IIAI International Journal Series
International Institute of Applied Informatics
2015, Vol. 1, No. 1, 1 – 99

A System to Support Memory Recall by Creating Tag Clouds from Calendars, Twitter, and Photos

Harumi Murakami * , Masaki Matsumoto,
Sho Matsuura, Shuta Watanabe †

Abstract

Human memory is crucial for many daily activities. However, human memory is unreliable. Since existing memory aids such as diaries and notebooks need some effort for writing and accessing, we present an approach to capture and display memory cues from the user's daily activities, such as writing schedules, posting SNS messages, and taking photos. We developed a system that supports memory recall by creating tag clouds from the user's calendars and Twitter posts. The main feature of this research is using the number of photos taken by users to weight the tags to recall impressive events. This is based on the hypothesis that impressive events are found on the days when users took many photos. We evaluated out tag cloud generation algorithms and our implemented prototype system and obtained the following results for user memory recall: (1) Our tag cloud (using our weighting method with calendar and Twitter data) is superior to other tag clouds. (2) Sorting by time in the tag cloud outperforms other sorting methods. (3) Our implemented prototype system is useful.

Keywords: human recollection, SNS, tag cloud, word cloud

1 Introduction

Human memory is crucial for many daily activities. For example, we often write summaries of meetings we attended or progress reports for a particular day or week. We plan anniversaries and recall how we celebrated last year or in previous years. Or we might simply want to reminisce about the day we saw our spouse for the first time. However, human memory is unreliable. As time passes, our ability to recall past memories deteriorates. In addition, the amount of information we manage is also increasing due to the ubiquity of the internet and smartphones. Since managing our memories is becoming more difficult, we need to support human memory recall.

Existing memory aids such as diaries and notebooks are often used to remember what we have done, observed, thought, and felt. To use these aids, we must make an effort to prepare the equipment and write down what we want to remember. However, there are times when it is not possible to prepare for and then write such a document. If you are very

* Osaka Metropolitan Universisy, Osaka, Japan
† Osaka City University, Osaka, Japan

busy with your work or personal life, you may not be able to record events or ideas during the day. After a few days, we are likely to forget what we did and what we thought. Harvey et al. [1] pointed out that (1) human memory is relatively poor, (2) human memory is cue-driven, and (3) a cue can be used to augment human memory. To solve human memory recall problems, we investigated a way to capture and display "memory cues" from daily activities without requiring much effort by the users.

Before computers and the internet, people often wrote plans on such paper media as organizers, planners, or diaries to manage their schedules. Due to smartphone proliferation, the number of people using digital media to manage such information has increased. Calendar applications on smartphones as well as online web-based calendars are also being used. Google Calendar is a very popular calendar service. Even though the growth of digital calendars is increasing, many people still prefer paper media and others use both paper and digital media. Social Networking Services (SNS), through which users can send and share information with others, are also growing. Users can easily post by smartphones their thoughts, activities, and feelings. Twitter is a popular SNS through which users can connect with others and post short messages. Although the primary purpose of writing schedules or posting messages on SNSs is not for recalling past events, we believe that schedules made in any medium and messages posted on SNSs are still useful sources of information for memory cues.

In this paper, we propose a memory recall support system that extracts useful keywords from the texts written by users on calendars and Twitter. We use *recall* here because this verb appropriately targets human memories and also suggests *remembering*, *recollecting*, and *reminiscing*. Since we believe that people take photos to remember something, days on which a user takes many photos probably contain memorable events. We therefore use not only the frequency of keywords but also the number of photos to weight the extracted keywords. We present the weighted keywords using tag clouds to jog user recall.

Below, in Section 2 we explain an overview of our approach. Our algorithms and implemented prototype system are described in Sections 3 and 4. We describe five experiments in Section 5. We show related work and discuss the significance of our research in Section 6.

