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

[1,*] Harumi Murakami, [2] Kenta Mitsuhashi
[1] *Graduate School for Creative Cities, Osaka City University, harumi@media.osaka-cu.ac.jp*
[2] *OGIS-RI Co., Ltd., kenta.mitsuhashi@gmail.com*
** Corresponding Author*

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, tweets, e-mails, calendars, and book purchases and create a user knowledge space. Our approach is based on an externalized-memory model inspired by a human memory model. 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. We show a case study in which a user can explore a desktop using the knowledge-space browser.

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, tweets, 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 show a case study in which a user can explore a desktop using the knowledge-space browser in Section 5. We discuss the significance of our research in Section 6.

2. Generating history structure

First, various information usages are stored as history structures from which the system creates user knowledge spaces. Figure 1 displays an overview of our approach.

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.

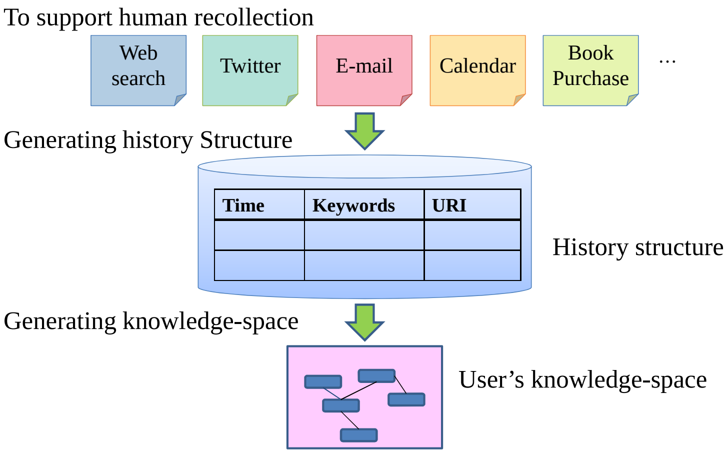


Figure 1. Our approach

2.1.1. 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 2.1.6.

2.1.2. 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 2.1.6.

2.1.3. 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 2.1.6.

2.1.4. 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.

2.1.5. 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 2.1.6.

2.1.6. 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.

The natural information usages of Japanese often contain non-Japanese noun phrases such as English nouns. 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 Figure 2.

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 noun phrase or as a Japanese noun phrase using heuristics. An example of the former is *artificial* and *intelligence*, which are concatenated into *artificial intelligence*, and becomes a keyword; an example of the latter is *jinko* (means artificial) and *chino* (means intelligence), which are concatenated into *jinkochino* (means artificial intelligence) and becomes a keyword.

The following is an example of the part of a heuristic algorithm that outputs a concatenated noun phrase using a space (a non-Japanese keyword). When the noun is a general noun, and (a) the previous noun is a general or a proper noun, or a number, and (b) there is no furigana, which is a Japanese reading aid, for either the previous or the current noun, the series of nouns are judged to be non-Japanese, and the previous noun and the current noun are concatenated with using a space.

When the noun is a noun adverbial or a noun adjective base (type 2), it directly becomes a keyword. For example, *asatte* (which means the day after tomorrow and is judged to be a noun adjective base) becomes a keyword.

When a term is an adjective and its type is a main adjective, the base form becomes a keyword. For example, a candidate is *tanoshiku* (means pleasant) changes to tanoshii (base form) and becomes a keyword. When a candidate 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. For example, *iku* (means go) becomes a keyword.

The detailed algorithm is described in [2]. In what follows, the examples in this paper were translated from Japanese into English for publication.

