History Structure for Exploring Desktop Data

Harumi Murakami
Osaka City University
3-3-138, Sugimoto, Sumiyoshi, Osaka,
558-8585 Japan
harumi@media.osaka-cu.ac.jp

ABSTRACT

We present a method of data integration and associative retrieval using a simple information structure called *history structure*, which is constructed from time, keywords, and URI sets. We developed a prototype that generates a user knowledge space from various information usages (e.g., web browsing, mail, twitter, diaries, and purchases) and helps users explore their desktop data.

Categories and Subject Descriptors

H.3.3 [**Information Storage and Retrieval**]: Information Search and Retrieval – *Retrieval models.*

General Terms

Design, Experimentation

Keywords

History structure, associative retrieval, knowledge space, externalized-memory model, integration

1. INTRODUCTION

We are surrounded by various kinds of information. Interest in Personal Information Management has increased in recent years [1], partly as a reaction to information overload, which is becoming a real problem in our daily lives. Much research has presented ideas for integrating such information [2, 3].

In this paper, we present a new method of data integration and associative retrieval by using a simple information structure called *history structure*, which is constructed from time, keywords, and URI sets. Our prototype generates a user knowledge space from the user's various information usages (e.g., web browsing, mail, twitter, diaries, and purchases) and helps users explore personal desktop data.

History structure is simply generated from existing information sources. Our approach resembles tagging; however, the manual tagging of personal information is time-consuming. We aim to automatically generate history structure.

First, various information is stored in the history structure. Next, the system creates knowledge spaces for individuals. Figure 1 displays an overview of our approach.

Our approach, which is based on an externalized-memory model inspired by a human memory model, is presented in Section 2.
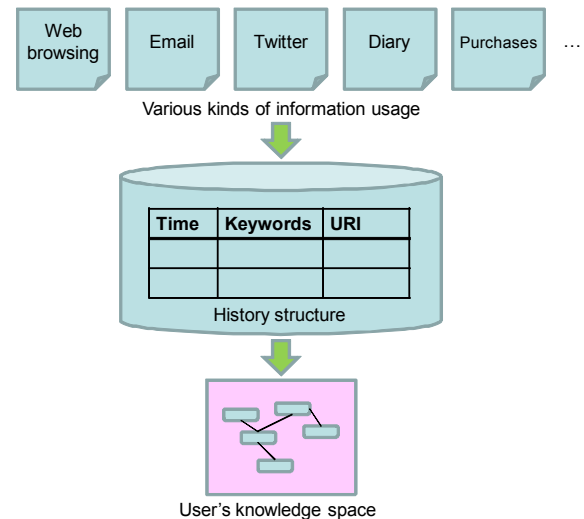
Figure 1. Overview

2. EXTERNALIZED-MEMORY MODEL

Externalized-memory is a concept that virtually externalizes and stores the contents of human working memory on computers. We capture information processed during human cognitive processes by working memory. History structure is a kind of externalized memory.

Figure 2 shows an overview of our proposed model. Such information processed by the user as browsing the web is accumulated into the user's externalized memory (history structure) by a sensory memory and working memory. The knowledge space is a simulated visualization of the history structure like semantic networks that are generated from the history structure. When a user selects a keyword in the knowledge-space browser, the system searches through the history structure, recalls related information, and displays it in the knowledge-space browser.

Figure 2. Externalized-memory model

3. ALGORITHMS

3.1 Generating History Structure

The basic algorithms for generating history structure extract time, keywords, and URIs from various information sources. When a URI does not exist, any information can be input by users. The details of the algorithms are different among information sources. The following is an example of generating history structure from a web browsing history: extracting a time from an accessed time, extracting keywords from a title of the accessed web page, and extracting an URI from the accessed URI. Table 1 shows an example of a generated history structure.

