

Creating User's Knowledge Space from Various Information Usages to Support Human Recollection

Harumi Murakami¹, Kenta Mitsuhashi², and Keita Senba¹

¹ Graduate School for Creative Cities, Osaka City University,
3-3-138, Sugimoto, Sumiyoshi, Osaka 558-8585 Japan
harumi@media.osaka-cu.ac.jp
<http://murakami.media.osaka-cu.ac.jp/>

² OGIS-RI Co.,Ltd.,
3-Minami 2-37, Chiyozaeki, Nishi, Osaka 550-0023 Japan

Abstract. To support human recollection, we present a data integration method using a simple information structure called a history structure, which is constructed from time, keywords, and URI sets. We also present algorithms that generate history structures from such information usages as Web searches, twitter, e-mails, calendars, and book purchases and create a user knowledge space. Based on our approach, we developed a system called a knowledge-space browser and evaluated whether it can help users recall a particular day by summarizing that day's history structure. Experimental results reveal the usefulness of our approach and our implemented system.

Keywords: knowledge space, history structure, human recollection, web search, twitter, e-mail, calendar.

1 Introduction

Memory is crucial for various activities. We think and feel using memory. We need to recall past memories of particular periods. For example, we may have to write progress reports about what we have done on particular days or weeks. We may have to plan anniversaries and recall what we did last year or in previous years. Or we might simply want to reminisce about the day when we saw our spouse for the first time. We aim to help users recall the past from particular periods.

We propose an approach that gathers pieces of the past memory of a day and visualizes them as a knowledge space to help recollection. We present a method of data integration using a simple information structure called a history structure, which is constructed from time, keywords, and URI sets. A history structure is simply generated from existing information sources. In this paper, we present a method that generates history structures from such information usages as web searches, twitter, e-mails, calendars, and book purchases, and create a user knowledge space from them.

Below, we explain our proposed algorithms in Sections 2 and 3. The implementation and examples of the knowledge-space browser are described in Section 3. Our experiments are described in Section 4. We discuss the significance of our research in Section 5.

2 Generating History Structure

2.1 Gathering Information from Various Information Usages

We gather texts that have time information and express user thoughts or activities to help them recall their memories.

Web Search. We believe that Web search histories often express user interests and are thus related to their thoughts or activities. We use Web search histories (Google queries and search results) to gather information that expresses user interests. For queries, we extract a time of Web search, query, and query URI by dividing the query using spaces and generate keywords. For search results, we extract a time of browsing, browsed page's title, and page's URI and generate keywords from the title of the browsed page using the algorithm described in Section 2.2.

Twitter. Since tweets generally express the user's thoughts or activities, we use all of them except for those starting with @ because they are mainly discourse and official RTs (Retweets) because they are mainly other's opinions. We extract a tweet time, tweet, and its URI and generate keywords from the tweet using the algorithm described in Section 2.2.

E-mail. E-mails sometimes express the user's thoughts or activities. We use the receivers and the subjects of the sent messages because sent messages often express thoughts or activities. We do not use received messages, because most are direct mails that are not related to the user activities. We extract the time e-mail was sent, its receivers and subject, and its URI. For receivers, we simply extract names and addresses. For subjects, we generate keywords using the algorithm described in Section 2.2.

Calendar. Calendars or diaries are obviously useful sources for user activities. We extract a "start time of event, event title, and event URI by simply dividing the event title into keywords using spaces and adding the original event title as keywords. We believe that the original event title is meaningful for users.

Book Purchases. Product purchases sometimes help users remember thoughts or activities. In this research, we focus on book purchases because book contents are probably related to user knowledge. We extract a time of order/purchase, title, and ISBN of the book. We generate keywords from the book title using the algorithm described in Section 2.2.

2.2 Generating Keyword Algorithm

We developed a generating keyword algorithm that creates a set of keywords from such texts as the titles of browsed Web pages, tweets, the subjects of sent e-mails, event titles in calendars, and the titles of books purchased.

Our algorithm extracts noun phrases, adjectives, verbs, and non-Japanese terms with MeCab, a Japanese morphological analysis tool [1], which outputs non-Japanese terms as nouns. See Fig. 1.

