

A STRATEGY ON PERSONAL WEB REVISITATION USING CONTEXT AND CONTENT KEYWORDS BY RELEVANCE FEEDBACK

¹Varsha Reddy, ²Hashika Lewis, ³Sainath Reddy, ⁴Nagarjun Chowdary, ⁵D. Krishna

^{1,2,3,4}Student, ⁵ Assistant Professor
Computer Science and Engineering
Martin's Engineering College, Hyderabad, India

Abstract : 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 paper leverages human's natural recall process of using episodic and semantic memory cues to facilitate recall, and presents a personal web revisitation technique called WebPagePrev through context and content keywords. Underlying techniques for context and content memories' acquisition, storage, decay, and utilization for page re-finding are discussed. A relevance feedback mechanism is also involved to tailor to individual's memory strength and revisitation habits. Our 6-month user study shows that: (1) Compared with the existing web revisitation tool Memento, History List Searching method, and Search Engine method, the proposed WebPagePrev delivers the best re-finding quality in finding rate (92.10%), average F1-measure (0.4318) and average rank error (0.3145). (2) Our dynamic management of context and content memories including decay and reinforcement strategy can mimic users' retrieval and recall mechanism. With relevance feedback, the finding rate of WebPagePrev increases by 9.82%, average F1-measure increases by 47.09%, and average rank error decreases by 19.44% compared to stable memory management strategy. Among time, location, and activity context factors in WebPagePrev, activity is the best recall cue, and context+content based re-finding delivers the best performance, compared to context based re-finding and content based re-finding.

IndexTerms - Web revisitation, access context, page content, relevance feedback .

I. 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 revisitation is to re-find the previously viewed web pages, not only the page url, but also the page snapshot at that access timestamp [1]. 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%. 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 [5]. Thus, when a user's web revisitation 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.) [6], [7]. Inspired by the psychological findings, this paper 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.

The Basic architecture in our work is follows

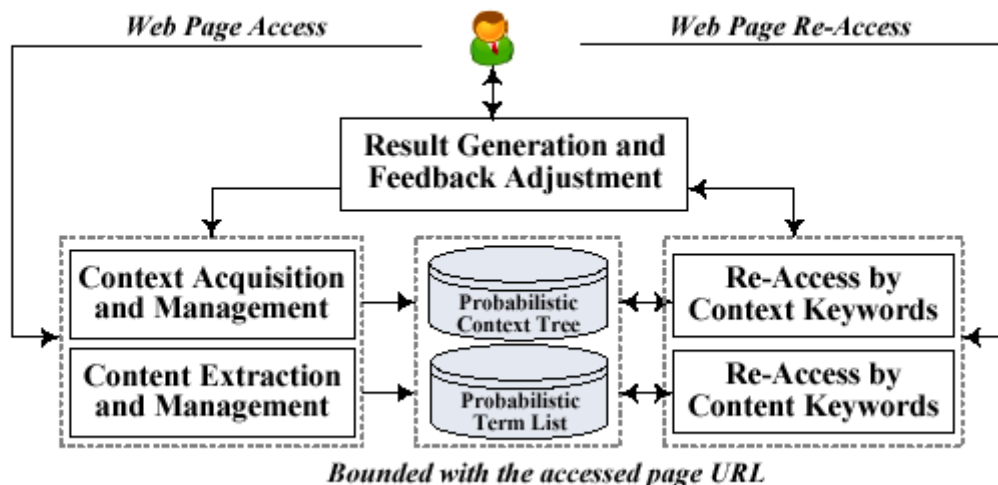


Fig 1- The Personal Web Revisitation Framework.

METHODOLOGY:

Our personal web revisitation framework with relevance feedback. It consists of two main phases.

(1) 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. Meanwhile, the content extraction and management module performs the unigram-based extraction from the displayed page segments and obtains a list of probabilistic content terms. The probabilities of acquired context instances and extracted content terms reflect how likely the user will refer to them as memory cues to get back to the previously focused page. (2) Web revisitation. 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. The relevance feedback mechanism dynamically tunes influential parameters (including memories' decay rates, page reading time threshold, interleaved window size threshold, weight vectors in computing the association and impression scores), which are critical to the construction and management of context and content memories for personal web revisitation. The main contributions of our paper thus lie in the following three aspects:

- We present a personal web revisitation 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.
- Dynamic tuning strategies to tailor to individual's memorization strength and recall habits based on relevance feedback (e.g., weight preference calculation, decay rate adjustment, etc.) are developed for performance improvement.
- We evaluate the effectiveness of the proposed technique WebPagePrev, and report the findings (e.g., the importance of context and content factors) in web revisitation through a 6-month user study with 21 participants.

