Personal Web Revisitation by Context and Content Keywords with Relevance Feedback

Mrs. Hajera Begum , Ms. Azzah Arshad , Ms. Gulnar Shagufta , Mr. Mir Mustafa Ali,

[1]Dept. of IT, Deccan College of Engineering and Technology, Hyderabad.

[2]Dept. of IT, Deccan College of Engineering and Technology, Hyderabad

[3]Dept. of IT, Deccan College of Engineering and Technology, Hyderabad

[4]Assistant Professor, Dept. of IT, Deccan College of Engineering and Technology, Hyderabad

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 WebPagePrevthrough 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 WebPagePrevdelivers 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 WebPagePrevincreases 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.*

Keywords: Web revisitation, access context, page content, relevance feedback.

1. Introduction

The rapid development of Multimedia Social Networks (MSNs) causes the tremendous growth of users and digital contents. It's also convenient for users to access digital contents in MSNs with a large-scale video dataset [1]. Meanwhile, the interaction between user and user, user and system increases. Therefore, providing users with timely and rapidly personalized services considering the complex interaction [2] is now a challenge in the study of multimedia social networks. Generally speaking, multimedia computing can be decomposed into three different stages, from datacentric multimedia compression, content-centric multimedia communication

and content analysis, to user-centric social media analysis till today, including user trust modeling [3, 4], propagation paths mining [5, 6] and digital right sharing [7], and digital forensics[8-10]. However, understanding and predicting what multimedia content users' real needs in different situations and contexts have not been well studied [11].

Context-Aware (CA) [12-15] was first proposed by Schilit et al in 1994. They defined context as the set of location, people nearby, objects, and the changes of the objects. Prof. Carl K. Chang [16] proposed the Situ theory by combining the service environment with situation awareness to handle the dynamic update or development of service at run time. Therefore the service can meet the changing needs of users and provide users with personalized service. In order to adapt to the dynamic service environment and make a timely respond to the feedback of service environment, social media services increasingly require situation awareness. In social media networks, the human being is a complex and open system. The individual's intention can change at any time, which also causes a change in the user's needs. Moreover, the user's context and behavior are dynamic. Some studies show that the characteristics of the dynamic change will have different effects in a user's potential needs [17, 18]. A user's intention can be reflected through the acquiring attributes of the user's situation awareness and feedback on resources. The system can formulate a timely personalized service for the user based on user's intention, which will increase the user's service experience. In social media networks, the user has different roles in different groups. The different identifications that the user has may cause the user's intention to change. The change of intention reflects the change in user's behavior. The Situ theory [16, 19] does not fully meet the analysis of the intention of users with different identities in the social media environment. This paper's motivation is to analyze the user's intention sequence mode(s) in social media networks. The major contributions of this paper are two folds. One is to enrich and extend the Situ theory outreaching for social domain, that is the social media ecosystem, through newly and comprehensively considering user's changeable identity (including role and group), and the other is to propose a novel algorithm for users' behavior pattern analysis and mining. The important vision of the work is to further predict users' more and deeper intention and mental based on a large volume of previous actions. The remaining parts of this paper are as follows: Section 2 shows the progress in related studies; the next section shows the extension of the Situ framework; Section 4 introduces the intention serialization algorithm; the experiment and its results of the serialization algorithm are in detail presented in Section 5; and finally conclusions are drawn.

2. Literature Survey:

Beyond the usual suspects context-aware revisitation support

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

International Journal of Research

Available at https://journals.pen2print.org/index.php/ijr/

e-ISSN: 2348-6848
p-ISSN: 2348-795X
Volume 06 Issue 03
March 2019

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.

3. System Analysis:

Existing System:

- Apart from back=forward buttons, manually or 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.

Disadvantages:

- 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.

Proposed System:

- Personal web revisitation framework with relevance feedback. It consists of two main phases.
- 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.

Advantages:

- 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 WebPagePrevduring web re-visitation phase, s/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.

International Journal of Research

Available at https://journals.pen2print.org/index.php/ijr/

e-ISSN: 2348-6848
p-ISSN: 2348-795X
Volume 06 Issue 03
March 2019

Modules:

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-Δ;we+Δ], denoted as c[ws-Δ;we+Δ] = (c:title; c:dur; c:freq), if and only if (c:dur> _c).