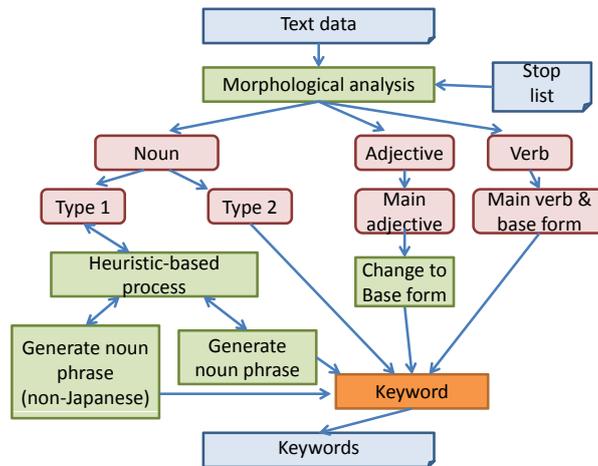


Fig. 1. Algorithm of generating keywords

When a term is a noun, a common noun, a proper noun, a noun verbal, a noun suffix, or a noun number (type 1), it is repeatedly concatenated with previous terms as a non-Japanese keyword or as a Japanese keyword using heuristics. When the noun is a noun adverbial or a noun adjective base (type 2), it directly becomes a keyword. When a term is an adjective and its type is a main adjective, the base form becomes a keyword. When a term is a verb, its type is a main verb, and it is equal to the base form, the base form of the term becomes a keyword. The detailed algorithm is described in [2].

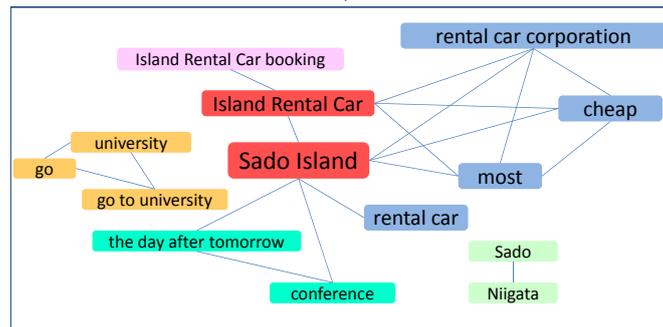
3 Generating Knowledge Space

3.1 Generating Knowledge-Space Algorithm

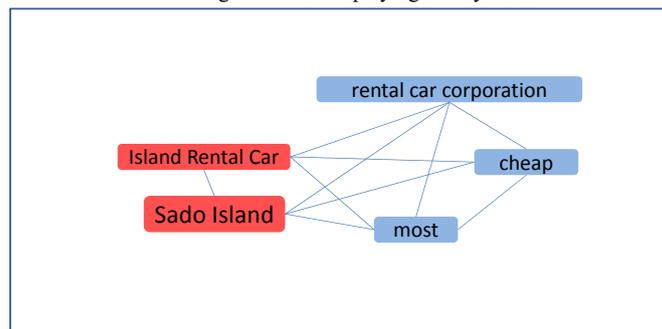
The idea of knowledge space is based on semantic networks that represent semantic memory in cognitive psychology [4]. We believe that displaying a user's knowledge space like semantic networks helps user recollection.

History Structure

Time	Keywords	URI	Note
Mon 13 00:00:00 JST 2010	Go to university, go, university	http://www.google.com/calendar/...	Calendar
Mon 13 00:28:00 JST 2010	Niigata, Sado	ISBN-10:4398153187	Book purchase
Mon 13 15:19:14 JST 2010	Sado island, rental car	http://www.google.com/search?q...	Web browsing (query)
Mon 13 16:06:18 JST 2010	Island Rental Car, Sado island, most, cheap, rental car corporation	http://park2wakwak.com/...	Web browsing (title of browsed page)
Mon 13 17:14:33 JST 2010	The day after tomorrow, conference, Sado island	http://twitter.com/...	Twitter
Mon 13 19:07:55 JST 2010	Island Rental Car, Sado island, most, cheap, rental car corporation	http://park2.wakwak.com/...	Web browsing (title of browsed page)
Mon 13 19:09:35 JST 2010	Island Rental Car, Island Rental Car booking	https://mail.google.com/mail...	Mail



Algorithm 1: Displaying all keywords



Algorithm 2: Displaying keywords that occur more than once and have one or more relations.

Fig. 2. Algorithm of generating knowledge space

The basic algorithms for generating knowledge space connect the keywords that co-occurred in the history structures. Clusters generated by natural connections help users recall their past.

