack, a SQL based MyLifeBits platform was built for recording, storing, and accessing a personal lifetime archive. It stored content and metadata for a variety of item types, including contacts, documents, email, events, photos, music and video, which were linked together implicitly using “time”, or explicitly linked with typed links such as a “person in photo” link between a contact and a photo, or a “comment” link between a voice comment and a document. With linking, the traditional folder (directory) tree was replaced by a more general “collections” function using a directed acyclic graph (DAG).

## II. problem statement

The motivation is to provide quick and easier re-access to web with the help of techniques based on human’s natural recall process. The main aim of our project is to provide revisitation of personal web by context and content keywords. The objective is to eliminate the drawbacks present in the existing systems and to provide better approaches for personal web revisitation

## III. LITERATURE SURVEY

### Beyond the usual suspects context-aware revisitation support

A considerable amount of our activities on the Web involves revisits to pages or sites. Reasons for revisiting include active monitoring of content, verification of information, regular use of online services, and reoccurring tasks. Browsers support for revisitation is mainly focused on frequently and recently visited pages. In this paper we present a dynamic browser toolbar that provides recommendations beyond these usual suspects, balancing diversity and relevance. The recommendation method used is a combination of ranking and propagation methods. Experimental outcomes show that this algorithm performs significantly better than the baseline method. Further experiments address the question whether it is more appropriate to recommend specific pages or rather (portal pages of) Web sites. We conducted two user studies with a dynamic toolbar that relies on our recommendation algorithm. In this context, the outcomes confirm that users appreciate and use the contextual recommendations provided by the toolbar.

### Improving web page revisitation: analysis, design and evaluation

Several years of research suggest improvement is needed in how people return to their previously visited Web pages. Web page revisitation is one of the most frequent actions in computer use, so any interface improvements in this area can have a very large effect. Five categories of revisitation research are involved: 1) Characterizations of user behavior; 2) System models of navigation and their impact on the user’s understanding; 3) Interface methods for increasing the efficiency of the Back button; 4) Alternative system models for navigation; and 5) Alternative methods for presenting Web navigation histories. Revisitation is a dominant activity, with an average of 80% of page visits being to previously seen pages. The Back button is heavily used, but poorly understood. Three interface strategies for improving Web page revisitation are described: 1) A gesture-based mechanism for issuing the frequent Back and Forward commands addresses low-level interface issues; it is shown to be both popular and effective; 2) A "temporal" behavior for the Back and Forward buttons aims to overcome the problems associated with poor understanding of the current behavior of Back, strongly

suggesting that revisitation can be improved by providing temporally ordered lists of previously visited pages; 3) Next-generation browsers could integrate the current tools for revisitation into a single utility, thus allowing simple visualization methods to aid users in identifying miniature target pages.

#### **A survey on the use of relevance feedback for information access systems .**

Users of online search engines often find it difficult to express their need for information in the form of a query. However, 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 need. In this paper we survey relevance feedback techniques. We study both automatic techniques, in which the system modifies the user's query, and interactive techniques, in which the user has control over query modification. We also consider specific interfaces to relevance feedback systems and characteristics of searchers that can affect the use and success of relevance feedback systems.

#### **Personalized social search based on the user's social network.**

This work investigates personalized social search based on the user's social relations -- search results are re-ranked according to their relations with individuals in the user's social network. We study the effectiveness of several social network types for personalization: (1) Familiarity-based network of people related to the user through explicit familiarity connection; (2) Similarity-based network of people "similar" to the user as reflected by their social activity; (3) Overall network that provides both relationship types. For comparison we also experiment with Topic-based personalization that is based on the user's related terms, aggregated from several social applications. We evaluate the contribution of the different personalization strategies by an off- line study and by a user survey within our organization. In the off- line study we apply bookmark-based evaluation, suggested recently, that exploits data gathered from a social bookmarking system to evaluate personalized retrieval. In the on- line study we analyze the feedback of 240 employees exposed to the alternative personalization approaches. Our main results show that both in the off- line study and in the user survey social network based personalization significantly outperforms non-personalized social search. Additionally, as reflected by the user survey, all three SN-based strategies significantly outperform the Topic-based strategy.