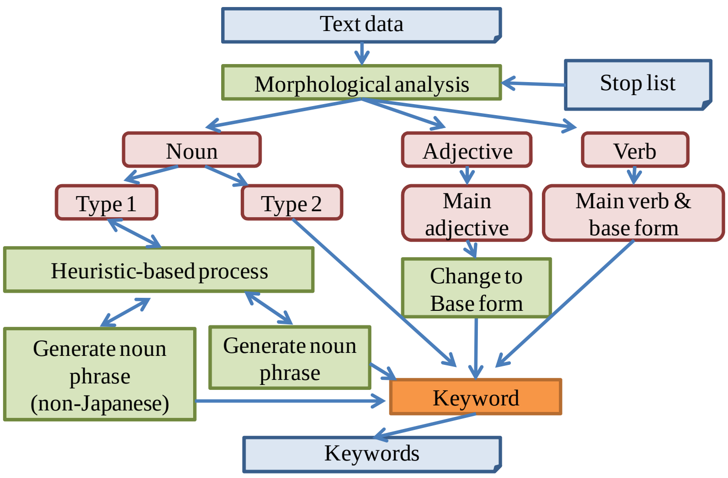


Figure 2. Generating keyword algorithm

3. Creating knowledge space

3.1. Externalized-memory model

Our approach is based on an externalized-memory model inspired by a human 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 3 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 wants to generate a knowledge space, the system searches through the history structure, recalls related information, and displays it in the knowledge-space browser.

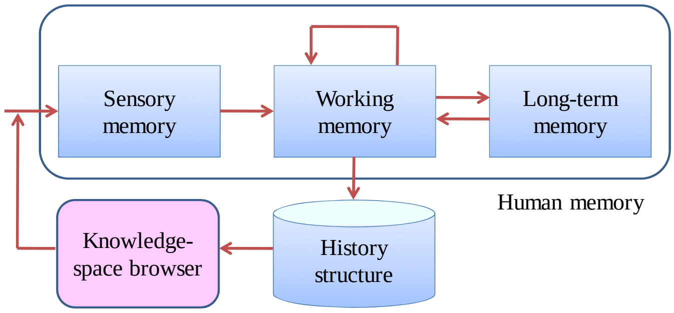


Figure 3. Externalized-memory model

3.2. Generating knowledge-space algorithm

The idea of knowledge space is based on semantic networks that represent semantic memory in cognitive psychology [3]. We believe that displaying a user's knowledge space like semantic networks helps user recollection.

We assume that a cognitive process does occur, as shown in Figure 3, 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 are displayed as larger.

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:

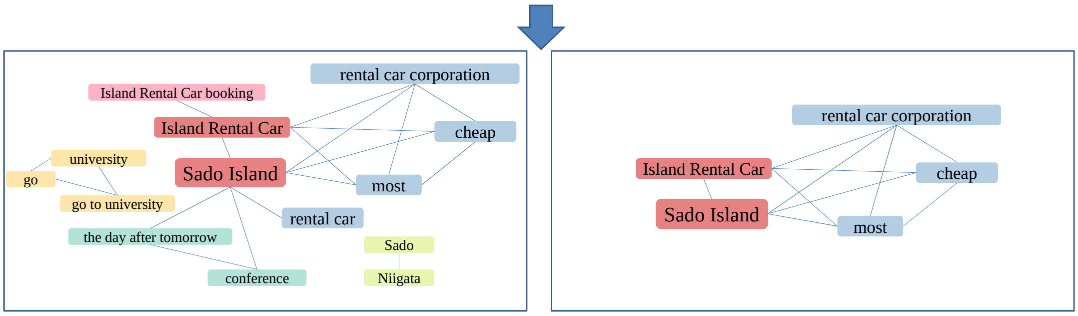
- Connect keywords that co-occurred in history structures.
- Keywords that occurred frequently in the history structure are displayed as larger.
- Different colors correspond to different information usages: light blue for web searches, aquamarine for tweets, pink for e-mails, orange for calendars, and light green for book purchases. These five colors for each information usage were designed through a preliminary investigation.
- Keywords that occurred in multiple information usages 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). In other words, when we define $score_1(k)$ as the frequency of the keyword and $score_2(k)$ as the number of the co-occurred keywords, keywords whose $score_1(k) > 0$ are displayed in visualization 1 and keywords whose $score_1(k) > 1$ and $score_2(k) > 0$ are displayed in visualization 2.

Figure 4 shows an overview of the two algorithms. This example is part of the next section.