The basic algorithm is as follows:

- Connect keywords that co-occurred in history structures.
- Keywords that occurred frequently in the history structure are displayed as larger.
- Different colors correspond to different information sources: light blue for Web searches, aquamarine for twitter, pink for e-mails, yellow for calendars, and purple for book purchases. These five colors for each information source were designed through a preliminary investigation.
- Keywords that occurred in multiple information sources are emphasized in red.

We designed and implemented two algorithms: displaying all keywords (visualization algorithm 1) and displaying keywords that occur more than once in history structures and have one or more relations (visualization algorithm 2). Fig. 2 shows an overview of the two algorithms. This example is part of the next section.

3.2 Knowledge-Space Browser

We implemented a knowledge-space browser that is comprised of four parts: (a) a knowledge space display that shows the user’s knowledge space, (b) a history structure display that lists the user’s history structures, (c) an operation display on which the user manages the knowledge space, and (d) a selected keyword display that lists the history structures on which the selected keyword is included. For the knowledge space, users can change either visualization algorithm 1 or 2 at anytime.

Fig. 3 is an example screen of a user’s knowledge space generated by visualization algorithm 1 on 13/Sep/2010. He prepared to go to a conference that will be held on 15/Sep/2010 on Sado Island. See Fig. 2. His calendar shows his plan to go to the university to prepare his presentation. He bought a guide book of Sado Island in Niigata prefecture. He searched “Sado Island, rental car” and browsed a page of a “rental car corporation named Island Rental Car” that is denoted as the cheapest on the island. He tweeted “I will go to Sado Island for a conference the day after tomorrow.” He booked a rental car at Island Rental Car by e-mail. The biggest cluster in the knowledge space contains *Sado Island*, *Island Rental Car*, *Sado*, *conference*, in red. This cluster shows the main activities of the user for that day. The biggest (i.e. the most frequent) keyword is *trial version* in a different cluster. Such keywords as *Photoshop* and *download* are included in this cluster, which shows that the user was preparing a presentation for the conference using photoshop.

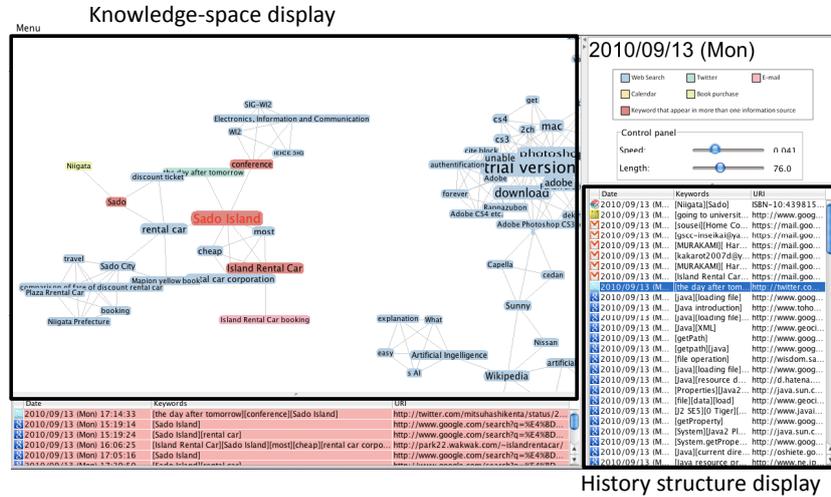


Fig. 3. Example of knowledge-space browser

4 Experiment

4.1 Experiment 1

We evaluated the usefulness of our algorithm that extracts keywords.

Method. The subjects were 10 male, computer and information science graduate students aged 22-24.

We generated history structures from five information usages (Web searches, twitter, e-mails, calendars, and book purchases) using four methods. The comparative method (a) extracts only nouns, (b) extracts nouns and concatenates them using TermExtract [3], which is a Japanese noun phrase extraction tool that orders the extracted Japanese phrases in importance, or (c) only extracts noun phrases by our proposed method. Our method (d) extracts concatenated noun phrases, adjectives, and verbs.

The subjects evaluated whether the generated keywords are appropriate among five information usages and four methods by five values (5: very appropriate; 4: appropriate; 3: OK; 2: not very appropriate; 1: inappropriate). Each subject evaluated 10 history structures * 5 information usages * 4 methods: 200 history structures.

Results and Discussion. Table 1 shows the average values in Experiment 1. Our method was the best among all methods. These results suggest the usefulness of our algorithm for generating keywords.

Except for calendars, there were significant differences among the four methods ($p < 0.01$) by a one-way anova. There also were significant differences between (a) and (c), (a) and (d), and (b) and (d) (each $p < 0.01$).