#### **How people revisit web pages: empirical findings and implications for the design of history systems.**

We report on users' revisitation patterns to World Wide Web (web) pages, and use the results to lay an empirical foundation for the design of history mechanisms in web browsers. Through history, a user can return quickly to a previously visited page, possibly reducing the cognitive and physical overhead required to navigate to it from scratch. We analyzed 6 weeks of detailed usage data collected from 23 users of a well-known web browser. We found that 58% of an individual's pages are revisits, and that users continually add new web pages into their repertoire of visited pages. People tend to revisit pages just visited, access only a few pages frequently, browse in very small clusters of related pages and generate only short sequences of repeated URL paths. We compared different history mechanisms, and found that the stack-based prediction method prevalent in commercial browsers is inferior to the simpler approach of showing the last few recently visited URLs with duplicates removed. Other predictive approaches fare even better. Based on empirical evidence, eight design guidelines for web browser history mechanisms were then formulated. When used to evaluate the existing hypertext-based history mechanisms, they explain why some aspects of today's browsers seem to work well, and other's poorly. The guidelines also indicate how history mechanisms in the web can be made even more effective.† ( 1997 Academic Press Limited

#### IV. SYSTEM ANALYSIS

Apart from back=forward buttons, manually/ automatically bookmarking favorite web pages in web browsers enables users to get back to the previously accessed pages. According to user's every visited web page and browsing preferences built bookmarks automatically and organized them into a recent list or layered structure , respectively. Gamezet al. [14] further used classifiers to forecast a few of the bookmarks that are more probably to be visited later and showed them in the browser bookmarks personal toolbar, so that the user can access the desired web page through a single mouse click.

Tyler and Teevan studied how search engines are used for re- finding previously found search results. It explored the differences between queries that had substantial/minimal changes between the previous query and the revisit query. Through observing the differences between re-finding behavior occurring within the same session and across multiple sessions, the results showed that cross-session re- finding may be a way to bridge a task between two different sessions.

This body of research emphasizes episodic context cues in page recall. How to grasp possibly impressive semantic content cues from user's page access behaviors, and utilize them to facilitate recall are not discussed.

The differences of users in memorizing previous access context and page content cues, a relevance feedback mechanism is involved to enhance personal web re-visitation performance.

Preparation for web revisitation. When a user accesses a web page, which is of potential to be revisited later by the user (i.e., page access time is over a threshold), the

context acquisition and management module captures the current access context (i.e., time, location, activities inferred from the currently running computer programs) into a probabilistic context tree.

Web re-visitation: Later, when a user requests to get back to a previously focused page through context and/or content keywords, the re-access by context keywords module and re-access by content keywords module search the probabilistic context tree repository and probabilistic term list repository, respectively. The result generation and feedback adjustment module combines the two search results and returns to the user a ranked list of visited page URLs.

We present a personal web re- visitation technique, called WebPagePrev that allows users to get back to their previously focused pages through access context and page content keywords. Underlying techniques for context and content memories' acquisition, storage, and utilization for web page recall are discussed.

Relevance feedback is an interactive approach that has been shown to work particularly well in classical information retrieval and more recently in web search domain. When a user interacts with WebPagePrev during web re-visitation phase, he can manually pick up suggested values from contextual hierarchies by clicking the time, location tree views. Each contextual hierarchy is dynamically maintained by analyzing the user's clicking behaviors and the statistical frequencies of captured context instances.

#### IV.IMPLEMENTATION

##### Admin

In this module, the Admin has to login by using valid user name and password. After login successful he can do some operations such as view all user and their details and authorize them, admin can upload data on different topics which can be viewed by user based on search keyword.

##### User

User can register to application and login to application by selecting location. User can search for files based on keyword and get results with most matched search results. When user view data there are process like Context acquisition, Content Extraction and Management and Construction of Probabilistic Context Trees.

##### Context Acquisition

Access context, i.e., access time, access location, and concurrent activities, are captured. While access time is determinate, access location can be derived from user registration details.