Content Extraction and Management:
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.
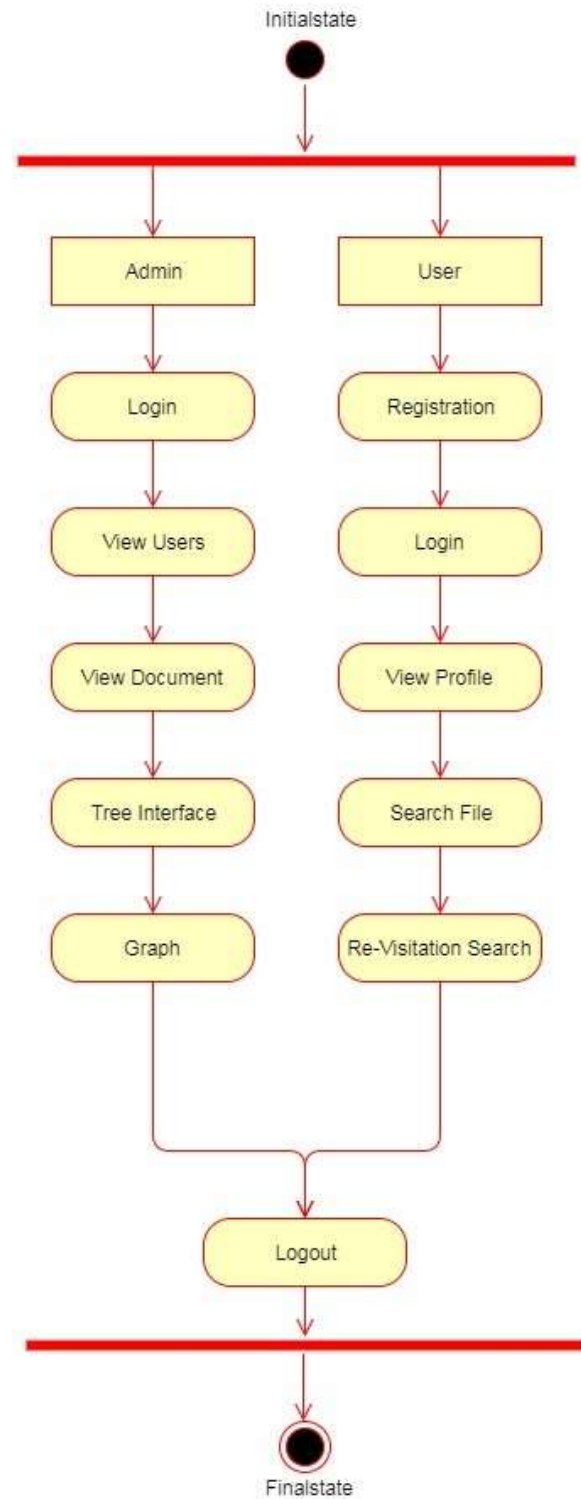
4. System Design:
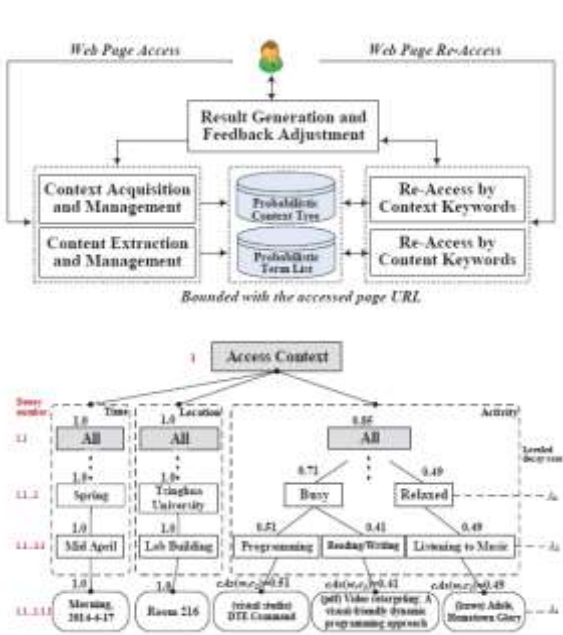
4.1 System Architecture

Image 4.2 Activity Diagram

5. Output Results:

Fig 5.1: Home Page

Fig 5.2: Admin Login

Fig 5.3: Admin Home

Fig 5.5: Upload Document

Fig 5.6: Tree Interface

Fig 5.7: Time Tree

Fig 5.8: Place Tree

Fig 5.9: Activity Tree

Fig 5.10: View Documents

Fig 5.11: Graph

Fig 5.12: web revisitation Interface

6. 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.

.

References

[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 historysystems. 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 surveyof existing approaches and directions. Foundations and Trends inHCI, 2(3):173–278, 2009.

[5] L. C. Wiggs, J. Weisberg, and A. Martin. Neural correlates ofsemantic and episodic memory retrieval. Neuropsychologia, pages103–118, 1999.

[6] M. Lamming and M. Flynn. "forget-me-not": intimate computingin support of human memory. In FRIEND21 Intl. Symposium onNext Generation Human Interface, 1994.

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

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

[9] J. Hailpern, N. Jitkoff, A. Warr, K. Karahalios, R. Sesek, andN. Shkrob. Youpivot: improving recall with contextual search.InCHI, 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 bycontextual keywords. In ICWE, pages 323–337, 2013.

[12] H. Takano and T. Winograd.Dynamic bookmarks for the WWW.In HYPERTEXT, pages 297–298, 1998.

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

[14] J. A. Gamez, J. L. Mateo, and J. M. Puerta.Improving revisitationbrowsers capability by using a dynamic bookmarks personaltoolbar. In WISE, pages 643–652, 2007.

[15] R. Kawase, G. Papadakis, E. Herder, and W. Nejdl. Beyond theusual suspects: context-aware revisitation support. In HT, pages27–36, 2011.