A System using Tag Cloud
for Recalling Personal Memories

Harumi Murakami and Ryutaro Murakami

Osaka City University, Sugimoto, Sumiyoshi, Osaka 558-8585, Japan
`harumi@osaka-cu.ac.jp`
http://murakami.media.osaka-cu.ac.jp/

Abstract. The research presented here extends a previous prototype that supported human recollection with tag clouds created from the use of a personal calendar and Twitter. That system weighted keywords by combining term frequency and the number of photos taken by users to recall memorable events. The aim in this paper is to improve upon our previous work and present a full system that uses tag clouds for recalling personal memories. The main differences from our previous work are as follows. (1) Multiple information sources such as SNSs or instant messengers can be used. (2) To handle multiple information sources, we present a new unified keyword-weighting algorithm. (3) We implemented new functions, such as keyword search, tag search, and photo display, to form a complete system. Preliminary experiments reveal the usefulness of our system in recalling personal memories.

Keywords: Personal memory support, Tag cloud, Calendar, Twitter, LINE, Number of photos

1   Introduction

Memory is crucial for various activities in our daily lives. 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 thus recall what we did last year or in previous years. Or we might simply want to reminisce about the day we saw our spouse for the first time. In addition, the amount of information that we manage is increasing. Consequently, we aim to support human memory.

We have presented a prototype that supports human recollection with tag clouds created from the use of calendar and Twitter [1]. Since we assumed that we could identify memorable events based on the days when a user took many photos, we weighted keywords by combining term frequency and the number of photos taken by users to generate tag clouds.

However, we found the following problems in the prototype. (1) Information sources are fixed in a calendar and Twitter. The trends of SNSs are changing, and their preferences are different. We need to easily cope with other information sources. (2) The keyword weighting algorithms for calendar and Twitter are

different because they are customized to each source. To cope with new information sources, we need to develop an effective simplified weighting algorithm. (3) There is no function of keyword search, which is obviously useful for recalling memory. (4) Photos were used for keyword-weighing algorithms, but the photos themselves were not displayed. Photos are also obviously effective for recalling the past.

The aim of this research is to improve the system developed in our previous work to cope with the above problems and to present a new complete system using tag cloud for recalling personal memories. The main differences from our previous work are as follows. (1) Multiple information sources such as SNSs or instant messengers can be used. (2) To handle multiple information sources, we presented a new unified keyword-weighting algorithm. (3) We implemented new functions, such as keyword search, tag search, and photo display, to build a complete system.

2   Tag Browser

Our research generates a tag cloud by extracting keywords from various kinds of information usage and weighting them using term frequency and the number of photos. First, we obtain the data written by users and generate history structures [2]. Next, we generate tag clouds from these history structures. We call our new system a tag browser. LINE, which is classified as an SNS but is actually the most frequently used instant messenger in Japan, is adopted as well as calendar and Twitter. The user can set the period and the number of tags (default number is 30) and the system displays a tag cloud, information logs, and photos.

Several new functions have been developed for the tag browser. (1) Multiple information sources such as SNSs or instant messengers can be used. (2) Log windows (display position of information sources) are defined automatically based on the amount of log data. (3) The same keyword-weighting algorithm is used for all information sources. (4) The user can change the weighting algorithm using a slider interface. (5) Tag and keyword search functions can be made. (6) Photos are displayed according to date. In what follows, we translated the examples in this paper from Japanese into English for publication.

2.1   Generating History Structure

The basic components constructing the history structure include time, keywords, and log (original text) sets[1]. Nouns and noun phrases are extracted from information logs as keywords. For example, from a LINE message "$\cdots$ Yes! In Kyoto every day is like a festival!" The system extracts "Kyoto," "every day," and "festival" as keywords. See Fig. 1 for example of history structures.

---

[1] We have changed from using URI [2] to log in this research. Other attribute information such as URI can be stored optionally. For LINE, the receiver and sender of the message are stored as To and From, respectively.

History Structure (Calendar)	Time	Keywords	Log
	Sat Aug 12 00:00:00 JST 2017	[Hiroshima]	going back to Hiroshima
	...		

History Structure (Twitter)	Time	Keywords	Log
	Sat Feb 18 06:15:35 JST 2017	[Randen][Arashiyama] [Kyoto]	I'm at Randen Arashiyama in Kyoto···
	...		