2 Approach

Our research supports human recall by extracting keywords from calendars and Twitter, weighting them using term frequency and the number of photos, and displaying them by tag clouds. An overview of our approach is shown in Figure 1. Note that the examples in this paper were translated from Japanese into English for publication. First, we obtain the data written by users on calendars and Twitter and generate files called history structures. A history structure is an information structure that is constructed from time, keywords, and URI sets for existing information integration [2]. Next, we generate tag clouds from history structures.

As a feature of this study, we support the recall of impressive memories that are often obtained on trips or at special events because people tend to take more photos on such occasions. To put it another way, impressive events can probably be identified around days on which a user takes many photos. The above hypothesis is an important element in our study. Tag clouds (or word clouds) are visual presentations of a set of words, typically a set of tags, in which such attributes of a text as size, weight, or color can be used to represent

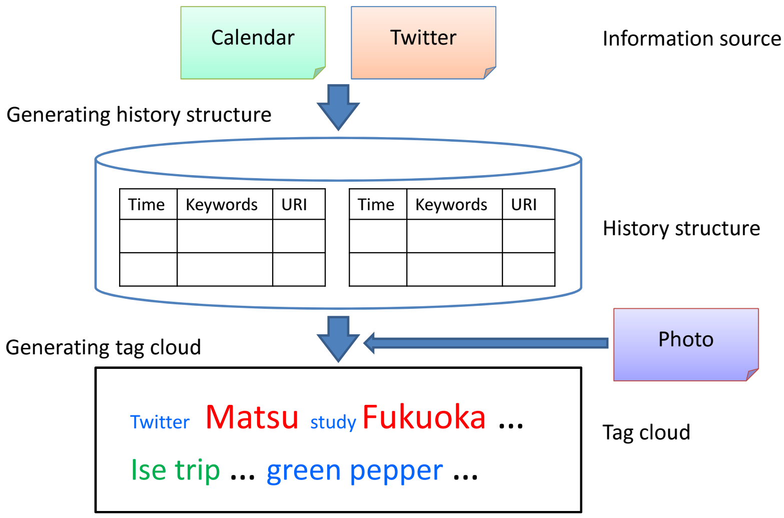

Figure 1: Overview of our approach

the features (e.g., frequency) of associated terms [3] [4]. The basic idea is that tag clouds provide navigational clues by aggregating tags and corresponding resources from multiple sources [5]. Tag clouds are useful for quickly perceiving the terms that appear in a target. We exploit this advantage and adopt tag clouds as memory clues, not as navigational clues for information resources.

Figure 2 shows an example of a tag cloud created with our system with a user's data for one month (September 2012). Figure 3 shows an example of a created tag cloud (with a comparative system) that only calculates the weights of the tags (keywords) by term frequency. In this period, he was cramming for a national exam of the application information technology, and he also traveled to Ise with a friend named Matsu for a few days and had lots of fun. In the comparative tag cloud (Figure 3), *study* and *application* appeared larger because he tweeted these terms many times. In our proposed tag cloud (Figure 2), *Ise trip* and *Matsu* are displayed larger and *Matsu* is highlighted in red. The word *happiness* reflects his tweets during his trip, and *green pepper* reminds him that he ate some incredibly hot peppers for the first time at a Japanese style barbecue restaurant that caused him to tear up. That food made a deep impression on him. He took some photos during the trip, and *green pepper* was displayed even though it just occurred on one day. In both tag clouds, *Matsu* and *Fukuoka* (place) are displayed in red since they are included in both the calendar events and tweets; however, they are larger in Figure 2 than in Figure 3. We believe that our proposed tag cloud outperforms the comparative scheme and effectively helps users recall impressive memories.

3 Algorithm

In this section, we describe our algorithms that obtain information from calendars, Twitter, and photos to create history structures and tag clouds.