From the comparison of (a) and (c), we found that the noun phrases extracted by our method are better than nouns; from the comparison of (b) and (c), our method is better than TermExtract for extracting noun phrases from our data; from the comparison of (c) and (d), adding adjectives and verbs is better than only using noun phrases. We found that noun phrases, adjectives, and verbs must be extracted as keywords rather than only nouns or noun phrases, and appearance is better than importance as the order of keywords.

Table 1. Result of Experiment 1

	(a)	(b)	(c)	(d) our method
Web searches	2.71	2.81	4.02	4.05
Twitter	2.56	2.43	3.56	3.91
E-Mails	2.85	3.26	4.10	4.15
Calendars	3.50	3.54	3.89	4.00
Book purchases	2.78	2.69	3.88	4.18

4.2 Experiment 2

We investigated the features of three visualization methods: a list of history structures, visualization algorithm 1, and visualization algorithm 2 when the users can recall memories of a day last week.

Method. The subjects were six male, computer and information science graduate students aged 22-24.

We gathered and generated history structures for each subject from five information usages: Web searches, twitter, e-mails, calendars, and book purchases. Three sets of history structures of three days from about one week before were assigned to three visualization methods.

We asked the subjects to use the history structures display (without showing other display parts) and explain what they recalled about the day in interviews. Next, we repeated the above processes using a knowledge-space display (visualization 1) and a knowledge-space display (visualization 2). After each experiment, we asked subjects to answer the pros and cons for each visualization method. Finally, the subjects ranked their answers for the three methods to these two questions: “Was this method useful to recall the past?” (Q1) and “Was this method useful to summarize the day?” (Q2).

Results and Discussion. For question 1, all subjects answered that using history structures was the most useful way to recall the past. Visualizations 1 and 2 were equal; three subjects ranked visualization 1 and three others ranked visualization 2 second. For question 2, three subjects answered that visualization 2 was the most useful, two answered visualization 1, and one answered the history structures. We think that a list of history structures is useful to recall the past and that knowledge space is useful to summarize one day. All subjects recommended displaying the history structures and the knowledge space together.

We combined user comments for the pros and cons that express the same meanings. For the list of history structures, “The keywords are ordered chronologically” (4: pros) and “Viewing was difficult” (2: cons). For visualization 1, “I can get a day’s overview by looking at the clusters” (6: pros), “One cluster corresponds to a topic” (3: pros), and “I can grasp a day intuitively” (3: pros) and “Similar clusters need to be combined” (3: cons). For a visualization 2, “The information is well-organized, and there are many keywords that help recall” (4: pros), “Visualization 2 is better arranged than visualization 1” (2: pros), and “This method doesn’t provide enough information” (2: cons).

We think that a list of history structure is beneficial for sequentially viewing chunks of memory, and a knowledge space is good to see grasp their overview.

4.3 Experiment 3

We evaluated the usefulness of our system to support human recollection when the users can recall the memories of a day of the previous week.

Method. The subjects were the same six male students from Experiment 2, which was conducted immediately before Experiment 3.

For Experiment 3, one set of the history structures of five information usages of one day gathered in Experiment 2 was used. The subjects used the knowledge-space browser and explained what they recalled about the day in interviews.

The subjects also evaluated the usefulness of the system by answering six questions on a five-point scale (5: I completely agree, 4: I agree, 3: OK, 2: I don’t agree, 1: I completely disagree).

Table 2. Results of Experiment 3

	Question	Mean
Q1	Could you recall your past faster using the system?	4.8
Q2	Could you recall your past easier using the system?	5.0
Q3	Did the system help you recall something that you had forgotten?	4.7
Q4	Do you think that the system is useful?	4.2
Q5	Do you want to use the system again?	4.5
Q6	Do you think that you can recall your past easier using the system?	4.7

Results and Discussion. The evaluation results are shown in Table 2. The subjects thought the system was useful to support human recollection. In a knowledge-space browser, a user can choose the visualization algorithm. In this experiment, three subjects selected visualization 1 and three others selected visualization 2. Two of the latter subjects changed their choices from visualization 2 to 1 because they thought visualization 2 lacked sufficient information.

The overall results suggest that the combination of display of history structure and knowledge space is useful to support human recollection.