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 search (query)
Mon 13 16:06:18 JST 2010	Island Rental Car, Sado Island, most, cheap, rental car corporation	http://park2wakwak.com/...	Web search (title of browsed page)
Mon 13 17:14:33 JST 2010	the day after tomorrow, Sado Island, conference	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 search (title of browsed page)
Mon 13 19:09:35 JST 2010	Island Rental Car, Island Rental Car booking	https://mail.google.com/mail...	Mail



Visualization algorithm 1: Display all keywords

Visualization algorithm 2: Display keywords that occur more than once and have one or more relations

Figure 4. Generating knowledge-space algorithm

3.3. Knowledge-space browser

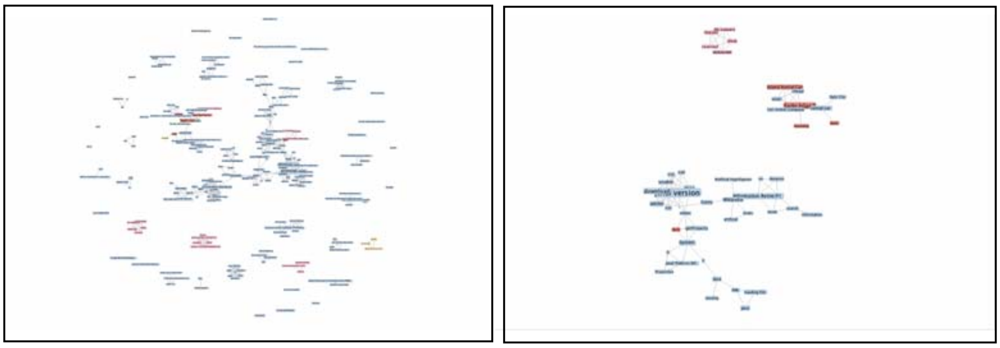
We implemented a knowledge-space browser that is comprised of four parts: (a) a knowledge space display that shows the user's knowledge space, (b) a history structure display that lists the user's history structures, (c) an operation display on which the user manages the knowledge space, and (d) a selected keyword display that lists the history structures on which the selected keyword is included.

The user can set a period (from date to date), select visualization algorithms 1 or 2, and display the knowledge space by the main menu. For the knowledge space, users can change either visualization algorithm 1 or 2 at anytime. The knowledge space generated by visualizations 1 and 2 to show all the keywords is displayed in Figure 5. Users can zoom in/out of the screen, move objects, and change the length of the links between the objects and their movement speed.

In the history structure display, they can change the order of the history structures based on the type of information usages, time, keywords, and URIs.

Figure 6 is an example screen of a knowledge-space browser in which a user's knowledge space was generated by visualization algorithm 1 on 13/Sep/2010 and then focused on by the user, who was preparing to go to a conference on 15/Sep/2010 on Sado Island. See Figure 4. 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 that "The day after tomorrow I will go to Sado Island for a conference." He booked a rental car at Island Rental Car by e-mail. The left cluster in the knowledge space in Figure 6, which contains *Sado Island*, *Island Rental Car*, *Sado*, *conference*, in red, 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.

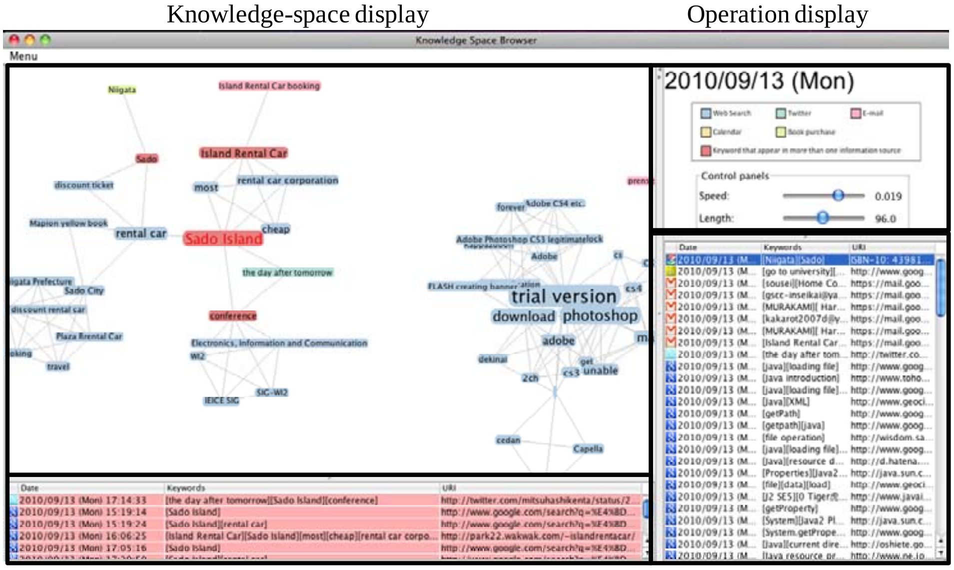
He selected *Sado Island* in a knowledge space and the history structures on which Sado Island is included are displayed in a selected keyword display.