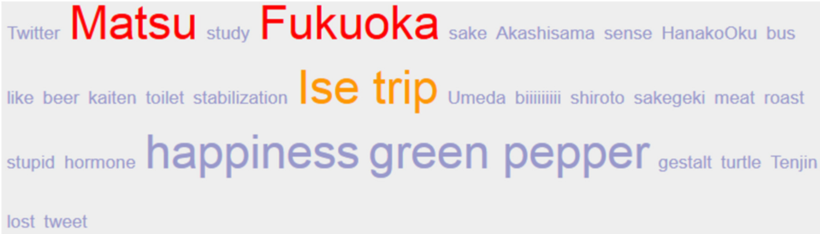

Figure 2: Tag cloud created by our method

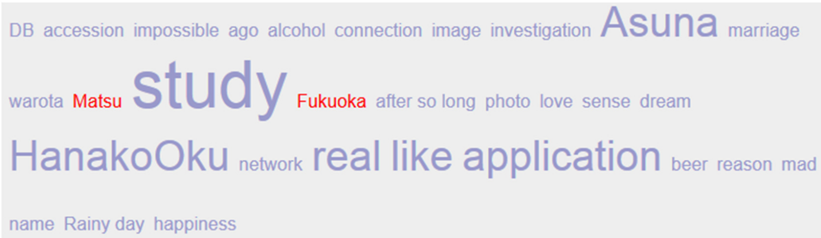

Figure 3: Tag cloud created by comparative method

3.1 Generating History Structures

A history structure is constructed from a time, keywords, and a uniform resource identifier (URI). Actually *keywords* in history structures accept any type of character strings, which may contain any kind of language: words, phrases, and even non-words. In this research, using MeCab, a Japanese morphological analysis tool from either a calendar event or a tweet, we extracted nouns and noun phrases that become keywords.

(1) CALENDAR We obtained an event time (start time) and its title from a calendar entry. Nouns and noun phrases were extracted from event titles and become keywords. For example, from an event title, “Ise trip with Matsu,” *Ise trip* and *Matsu* became keywords.

(2) TWITTER Since tweets generally express user thoughts or activities, we use all of them except those starting with @, which are mainly discourse, and official retweets (RTs), which are usually the opinions of others. We obtained the tweet times and their texts from the above tweets. Nouns and noun phrases were extracted and become keywords. For example, from a tweet, “Too spicy! (crying) with green pepper!,” *crying* and *green pepper* became keywords.

GENERATING KEYWORD ALGORITHM: We describe a generating keyword algorithm that generates a set of keywords from such texts as event titles in calendars ane tweets. First, it extracts terms with MeCab, a Japanese morphological analysis tool. When an extracted term is a noun, a common noun, a proper noun, a verbal noun, a noun suffix, or a noun as a number (type 1), it is repeatedly concatenated using heuristics with previous terms as a non-Japanese noun phrase or a Japanese noun phrase. When the noun is a noun adverbial or a noun adjective base (type 2), it directly becomes a keyword. During the

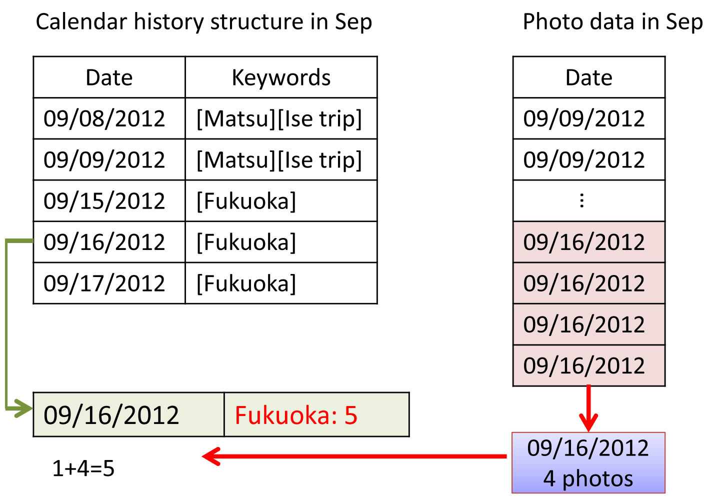

Figure 4: Example of calculating weights of calendar terms

above process, we removed unnecessary text strings using 24 patterns or words. A detailed keyword generation algorithm was described in a previous work [6].