History Structure (LINE)	Time	Keywords	Log	To	From
	Sat Feb 04 21:58:00 JST 2017	[Kyoto][every day] [festival]	··· Yes! In Kyoto, every day is like a festival!	Ichi	Ryu
	...				

Fig. 1. Example of history structures

2.2   Displaying Tag Clouds

Based on the results of experiments, it was decided to display tag cloud to the left and information logs to the right. First, we describe how to display a tag cloud. According to the input (period with or without a keyword), weights of keywords are calculated using Eqs. (1) and (2).

First, W_{i,t_k} defines the weight of term t_k in each information source i:

$$W_{i,t_k} = \sum \frac{1 + C_{photo}(G_{date}(t_k))}{C_{keyword}(G_{date}(t_k))}, \tag{1}$$

where the addition occurs for each term t_k appearing in each history structure. Here, $G_{date}(t_k)$ is a function that obtains the date of term t_k, $C_{photo}(R)$ is a function that obtains the number of photos in range R, and $C_{keyword}(R)$ is a function that obtains the number of keywords (terms) in range R. The denominator works as a normalizing factor for the number of information logs.

Next, W_{t_k} defines the weight of term t_k:

$$W_{t_k} = \sum_{i=1}^{n} \alpha_i W_{i,t_k}, \tag{2}$$

where n is the number of information sources and the default value for α_i is 0.5. In this research, we assigned the following values to i: calendar: 1, Twitter: 2, and LINE: 3.

The size of tags (keywords) is based on the weights determined by Eq. (2). Font colors are designed according to the information sources in which they appear. They reflect the original color image of the applications except for the calendar. If term t_k only appears in the calendar, it is orange, while it is blue in Twitter and green in LINE. If term t_k appears in more than one log, it is red. Tags are sorted by the time of first appearance.

2.3   Displaying Information Logs and Photos

According to the amount of data, information logs are displayed from left to right. In each information source, logs are displayed according to time. When only a period is input, all logs during the period are displayed. In tag or keyword searches, logs that contain the tag or keyword are displayed. Photos are displayed in ascending order according to date. When only a period is input, all photos during that period are displayed. In tag or keyword search, the photos taken on the date matching the date of logs containing the searched tag or keyword are displayed.
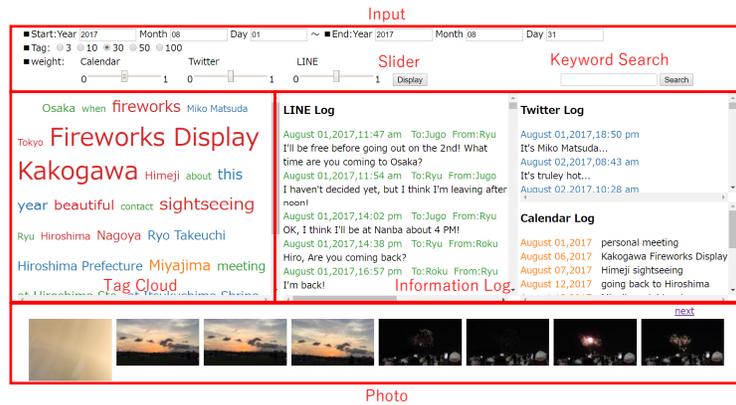
2.4   Example

Fig. 2. Screenshot of display for one-month period

Fig. 2 shows an example of the system's basic usage. The user can set the period and number of tags, and a tag cloud is displayed on the left while information logs are displayed on the right. In this case, LINE, Twitter, and calendar logs are displayed from left to right according to the amount of data. Logs and photos are displayed according to time order, and the user can scroll them. The user can change parameter α_i of Eq. (2) ($0 \leq \alpha_i \leq 1$) by using the slider.

Fig. 2 shows an example screen of one month (Aug. 2017) for a user. "Fireworks Display" is the largest tag and is shown in red. 10 out of 30 (33%) tags are related to places. In this month' case, many tags work as clues to remember salient events.

When the user clicks a tag in the tag cloud, history structures that contain the keyword are extracted, a new tag cloud is generated, and the information logs of these history structures are displayed. For example, when the user clicks "fireworks," information logs containing "fireworks" are searched and displayed.