The rest of the paper is organized as follows. In Section 2, we review closely related work. Then we address the acquisition and management of user's previous access context and content-related information in Section 3, and describe our personal web revisitation approach in Section 4. A relevance feedback mechanism is detailed in Section 5. We evaluate the performance from a 6-month user study in Section 6, and discuss further issues in Section 7. Finally, Section 8 concludes the paper.

CONTEXT ACQUISITION AND MANAGEMENT MODULE

Three kinds of user's access context, i.e., access time, access location, and concurrent activities, are captured. While access time is determinate, access location can be derived from the IP address of user's computing device. By calling the public IP localization API, we can map the IP address (e.g., "166.111.71.131") to a region (e.g., "Beijing, Tsinghua University"). In order to get a high-precision location, we further build an IP region geocoding database, which could translate a static IP address to a concrete place like "Lab Building, Room 216". If the user's GPS information is available, a public GPS localization application could also help localize the user to a Point of Interest (POI) in the region. 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 either a web page, a word file, or a chatting program window, etc., during user's interaction with the computer. Once a user visits a web page longer than a threshold t_c , computer programs that run interleaving with the current web access program for over t_c time are taken as the associated computer programs (i.e., context activities).

CONTENT EXTRACTION AND MANAGEMENT MODULE

Apart from access context, 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 only consider the page segments shown on the screen. There are many term weighting schemes in the information retrieval field. The most generic one is to calculate term frequency-inverse document frequency (tf-idf) [36]. For personalized web revisitation, merely counting the occurrence of a term in the presented page segment is not enough. Also, user's web page browsing behaviors (e.g. visitation time length and highlighting or not), as well as page's subject headings, are counted as user's impression and potential interest indicators for later recall. In a similar manner as access context, we bind an impression score to each extracted content termed, showing how likely the user will refer to it for recall based on the four normalized features.

RESULTS

- (1) Performance Comparison with Existing Approaches.
- (2) User History of Context keywords and Content keywords
- (3) Analysis of used words.