3.2 Generating Tag Clouds

3.2.1 Calculating Tag Weights

Hereafter, *term* denotes a keyword in the history structures since the notation naturally expresses character strings in information retrieval. Calculating a term's weight depends on the information source. In addition, we calculate the weight based on the number of photos taken by users, since we assume that users take many photos on special days.

Next we define how to weight the terms in the calendars and those in the tweets and add the weights.

(1) WEIGHTING CALENDAR TERMS We define weighting function $CalW(t)$ to weight term t that appeared in the history structure of the calendars:

$$CalW(t) = \sum_{t \in HS} (1 + C_{photo}(G_{date}(t))). \tag{1}$$

$G_{date}(t)$ is a function that gets the date of term t. $C_{photo}(R)$ is a function that counts the number of photos of range R. HS means the history structure.

The more photos taken by users, the more the weights of the terms increase. Figure 4 shows an example of calculating calendar terms. The example data are a user's usage on September 2012. We calculated the weight with these example data. *Fukuoka* appears three times in this example. Each *Fukuoka* is weighted one plus the number of the day's photos. For example, *Fukuoka* on September 16 has a weight of 5 points because there are four photos (1+4). The other *Fukuoka* examples are calculated in the same way, and all the points for *Fukuoka* in the history structure are added.

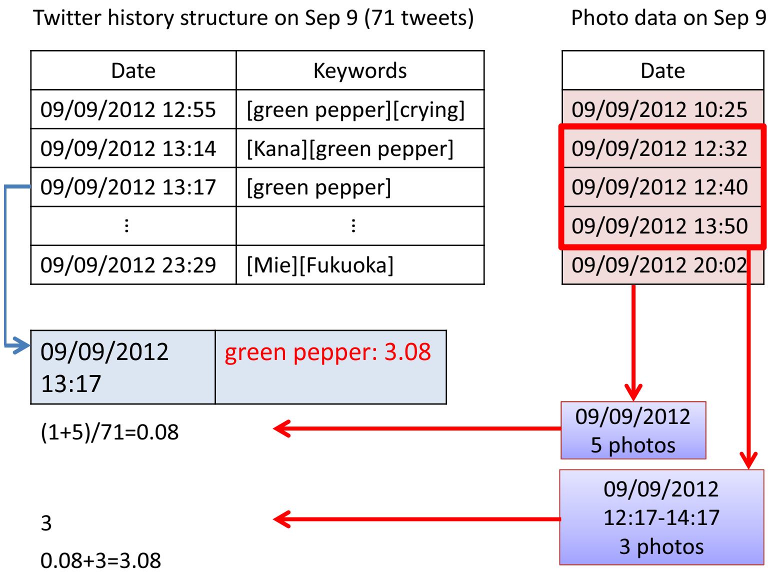

Figure 5: Example of calculating weights of Twitter terms

(2) WEIGHTING TWITTER TERMS We define weighting function $TwiW(t)$ to weight term t that appears in the history structure of the tweets:

$$TwiW(t) = \sum_{t \in HS} \left(\frac{1 + C_{photo}(G_{date}(t))}{C_{tweet}(G_{date}(t))} + C_{photo}(G_{time-ba}(t))\right). \quad (2)$$

$C_{tweet}(R)$ is a function that counts the number of tweets of range R, and $G_{time-ba}(t)$ is a function that gets the time one hour before and after term t (i.e., two hours).