Photos taken on the dates of the retrieved logs are displayed, and thus the user can recall pleasant memories of "Fireworks Display at Kakogawa."
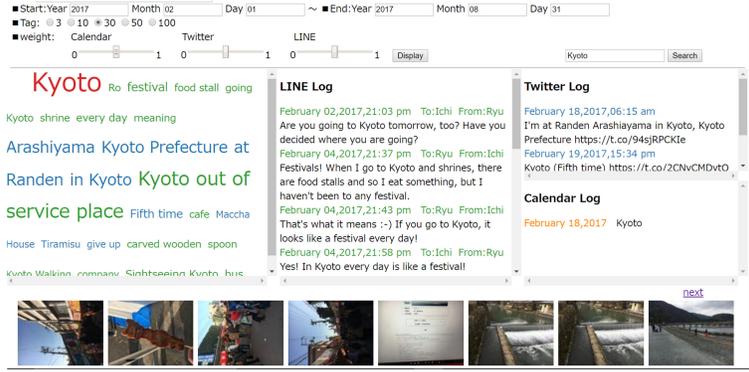

Fig. 3. Screenshot of display for keyword search

Fig. 3 shows an example of keyword search. When the user inputs "Kyoto," information logs containing "Kyoto" are retrieved. In this case, the period is set to about half a year. Although the user went to Kyoto just once in February as in calendar log, many message exchanges using LINE related to Kyoto are retrieved, and various types of information (plan, memory, etc.) about Kyoto can be recalled. Photos related to Kyoto are also displayed.

The tag browser can be used for business as well. For example, the user can look for photos of certain presentation slides taken at conferences. Although there is no manual tag for the photos, they can be found using a date or information logs searched via tag or keyword search.

3   Experiment

We recruited subjects who satisfied the following conditions: (a) those who take photos and (b) those using more than one of the following: writing a calendar schedule on any device, using Twitter, or using LINE. We gathered data for more than three months, and each experiment used the data of one month.

3.1   Experiment 1: Generating Tag Cloud

We evaluated the usefulness of our algorithm for creating tag clouds. Our subjects were ten males and one female, with an average age of 22.8. We prepared four tag clouds from four weighting algorithms (our algorithm, raw frequency (number of t_k in each history structure), relative frequency (number of t_k divided by number of keywords in each history structure), and a previous algorithm [1])

that display 30 keywords for comparison. The subjects performed the following task: "Rank the four tag clouds by the ease with which they helped you recall your memories." Five subjects selected our algorithm as 1st rank among the four algorithms. Three subjects selected raw frequency, two selected the previous algorithm, and one selected relative frequency. No subject selected our algorithm as 4th (worst) rank. Although the difference was slight, our algorithm was best among the four algorithms, and we found that it was possible to unify a keyword-weighting algorithm.

3.2 Experiment 2: Tag Browser

We evaluated the usefulness of our system. Our subjects were five males and one female, with average age of 23.2. After recalling a one-month-old memory using the system, the subjects answered questions at five levels (5 to 1; 5 is best). Here, we extract some results of the questions in Table 1.

Table 1. Some results of questionnaire

Questions	Mean	SD
Q1 Was the system useful for recalling your memories?	4.50	0.50
Q2 Do you want to use this system in the future?	4.33	0.75
Q3 Did you feel fun while using this system?	4.33	0.75

For the question "Which part was the most useful for recalling your memories? - A: Tag cloud, B: Logs, C: Photos," five subjects answered A: Tag cloud and one subject answered C: Photos. Overall, the results show the usefulness of our system for recalling personal memories.

4   Related Work

This research is a part of our work on personal memory support. Murakami [2] presented the concept of information structure called history structure. The history structure integrates various kinds of information usage. Murakami et al. [3] developed a knowledge-space browser that displays a network rather than a tag cloud. Matsumoto et al. [1] developed the initial prototype of the tag browser displaying a tag cloud. This research improved upon the previous work [1] by further refining the tag browser to its current form.

Our research uses tag clouds for human memory recall, even though little research has used them for this purpose. Chen and Jones [4] developed a prototype system called iCLIPS that searched through personal lifelogs for memory support. In iCLIPS, computer activities and the names of locations and people were displayed in term clouds, which resemble tag clouds. No detailed algorithms for generating term clouds and user studies of the prototype have been reported.

We focus on generating better tag clouds rather than accumulating all human activities. We also conducted preliminary experiments that demonstrated our system's usefulness.