(a) Visualization 1 (b) Visualization 2

Figure 5. Knowledge-spaces generated by visualizations 1 and 2 to show all keywords

Knowledge-space display Operation display



Selected keyword display History structure display

Figure 6. Knowledge-space browser with visualization 1

4. Experiment

4.1. Experiment 1

We evaluated the usefulness of our algorithm that extracts keywords.

4.1.1. 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, tweets, e-mails, calendars, and book purchases) using four methods. The comparative method (a) extracts only nouns, (b) extracts nouns and concatenates them using TermExtract [4], 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.

4.1.2. Results and analysis

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 ($p < .01$) by a one-way anova. There also were significant differences between (a) and (c), (a) and (d), and (b) and (d) (each $p < .01$).

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.

We analyze that since calendar data contain few noun phrases, adjectives or verbs, no significant difference were found among methods.

Table 1. Results of Experiment 1

	(a) **Nouns only**	*(b)* **NP by other method**	*(c)* **NP only by our method**	*(d)* **Our method (NP + Adj + V)**
Web Search **	2.71	2.81	4.02	4.05
Twitter **	2.56	2.43	3.56	3.91
E-mail **	2.85	3.26	4.10	4.15
Calendar	3.50	3.54	3.89	4.00
Book **	2.78	2.69	3.88	4.18

Note: *: $p < .05$; **: $p < .01$

4.2. Experiment 2

We investigated which information usage is most useful to generate a knowledge space.

4.2.1. 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 for about three weeks: web searches, tweets, e-mails, calendars, and book purchases. Since book purchases did not appear very often, we selected history structures that contain four or five information usages and treated them as a data set.

We selected one day from the data set and conducted Kendall's correlation analysis.

First, we showed the subjects the day's history structures and asked them to recall that day.

We then listed the keywords included from the history structure and asked them to evaluate how appropriately each keyword expressed their thoughts, activities, and what they saw/heard on that day (5: very appropriate; 4: appropriate; 3: OK; 2: not very appropriate; 1: inappropriate). To reduce the variations, the subjects rated the values based on normal distribution (i.e., 5: 7%; 4: 24%, 3: 38%; 2: 24%; 1: 7%).

For the five information usages, the keyword frequency was set as values. We set the number of the total frequencies of the keywords as frequency keywords. We set 1 when the keyword occurs in multiple information usages as multiple keywords and 0 when it occurs in only one information usage.

Since the number of keywords for the web searches was large, subjects rated keywords that occurred more than once.

4.2.2. Results and analysis

Table 2 shows Kendall's tau-b correlation coefficients with their significance levels. No information usage was directly related to the user evaluations. The frequency keywords and those from web searches were correlated (.85**), and the keywords from the multiple information usages and those from tweets were correlated (.49**).

Table 2. Results of Experiment 2

	Web	Twitter	E-mail	Cal	Book	Freq	Multiple	Eval
Web Search	1.00	.02	-.22**	-.11**	-.14**	.85**	.04	.12**
Twitter		1.00	-.04	-.02	-.02*	.10*	.49**	.04
E-mail			1.00	-.04	-.05	-.24**	.06	-.01
Calendar				1.00	-.02	-.13**	-.01	.10*
Book					1.00	-.17**	-.02	-.08*
Frequency						1.00	.09*	.11**
Multiple							1.00	.02
Evaluation								1.00

Note: *: $p < .05$; **: $p < .01$

The results vary by people. For example, for subject A, web search was not related (-.53*) to evaluation but e-mails was (.54*) ; for subject B, only web search was related (.49*) to evaluation.

The above results suggest that (a) important information usages vary based on cases, and therefore we need to deal with various information sources. In addition, (b) web search should be used as a basic information usage and (c) twitter should be useful to combine various information usages.

Keywords from web searches and calendars and frequency keywords showed significance with the evaluations, although the coefficient value was low (under .20). We should investigate potential of these keywords in the future.

4.3. Experiment 3

We investigated the features of three visualization methods: a list of history structures and visualization algorithms 1 and 2 when users recalled memories of a particular day within one month.