We assume that the more tweets in a day, the more both noisy and good terms increase. We need to normalize the weights based on the number of tweets. The denomination of the first part inside sigma shows the normalization. This part corresponds to the calendar weights (Eq. 1). Additionally, Twitter disseminates information in real time. Since we believe that both a photo and a tweet made in a close time proximity are highly related, the number of photos taken one hour before and after the tweet time is added to the weight (the second part inside the sigma of Eq. 2). Figure 5 shows an example of the calculation of twitter terms. We calculated the weight of *green pepper* that was included in a tweet at 13:17 (hereinafter gp-1317). *Green pepper* appears six times in September, and there are 71 tweets and five photos on September 9. First, the weight of the first part of gp-1317 becomes (1+5)/71=0.08. Second, there are three photos one hour before and after its tweet time. Thus, the weight of the second part of gp-1317 becomes 3. From the above results, the weight of gp-1317 becomes 0.08+3=3.08. The other *green paper* examples are calculated in the same way, and all of the points in the history structure are added.

Next, we define weighting function $Weight(t)$ of term t by combining the calendar and Twitter weights:

$$Weight(t) = \alpha CalW(t) + (1-\alpha) TwiW(t). \quad (3)$$

We set 0.5 to α. In other words, the weights for calendars and Twitter are added equally. This is equivalent to the default weight of the term:

$$Weight(t) = CalW(t) + TwiW(t). \tag{4}$$

3.2.2 Presenting Tag Clouds

We aim to create a tag cloud that is useful for memory recall. Here we describe the design principles: font size, the tag colors, and the tag orders.

(1) FONT SIZE As the score of the terms defined in the previous section increases, the tag's size becomes larger. The size of the tags is defined using PEAR::HTML_TagCLoud function.

(2) FONT COLOR We designed font colors for the tags based on the information sources in which they appear. We chose red, blue, and green as the basic colors [1]. Red is the most eye-catching color. If term *t* appears both in the calendars and tweets, it must be important, and therefore red is assigned to it. Since Twitter's basic color is blue, if term *t* appears only in Twitter, blue is assigned to it. If term *t* appears only in the calendars, the remaining color green is assigned to it.

(3) TAG ORDER Since we believe that the time order is important for recalling memories, the tags are sorted by the time of the term that first appeared.

4 Tag Browser

We show a complete image of our system used in Experiment 5 in Figure 6. We call the system *Tag browser*.

Users can input a period (start and end dates) to display a tag cloud, calendar events, and tweets for it. The number of tags can be selected from five choices: 10, 30, 50, 70, or 100. We set the default value to 30, based on a previous work [8]. Users can also change the tag weight by a slider (0.0-1.0). The number corresponds to α in Eq. 3. If the user selects 0.0, a tag cloud is generated only by tweets, and if 1.0 is chosen, it is generated only by calendar events. The default value is 0.5, which means the ratio of calendar events and tweets is 1:1. Outputs are displayed in three parts: a tag cloud, a calendar part, and a Twitter part. The tag cloud part displays a tag cloud, the calendar part displays a calendar's event titles, and the Twitter part displays tweets. Users can also access detailed contexts by clicking on tags.

Figure 6 shows a user's screen from August 1 to 31 2014. *Kaiyukan* (aquarium) is displayed as the largest text and in red, *Namba* (place) is the second largest display in green, and *spitted cutlet* is the third largest in blue in the tag cloud. These tags are enlarged since some photos have them. Since *Kaiyukan* appeared both in the calendar events and in the tweets, it is highlighted in red. With just a glance, he recalls a fun memory at Kaiyukan and Namba from his August memory. Although *Namba* only appeared once in the calendar

[1] In the initial prototype [7], we used different colors: orange for calendar only and light blue for Twitter only, as in the Figures 2 and 3. Experiments 1-4 were conducted using the initial prototype, and Experiment 5 was conducted using the complete system.