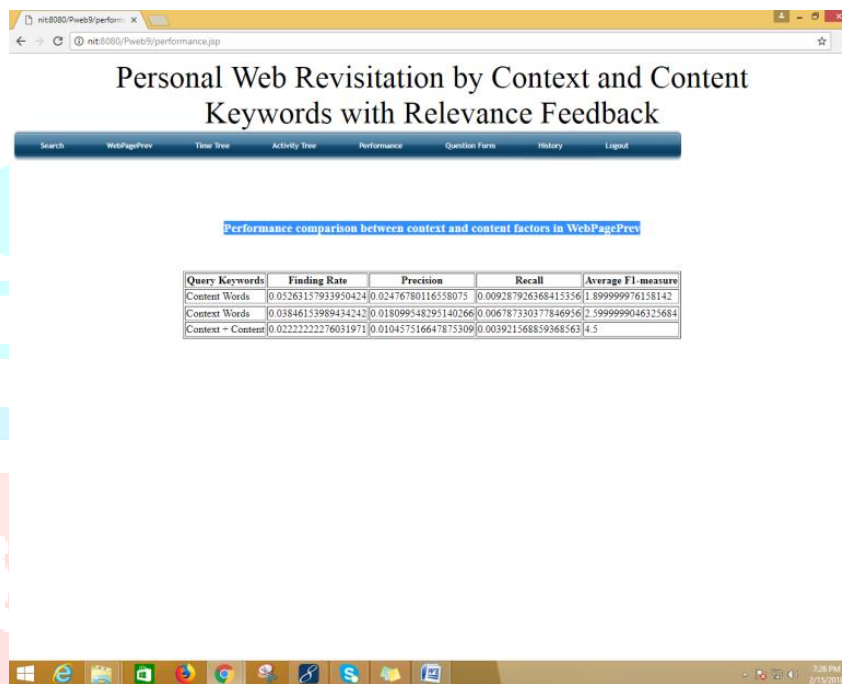


Fig 2: Performance Comparison

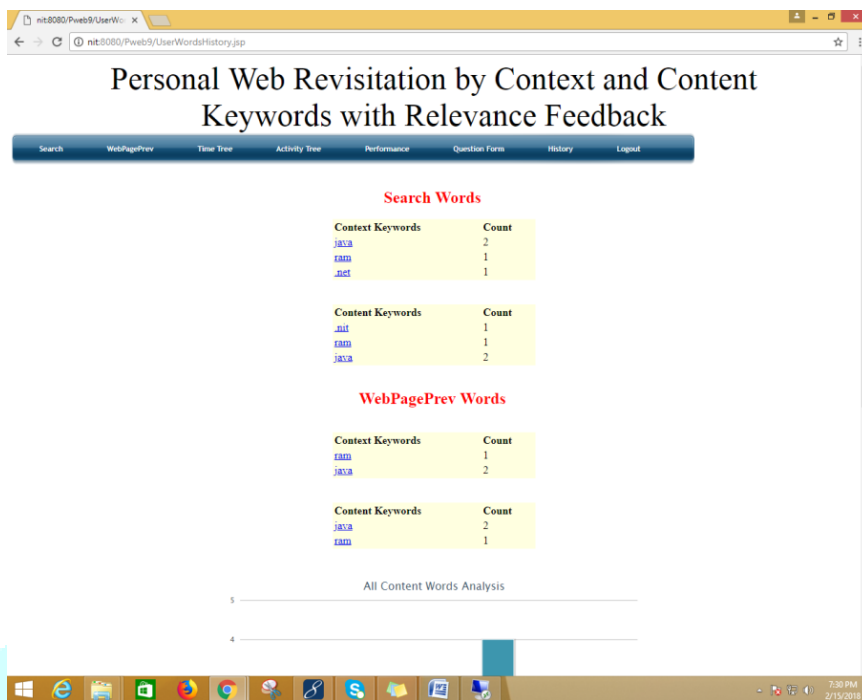


Fig 3: User History

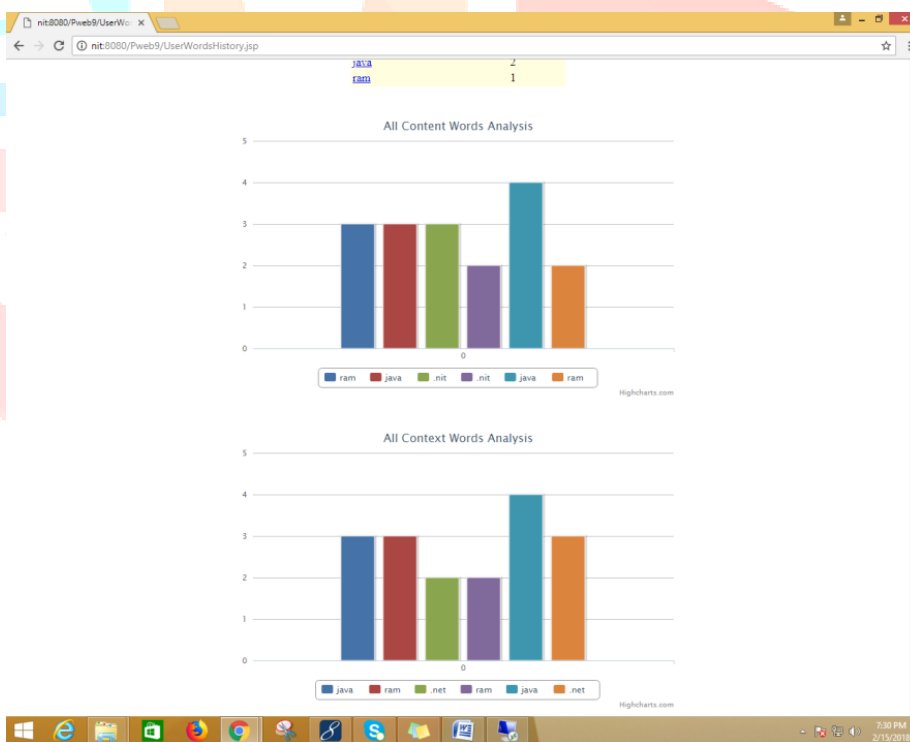


Fig 4: Analysis Of Used Words

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 users' revisitation, 2) extending the technique to support users' ambiguous re-finding requests, and 3) incorporating social context factors in information re-finding.

BIBLIOGRAPHY

- [1] 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.
- [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] J. Teevan, E. Adar, R. Jones, and M. Potts. Information re-retrieval: repeat queries in yahoo's logs. In *SIGIR*, pages 151–158, 2007.
- [4] 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.
- [5] L. C. Wiggs, J. Weisberg, and A. Martin. Neural correlates of semantic and episodic memory retrieval. *Neuropsychologia*, pages 103–118, 1999.
- [6] M. Lamming and M. Flynn. "forget-me-not": intimate computing in support of human memory. In *FRIEND21 Intl. Symposium on Next Generation Human Interface*, 1994.
- [7] E. Tulving. What is episodic memory? *Current Directions in Psychological Science*, 2(3):67–70, 1993.
- [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] J. Hailpern, N. Jitkoff, A. Warr, K. Karahalios, R. Seseck, and N. Shkrob. Youpivot: improving recall with contextual search. In *CHI*, pages 1521–1530, 2011.
- [10] 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.
- [11] T. Deng, L. Zhao, and L. Feng. Enhancing web revisitation by contextual keywords. In *ICWE*, pages 323–337, 2013.
- [12] H. Takano and T. Winograd. Dynamic bookmarks for the WWW. In *HYPertext*, pages 297–298, 1998.