4.3.1. Method

The subjects were the same six male students from Experiment 2.

Three sets of history structures of three days from a data set (not used in Experiment 2) were assigned to three visualization methods.

The subjects used the history structures display (without showing other display parts) and explained what they recalled about the day in interviews. Next, we repeated the above processes using a knowledge-space display with visualization 1 (without showing the other display) and a knowledge-space display with visualization 2 (without showing the other display).

After each experiment, we measured the time from seeing a system to starting to talk about it (time to recall) and the length of the talking time about the memory (time to talk) for each visualization method. We also asked the subjects whether they thought the number of clusters (a cluster is defined as a network that connects more than two keywords) was appropriate in visualizations 1 and 2. In addition, we asked subjects to provide 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).

4.3.2. Results and analysis

The average number of clusters was 10.5 in visualization 1 and 4.2 in visualization 2. Four answered that the number of clusters was appropriate in visualization 1 and three in visualization 2.

Table 3 shows the average number of history structures, time to recall, and time to talk for history structures and visualizations 1 and 2.

In this experiment, the data sets (history structures) were different and their averages were largest in visualization 2. For visualization 2, although the number of history structures was large, both time to recall and talk were the shortest among the three. In contrast, for visualization 1, although the number of history structures was small, both time to recall and talk is the longest.

If we assume that shorter is better for time to recall for summarizing a day, visualization 2 is better than visualization 1 and the history structures. If we assume that longer is better for time to talk for recalling a day, visualization 1 is better than visualization 2 and the history structures.

Table 3. Time to recall and talk in Experiment 3

	List (History)	Visualization 1	Visualization 2
Average number of history structures	62.7	42.5	95.7
Average time to recall	2 min 06 sec	2 min 48 sec	1 min 35 sec
Average time to talk	1 min 59 sec	4 min 33 sec	1 min 48 sec

We combined the user comments for the pros and cons that express the same meanings. Comments made by more than one person are described in Table 4.

Table 4. User comments in Experiment 3

	List (History Structure)	Visualization 1	Visualization 2
Pros	Keywords are ordered chronologically [4].	I could get a day's overview by looking at the clusters [6]. One cluster corresponds to a topic [3].	The information was well-organized, with many keywords that helped recall [4]. Visualization 2 is better arranged than visualization 1 [3].
Cons	Viewing was difficult [2].	Similar clusters need to be combined [3].	This method doesn't provide enough information [2].

Note: [] indicates number of people.

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 believe that a list of history structures is useful to recall the past and that knowledge spaces are useful to summarize one day. A list of history structures is beneficial for sequentially viewing chunks of memory, and a knowledge space is good to grasp their overview.

All subjects recommended displaying the history structures and the knowledge space together.

4.4. Experiment 4

We evaluated the usefulness of our system to support human recollection when the users recalled the memories of a day within one month.

4.4.1. Method

The subjects were the same six males from Experiments 2 and 3. The experiment was conducted immediately before Experiment 3.

For Experiment 4, one set of the history structures of five information usages of one day gathered as a data set in Experiment 2 but not used in Experiment 2 or 3 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).

4.4.2. Results and analysis

The evaluation results are shown in Table 5. 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.

Table 5. Results of Experiment 4

	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

We also found that the subjects used the system in the following similar ways:

- They recalled the day's overview by looking at the knowledge space.
- They browsed the list of history structures to review or to find the details of the day; they found keywords that they did not find in the knowledge space.
- They looked at the knowledge space and found keywords.
- Some repeated the above processes; some changed the visualization algorithm.
- They started to talk about their memories of the day.