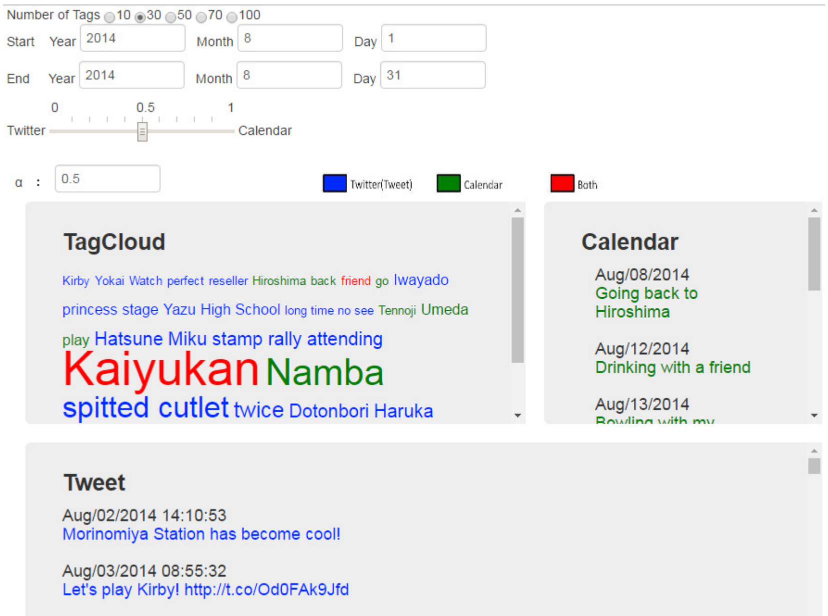

Figure 6: Tag browser: complete image of system

Figure 7: Tag cloud created only from calendar events

events and *spitted cutlet* only appeared twice in tweets, our system selected these terms based on the number of photos and displayed them together based on time sorting. Since the tag cloud is sorted by time, the first tag is *Kirby* (game), which is included in the August 3 tweets.

Figure 7 shows a tag cloud just created from a calendar where the user set the system parameter to 1. Figure 8 shows a tag cloud just created from Twitter where the user set the system parameter to 0.

5 Experiment

5.1 Overview

We recruited subjects under the following conditions: (a) they write schedules based on their media choice, (b) they have posted tweets for more than three months on Twitter,

Figure 8: Tag cloud created only from tweets

and (c) they take photos on digital devices. Our 21 subjects were all males, aged 19-25 (average 22.7 years old, SD=1.78). 16 were undergraduate or graduate students, and five subjects worked. The subjects who did not use Google Calendar copied their schedules onto it. Twelve subjects were allocated to Experiments 1-3 (continuously done) and 10 to Experiments 4 or 5. The subjects for Experiments 1-3 are the same. We used one month-long data for recalling their memories.

5.2 Experiment 1

We evaluated the usefulness of our algorithm that extracts and weights the terms.

5.2.1 Method

We extracted the top 30 terms by two algorithms from the usage of one month: our method and a comparative method. Our method is described in Section 3.2. Weighting function $Comparative(t)$ for term t in the comparative method is defined as follows:

$$Comparative(t) = \sum_{t \in HS} tf(t). \tag{5}$$

$Comparative(t)$ is calculated by the frequency of the term occurrence. That is, the more it appears, the more its weight increases.

For both methods, we combined the extracted 60 terms and sorted them alphabetically to form a list so that the subjects cannot guess the weighting algorithms. Our subjects evaluated whether the terms on the list helped them recall their memories at the following five levels: 5: very useful; 4: useful; 3: neutral; 2: not very useful; 1: not useful.

5.2.2 Results and Analysis

The data obtained by each subject are shown in Table 1. The average number of calendar events was 11.8 and 250.8 for the tweets. The average number of different terms for the calendar events was 10.5 and 391.2 for the tweets. The average number of photos was 35.3.

Table 2 shows the top terms for each subject in Experiment 1. More named entities such as *Nagasaki* (place), *Jingu Stadium*, *Moss* (shop), and *Icho Festival* (university festival) are included in our method. In contrast the top term for the four subjects in the comparative method were the same: *part-time job*. Generally, our method outputs impressive events, but the comparative methods output repetitive and persistent daily activities.