Mathur et al. [5] presented a prototype system of a tool called LifeView, which visualizes textual lifelogs for Sentimental Recall and Sharing. In this system, events are manually created by users who manually annotate tags. A tag cloud (these tags) for one event was displayed. On the other hand, our approach automatically extracts keywords from information sources and generates tag clouds.

Aires and Goncalves [6] presented Personal Information Dashboard, a web application that allows users to see, at a glance, various facets of their lives. In this system, Keywords Cloud is a tag cloud-like visualization that shows the most important words from a set of emails, posts, and/or tweets. To calculate the important words, they used tf-idf. The presentation (layout) is spacial (not sorting). Keywords Cloud can be configured to show data from a specific time period. Our tag cloud's algorithm, presentation, and information sources are different from Keywords Cloud.

Many systems and researches have generated tag clouds. The two main purposes are summarizing and navigating certain content. Rivadeneira et al. [7] classified tasks that tag clouds can support into four: (a) search, (b) browsing, (c) impression formation or gisting, and (d) recognition/matching. In general, (a) tag selection algorithms are based on the frequency of the terms or objects, (b) important tags are represented by size or color, and (c) tags are ordered alphabetically.

Venetis et al. [8] evaluated existing algorithms for exploring and understanding a set of objects against tf-idf-based algorithms and presented a maximum covering algorithm that seems a very good choice for most scenarios in their experiment and a popularity algorithm, which is easier to implement and performs well in specific contexts. Both algorithms are based on the number of objects associated with tags. Our research presented a unique algorithm based on the frequency of terms and photos.

Much research has presented ideas for integrating information in the light of Personal Information Management (PIM) [9], to overcome information overload (e.g. [10]). History structure is simply generated from existing information sources. Our research resembles text-based lifelog research. Our approach is simple and does not need special devices to capture information from the world.

5 Summary

We improved upon what we developed in our previous work and presented a complete new system using tag clouds for recalling personal memories. The main differences from our previous work are as follows. (1) Multiple information sources such as SNSs or instant messengers can be used. (2) To handle multiple information sources, we presented a new unified keyword-weighting algorithm. (3) We implemented new functions, such as keyword search, tag search,

and photo display, to build a complete system. Preliminary experiments reveal the usefulness of our system for recalling personal memories.

References

1. Matsumoto M., Matsuura S., Mitsuhashi K. Murakami H.: Supporting Human Recollection of the Impressive Events using the Number of Photos. In: Proceedings of the 6th International Conference on Agents and Artificial Intelligence - Volume 1: ICAART, pp. 538-543. (2014)
2. 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)
3. Murakami H., Mitsuhashi K., Senba K.: Creating User's Knowledge Space from Various Information Usages to Support Human Recollection. In: Jiang H., Ding W., Ali M., Wu X. (eds.) IEA/AIE 2012. LNCS, vol. 7345, pp. 596-605. Springer, Heidelberg (2012) https://doi.org/10.1007/978-3-319-60045-1_6
4. Chen, Y., Jones, G. F.: Augmenting Human Memory using Personal Lifelogs. In: Proceedings of the 1st Augmented Human International Conference. ACM, New York (2010)
5. Mathur, A., Majumder, A., Datta, S., Menon, S., Malhorta, S., Dahiya, A.: LifeView: a lifelog visualization tool for supporting sentimental recall and sharing. In: OzCHI'12: Proceedings of the 24th Australian Computer-Human Interaction Conference, pp. 371-380. ACM, New York (2012)
6. Aires, J., Goncalves, D.: Personal Information Dashboard - Me, At a Glance. In: PIM 2012, http://pimworkshop.org/2012/. (2012)
7. Rivadeneira, A. W., Gruen, D. M., Muller, M. J., Millen, D. R.: Getting our head in the clouds: toward evaluation studies of tagclouds. In: CHI '07 Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 995-998. ACM, New York (2007)
8. Venetis, P., Koutrika, G., Garcia-Molina, H.: On the selection of tags for tag clouds. In: WSDM '11 Proceedings of the fourth ACM international conference on Web search and data mining, pp. 835-844. ACM, New York (2011)
9. Jones, W.: Personal Information Management, ARIST 41 Annual Review of Information Science and Technology, 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: SIGIR '03: Proceedings of the 26th annual international ACM SIGIR conference on Research and development in information retrieval, pp. 72-79. ACM, New York (2003)