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

5. Case study

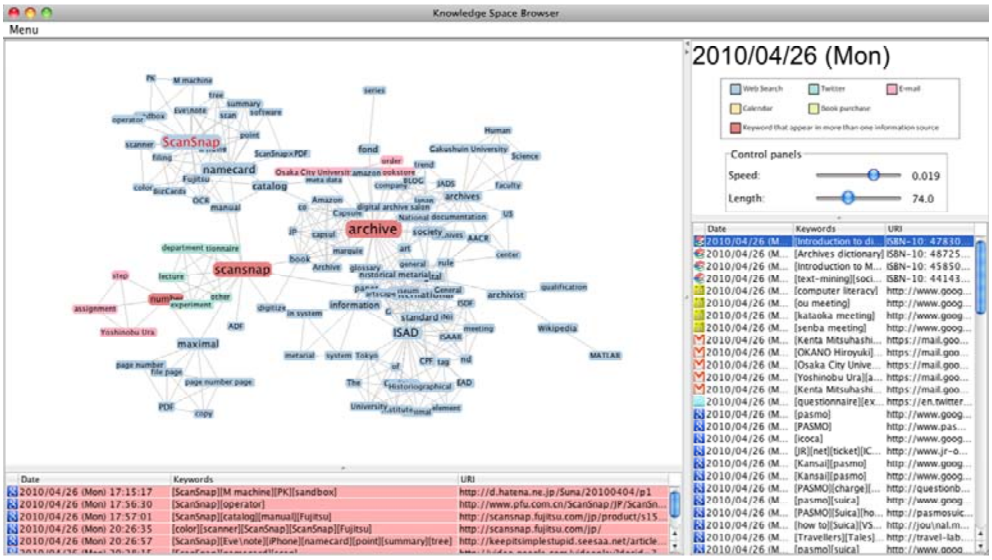
We show a case study in which a different user explores desktop by using the knowledge-space browser. The user went to a conference of JSAS (The Japan Society for Archival Science) on April 24 and 25 in Tokyo and gave a lecture on computer literacy on April 26.

Figure 7(a) shows an example screen of the knowledge-space browser with visualization 1.

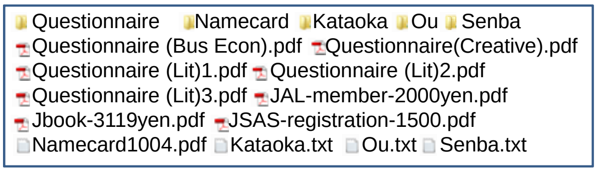
Most of the keywords were generated from web search histories. *Scansnap* (a scanner's name), *archive*, and *number* are displayed in red in the knowledge-space display. *Archive and scansnap* were the two most frequent keywords in the overall history structures and the web searches. The user searched for information about how to use ScanSnap on the web. *Questionnaire*, *experiment*, *other*, *department*, and *lecture* are displayed from twitter logs. *Computer literacy* is the name of a lecture in which the user conducted a questionnaire survey (see calendar data in the history structure display). She wanted to scan the completed questionnaires. After returning from the conference, the user searched for information about archives and bought some books (see book purchase data in the history structure display).

Figure 7(b) shows Microsoft windows 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).

Thus, user can explore the knowledge-space display and access the desktop data by recalling the contexts of files (e.g. she divided questionnaires by department due to the scanner's page limitations; the nameacard folder includes the scanned namecards of people met at it; she bought a book for 3,119 yen about archive).



(a) Knowledge-space browser with visualization 1



(b) Windows search results

Figure 7. Case study

6. Related work and discussion

6.1. Related work

This research is a part of a system called *Memory-Organizer* that helps users construct "externalized-memory" [5, 6]. The system consists of six browsers: a thinking space browser, an overlay web browser, an interest space browser, a map browser, a calendar browser, and the knowledge-space browser developed in this research. We previously proposed externalized-memory as a key concept of Memory-Organizer [5]. We enhanced this idea to an externalized-memory model to design a knowledge-space browser. 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, 10], to overcome information overload [11, 12]. 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.

Some systems offer visual interfaces to help users explore desktops [13]. iMecho [14] is such an example of associate memory based on a desktop search system. It resembles our system in the light of concept and semantic networks; however, our system links keywords and iMecho connects desktop sources such as folders or files.

The Digital Parrot [15], which resembles our approach, is based on the characteristics of human memory and developed a new interface (graph-based data model) for PIM. Although they presented an interesting data model based on semantic networks, they did not describe how to generate them. We proposed history structures to easily capture various information usages even when it is difficult to obtain semantic information.

Our research and PIM are related to a research topic called lifelogs that capture a person's entire lives [16]. Mylifebits [17] 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, such as SenseCam [18]. 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.

Finally, based on semantic networks (e.g., [19, 20]), many systems resemble our system in the light of visual interfaces, but they have different objectives such as intelligent information processing or ontology development. Our primary purpose is to support human recollection.

6.2. Discussion

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.