Table 1: Basic data obtained in Experiment 1

Subject	Events (Diffrent terms)	Tweets (Different terms)	Photos
Subject 1	15 (7)	14 (28)	19
Subject 2	20 (8)	224 (462)	73
Subject 3	16 (25)	36 (50)	41
Subject 4	13 (15)	332 (735)	27
Subject 5	3 (3)	15 (49)	124
Subject 6	3 (1)	62 (51)	5
Subject 7	43 (43)	198 (554)	28
Subject 8	3 (4)	251 (554)	25
Subject 9	2 (2)	30 (78)	37
Subject 10	1 (1)	22 (51)	20
Subject 11	20 (14)	1628 (1696)	15
Subject 12	2 (3)	197 (439)	9
Mean	11.8 (10.5)	250.8 (391.2)	35.3

Table 2: Top terms in each subject in Experiment 1

Subject	Top term by proposed method	Top term by comparative method
Subject 1	cake [5]	part-time job [5]
Subject 2	Nagasaki (place) [5]	part-time job [1]
Subject 3	fireworks display [3]	part-time job [3]
Subject 4	lunch [5]	hatebu (social bookmark site) [3]
Subject 5	sister [5]	camp [5]
Subject 6	Vietnam trip [5]	Vietnam trip [5]
Subject 7	Jingu Stadium [5]	part-time job [1]
Subject 8	salad bread [5]	Yahoo [3]
Subject 9	Kyoto (place) [5]	— [1]
Subject 10	Moss (shop) [5]	Moss (shop) [1]
Subject 11	Icho Festival (university festival) [5]	Anetai (anime character) [5]
Subject 12	e [1]	lecture [1]

Note: [] means evaluation values.

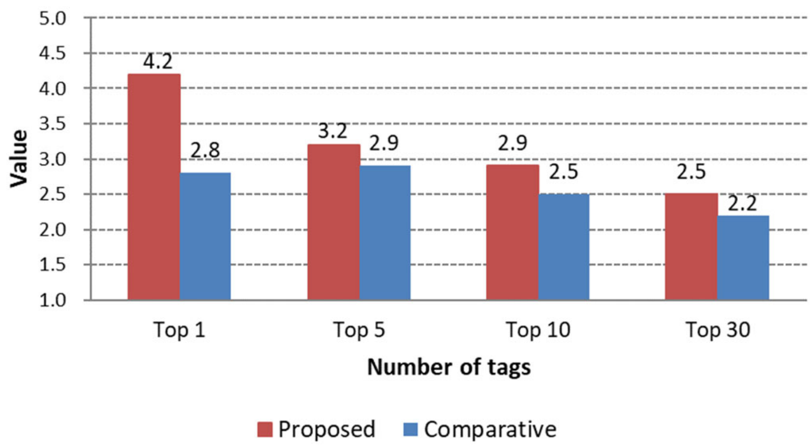

Figure 9: Average values of each method in Experiment 1

Table 3: Combination of methods and information sources in Experiment 2

	Calendar + Twitter	Calendar	Twitter
Proposed	A	B	C
Comparative	D	E	F

The average values of each algorithm for the top 1, 5, 10, and 30 terms are shown in Figure 9. We used the Wilcoxson signed-rank test to compare our method and the comparative one for each top 1, 5, 10, and 30 terms. Although not statistically significant except for the top term ($Z = 2.023, p < .05$), our method outperformed the comparative method.

These results suggest the basic usefulness of our weighting algorithm using the number of photos and two information sources.

5.3 Experiment 2

We evaluated the usefulness of our algorithm to create tag clouds.

5.3.1 Method

We prepared six tag clouds that display 30 terms for comparison: (a) our method and a comparative method and (b) information sources (calendar + Twitter, calendar, Twitter). The following are the six tag clouds (Table 3). Tag cloud A uses our weighting method and is composed of calendar and Twitter (our tag cloud). Tag clouds B-F are comparative tag clouds. Tag cloud B uses our weighting method and is only composed of calendar. Tag cloud C uses our weighting method and is only composed of Twitter. Tag cloud D uses the comparative weighting method and is composed of calendar and Twitter. Tag cloud E uses the comparative weighting method and is only composed of calendar. Tag cloud F uses the comparative weighting method and is only composed of Twitter.