5 Related Work and Discussion

This research is a part of a system called *Memory-Organizer* that helps users construct “externalized memory” [5][6]. We previously proposed knowledge space to help users explore their knowledge spaces created from Web browsing history [7]. We extended this idea and proposed history structures to integrate various information sources [8]. The differences between the previous research and this paper are that we presented new algorithms for generating history structures for five information usages and two visualization algorithms for generating knowledge space. We also developed a new version of the knowledge-space browser and conducted experiments to determine what users can recall on a particular day. We found that the overall usefulness of our approach was good and that the system helped user recollection.

Much research has presented ideas for integrating such information in the light of Personal Information Management (PIM) [9], to overcome information overload [10][11]. 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. Our research and PIM are related to a research topic called lifelogs that capture a person’s entire lives. Mylifebits [12] is well-known not only as PIM but also as lifelog research. In typical lifelog research, lifeloggers wear computers to capture what they have seen and heard. Much research has been conducted in the context of multimedia or image processing. Our research resembles text-based lifelog research. Our approach is simple and does not need special devices to capture information from the real world. Although much research extracts information from each information source such as web searches, twitter, e-mails, and calendars, little research extracts and integrates information from those sources.

The following are our paper’s main contributions. First, we presented a data integration method using a simple information structure called a history structure that is constructed from time, keywords, and URI sets. We also developed algorithms for generating history structures from various information sources and visualizing user knowledge spaces from them. Second, we developed a knowledge-space browser based on this approach and combined history structure and knowledge-space displays to help human recollection. Third, we evaluated whether our system helps users recall a particular day by summarizing its history structures. The experimental results revealed the usefulness of our approach and the implemented system.

Future work is listed below. First, although we analyzed the difference between visualization algorithms 1 and 2, we did not find out which is better in which user situation. Second, we need to improve our algorithms for generating keywords and visualization. Third, we need to examine the system for different periods, such as a day or a week within the past month.

6 Conclusions

We presented a data integration method using a simple information structure called history structure that is constructed from time, keywords, and URI sets and developed the following: (a) heuristic based keyword generation algorithms that extract noun phrases, adjectives, and verbs from various information sources; and (b) two visualization algorithms that create user knowledge spaces from history structures. We developed a system based on our approach to support human recollection and evaluated whether it can help users recall a particular day by summarizing that day's history structure. The experimental results revealed the usefulness of our approach and the implemented system.

References

1. MeCab: Yet Another Part-of-Speech and Morphological Analyzier, <http://mecab.sourceforge.net/>
2. Mitsuhashi, K.: A Human Recollection Support System by Integrating Diverse Information and Creating Knowledge Space, Master's Thesis, Graduate School for Creative Cities, Osaka City University (2011) (in Japanese)
3. <http://gensen.dl.itc.u-tokyo.ac.jp/termextract.html>
4. Collins, A.M., Loftus, E.F.: A Spreading-Activation Theory of Semantic Processing. *Psychological Review* 82(6), 407–428 (1975)
5. Murakami, H., Hirata, T.: Information Acquisition and Reorganization from the WWW by using Memory-Organizer. *Bulletin of Osaka City University Media Center* 3, 9–14 (2002)
6. Murakami, H., Hirata, T.: A System for Generating User's Chronological Interest Space from Web Browsing History. *International Journal of Knowledge-Based and Intelligent Engineering Systems* 8(3), 149–160 (2004)
7. Murakami, H., Hirata, T.: Recall of Web Browsing History by Generating User's Knowledge Space. *Bulletin of Osaka City University Media Center* 5, 21–26 (2004)
8. Murakami, H.: History Structure for Exploring Desktop Data. In: *Proceedings of the SIGIR 2010 Workshop on Desktop Search (Understanding, Supporting and Evaluating Personal Data Search)*, pp. 25–26 (2010)
9. Jones, W.: Personal Information Management. *ARIST* 41, 453–504 (2007)
10. Dumais, S., Cutrell, D., Cadiz, J.J., Jancke, G., Sarin, R., Robins, D.C.: Stuff I've Seen: A System for Personal Information Retrieval and Re-Use. In: *Proceedings of SIGIR 2003*, pp. 72–79 (2003)
11. Cutrell, E., Dumais, S., Teevan, J.: Searching to Eliminate Personal Information Management. *CACM* 49(1), 58–64 (2006)
12. Gemmel, J., Bell, G., Lueder, R., Drucker, S.: Mylifebits: Fullfilling the Memex Vision. In: *Proceedings of Tenth ACM Multimedia*, pp. 235–238 (2002)