User's concurrent activities are inferred from his/her computer programs, running before and after the page access. We continuously monitor the change of user's focused program windows, which can be a web page. Once a user visits a web page longer than a threshold  $\_c$ , computer programs that run interleaving with the current web access program for over  $\_c$  time are taken as the associated computer programs

Assume a web page access program  $w[ws;we] = (w:title;w:dur;w:freq)$  accesses a web page at time  $ws$ , leaves the page at time  $we$ , and the total visit time of the page (i.e., the total focus duration time of program  $w$ ) is longer than  $\_c$ . Computer program  $c$  is called an **associated computer program (context activity)** within the  $w$ 's interleaving window.  $[ws- ;we+\Delta]$ , denoted as  $c[ws- ;we+\Delta] = (c:title; c:dur; c:freq)$ , if and only if  $(c:dur > \_c)$ .

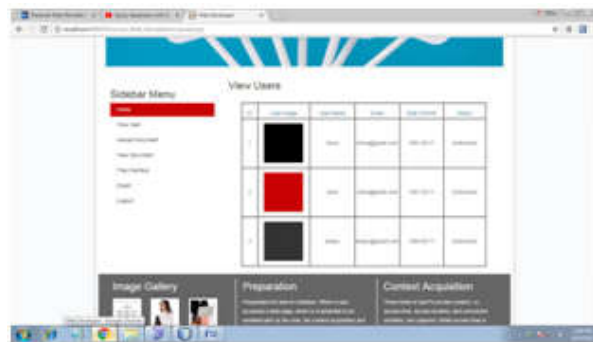
### Content Extraction and Management Module

Users may also get back to the previous viewed pages through some content keywords. Instead of extracting content terms from the full web page, we also extract data from visited pages

### Construction of Probabilistic Context Trees

Access context (i.e., time, location, and concurrent computer programming activities) is organized in a **probabilistic context tree** to support generalized revisit queries due to human user's cognitive understanding and progressive decay during learning and recalling processes.

## V. EXPERIMENTAL RESULTS





## CONCLUSION

Drawing on the characteristics of human brain memory in organizing and exploiting episodic events and semantic words in information recall, this paper presents a personal web revisitation technique based on context and content keywords. Context instances and page content are respectively organized as probabilistic context trees and probabilistic term lists, which dynamically evolve by degradation and reinforcement with relevance feedback. Our experimental results demonstrate the effectiveness and applicability of the proposed technique. Our future work includes 1) prediction of user's revisitation, 2) extending the technique to support users' ambiguous re-finding requests, and 3) incorporating social context factors in information re-finding.

## REFERENCES

1. S. Yu, C. Wang, K. Ren, "Achieving secure, scalable, and fine-grained data access control in cloud computing," Proc. IEEE INFOCOM, pp. 1-9, 2010.
2. J. Bethencourt, A. Sahai, B. Waters, "Cipher text-policy attribute-based encryption,"
3. J. Hur, D.K. Noh, "Attribute-based access control with efficient revocation in data outsourcing systems," IEEE Transactions on Parallel and Distributed Systems, vol. 22, no.7 pp. 1214-1221, 2011.
4. A. Lewko, B. Waters, "Decentralizing attribute-Based encryption," Proc. Advances in Cryptology-EUROCRYPT, pp. 568-588, 2011.
5. M. Li, S. Yu, Y. Zheng, "Scalable and secure sharing of personal health records in cloud computing using attribute-Based Encryption," IEEE Transactions on Parallel and Distributed System, vol. 24, no. 1, pp. 131- 143, 2013.
6. C.K. Chu, S.S.M. Chow, W.G. Tzeng, "Key-aggregate cryptosystem for scalable data sharing in cloud storage," IEEE Transactions on Parallel and Distributed Systems, vol. 25, no. 2, pp.468-477, 2014.
7. J. Li, K. Kim, "Hidden attribute-based signatures without anonymity revocation," Information Sciences, vol. 180, no. 9, pp. 1681-1689, 2010.
8. H.K. Maji, M. Prabhakaran, M. Rosulek, "Attribute-Based Signatures," Proc. Topics in Cryptology - CT-RSA, pp. 376-392, 2011.
9. S. Kumar, S. Agrawal, S. Balaraman, "Attribute based signatures for bounded multi-level threshold circuits," Proc. Public Key Infrastructures, Services and Applications, pp. 141-154, 2011.