Table 1. Example of generated history structure

Time	Keywords	URI	Note
Mon Apr 26 17:12:02 JST 2010	book, digitize, scansnap, S1300, S1500	URI of accessed web page	web browsing
Mon Apr 26 17:39:01 JST 2010	experiment, questionnaire, scansnap, maximal number	URI of twitter logs	twitter

3.2 Generating Knowledge Space

The basic algorithms for generating the knowledge space connect the keywords that co-occurred in a history structure. For example, in Table 1, *scansnap* (a scanner's name) is connected with *book*, *digitize*, *S1300*, *S1500*, *experiment*, *questionnaire*, and *maximal number*. We assume that a cognitive process does occur, as shown in Figure 2, and consider that memory is activated based on the frequency of the stimuli in the working memory. Thus, we change the size of the nodes based on the activity level. Keywords that occurred frequently in the history structure and that were frequently accessed by the operation are displayed as larger.

The user or the system can generate a knowledge space. The user can set the information sources, input or select keywords, and explore co-occurred keywords. The system can generate a knowledge space by using several algorithms. For example, the system can generate a *summary* of a designated period: a day, a month, a year. Frequently occurring keywords are displayed as seeds. The order of the information sources can be set prior to exploration.
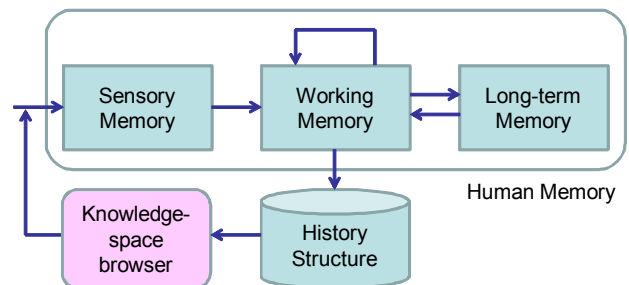
4. CASE STUDY

Figure 3 shows Microsoft windows desktop search results of folders and files modified on 26/Apr/2010. *Questionnaire*, *namecard*, *Kataoka*, *Ou*, and *Senba* are folders. *Kataoka*, *Ou*, and *Senba* are the names of the user's students. Files were: completed questionnaires (pdf files), purchase receipts (pdf files), and records of meetings with students (txt files).

Figure 3. Folders and files searched by desktop search

The knowledge-space browser helps users to recall the day's context.

Figure 4 shows an example screen generated by web browsing and book purchase histories, diaries, emails, and twitter logs of a day's summary. Different colors correspond to different information sources: aquamarine for Web browsing, purple for book purchases, gray for diary, yellow for email, and pink for twitter. When the same keywords exist in different information sources, the colors of the keywords change based on the last operation.

Figure 4. Summary of a day presented in the knowledge-space browser

The user went to a conference called JSAS on April 24 and 25 in Tokyo and gave a lecture on computer literacy on April 26.

Pasmo, *scansnap*, *Suica*, and *archive* were the four most frequent keywords in the web browsing history. The user searched for information about how to use scansnap on the web. *Questionnaire*, *experiment*, and *maximal number* are displayed from twitter logs. *Computer Literacy* is the name of the lecture in which the user conducted a questionnaire survey. He wanted to scan the completed questionnaires. After coming back from the conference, the user searched for information about archives and bought some books. *Pasmo* is a contactless smart card that can be used for train fares in the Tokyo area. He searched for information on how to buy it for his next business trip to Tokyo.

Thus users can explore the desktop data themselves (e.g., Web browsing history), recall the contexts of files (e.g. he divided questionnaires due to the scanner's page limitations), and recall future plans (e.g., buying a Pasmo for his next trip).

5. SUMMARY

We presented a method of data integration and associative retrieval using a simple information structure called history structure, which is constructed from time, keywords, and URI sets. We developed a prototype that generates a user knowledge space from various information usages and helps users explore desktop data.

REFERENCES

[1] Jones, W. 2007. Personal information management, ARIST. 41, 453-504.

[2] Gemmel, J., Bell, G., Lueder, R., Drucker, S. 2002. Mylifebits: Fullfilling the memex vision, Proc. Tenth ACM Multimedia. 235-238.

[3] Cutrell, E., Dumais, S., and Teevan, J. 2006. Searching to eliminate personal information management, CACM. 49, 1, 58-64.