The history structures and associated algorithms are rather simple and can be implemented easily. This is a strong benefit of our approach. However, we lose much semantically useful information when generating history structures. We need to consider how to keep semantic information in the information usages to create such semantic networks as Digital Parrot [15] to help users recollect more intelligently.

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. Fourth, although we describe a case study of desktop search, we have not evaluated the effectiveness, which will be adopted as future research themes.

7. 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. Our approach is based on an externalized-memory model inspired by a human memory model. 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. We described a case study in which users explore desktops using the knowledge-space browser.

Some parts of all of our algorithms require improvement, including those for generating keywords and visualization. We will also conduct studies for different periods.

8. References

[1] MeCab: Yet Another Part-of-Speech and Morphological Analyzer, http://mecab.sourceforge.net/.
[2] Kenta Mitsuhashi, "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] Allan M. Collins, Elizabeth F. Loftus, "A Spreading-Activation Theory of Semantic Processing", Psychological Review, vol. 82, no. 6, pp. 407-428, 1975.
[4] http://gensen.dl.itc.u-tokyo.ac.jp/termextract.html.
[5] Harumi Murakami, Takashi Hirata, "Information Acquisition and Reorganization from the WWW by using Memory-Organizer", Bulletin of Osaka City University Media Center, vol. 3, pp. 9-14, 2002.
[6] Harumi Murakami, Takashi Hirata, "A System for Generating User's Chronological Interest Space from Web Browsing History, International Journal of Knowledge-Based and Intelligent Engineering Systems", vol. 8, no. 3, pp.149-160, 2004.
[7] Harumi Murakami, Takashi Hirata, "Recall of Web Browsing History by Generating User's Knowledge Space", Bulletin of Osaka City University Media Center, vol. 5, pp. 21-26, 2004.
[8] Harumi Murakami, "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] William Jones, "Personal Information Management", ARIST, vol. 41, pp. 453-504, 2007.
[10] William Jones, "The Future of Personal Information Management, Part I: Our Information, Always and Forever", Morgan & Claypool Publishers, 2012.
[11] Susan Dumais, Edward Cutrell, JJ Cadiz, Gavin Jancke, Raman Sarin, Daniel C. Robbins, "Stuff I've Seen: A System for Personal Information Retrieval and Re-Use", In Proceedings of SIGIR 03, pp. 72-79, 2003.
[12] Edward Cutrell, Susan T. Dumais, Jaime Teevan, "Searching to Eliminate Personal Information Management", CACM, vol. 49, no. 1, pp. 58-64, 2006.
[13] Tye Rattenbury, John Canny, "CAAD: An Automatic Task Support System", In Proceedings CHI 2007, pp.687-696, 2007.
[14] Jidong Chen, Hang Guo, Wentao Wu, Wei Wang, "iMecho: An Associative Memory Based Desktop Search System", In Proceedings of ACM CIKM'09, pp. 731-740, 2009.
[15] Andrea Schweer, Steve Jones, Annika Hinze, "Trails of experiences: Navigating Personal Memories", In Proceedings Chinz 09, pp. 105-106, 2009.
[16] Abigail J. Sellen, Steve Whittaker, "Beyond Total Capture: A Constructive Critique of Lifelogging", CACM, vol. 53, no. 5, pp. 70-77, 2010.
[17] Jim Gemmel, Gordon Bell, Roger Lueder, Steven Drucker, Curtis Wong, "Mylifebits: Fullfilling the Memex Vision", In Proceedings of Tenth ACM Multimedia, pp. 235-238, 2002.
[18] Steve Hodges, Lyndsay Williams, Emma Berry, Shahram Izadi, James Srinivasan, Alex Butler, Gavin Smyth, Narinder Kapur, Ken Wood, "SenseCam: A retrospective memory aid. In Proceedings of Ubicomp", Springer, pp. 177-193, 2006.
[19] Junsheng Zhang, Yunchuan Sun, "An Analogy reasoning model for Semantic Link Network", International Journal of Digital Content Technology and its Applications, vol. 4, no. 7, pp. 128-139, 2010.
[20] Reena T. N. Shetty, Pierre-Michel Riccio, Joël Quinqueton, A Dynamic Knowledge Representation Model based on Hybrid Approach, Journal of Convergence Information Technology, vol. 2, no. 3, pp. 29-38, 2007.