The following are the questions:

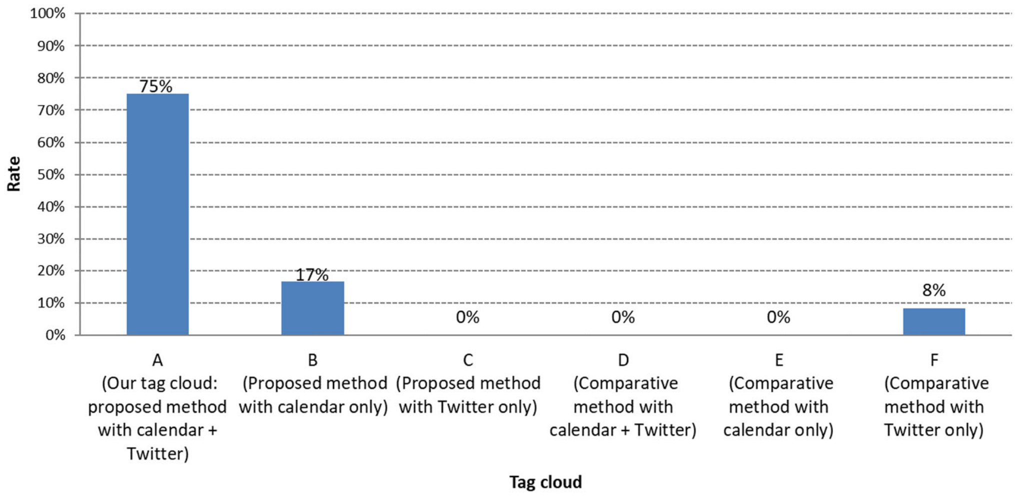

Figure 10: Most useful tag clouds in Experiment 2 (Q6)

Q1: Which is more useful to recall your memories, A or D?
Q2: Which is more useful to recall your memories, B or E?
Q3: Which is more useful to recall your memories, C or F?
Q4: Which is the most useful to recall your memories, A, B, or C?
Q5: Which is the most useful to recall your memories, D, E, or F?
Q6: Which of the six tag clouds is the most useful to recall your memories?

Qs 1-3 compare our method and the comparative method. Q4 and Q5 compare the source with each method. Q6 compares all of them.

5.3.2 Results and Analysis

For Q1 to Q3, we performed a chi-square goodness of fit test. For Q1, 92% (11/12) selected A and 8% (1/12) selected D ($\chi^2(1) = 8.333, p < .01$). For Q2, 92% (11/12) selected B and 8% (1/12) selected E ($\chi^2(1) = 8.333, p < .01$). For Q3, 83% (10/12) selected C and 17% selected F ($\chi^2(1) = 5.333, p < .05$). For Q4 to Q6, Friedman's test was performed. Since the subjects were only asked about the best tag clouds, 1 (rank) was assigned to them, and 2 was assigned to the remainder. Scheffe's method was used for multiple comparisons. For Q4, 75% (9/12) selected A and 25% (3/12) selected B ($\chi^2(2) = 7.875, p < .01$). There was a significant difference between A and C ($p < .01$). For Q5, 50% (6/12) selected D, 42% (5/12) selected E, and 8% (1/12) selected F (*ns*). For Q6, 75% (9/12) selected A, 17% (2/12) selected B, and 8% (1/12) selected F ($\chi^2(5) = 13.286, p < .05$), (Figure 10). There were significant differences between A and B ($p < .05$) as well as A and C, D, E, F ($p < .01$). Our algorithm outperformed the comparative method regardless of the information source, based on the results of Qs 1-3. From Q6, combined tag cloud A with our method and two sources was superior to the other tag clouds. However, tag cloud B (our method with calendar only) was also selected by two subjects in Q6, and there was no significant difference between tag clouds A and B in Q5.

Concerning the Q6 result, the following are the reasons for choosing our tag cloud A: "More important words and different colors were displayed"; "It was easy to visually recall

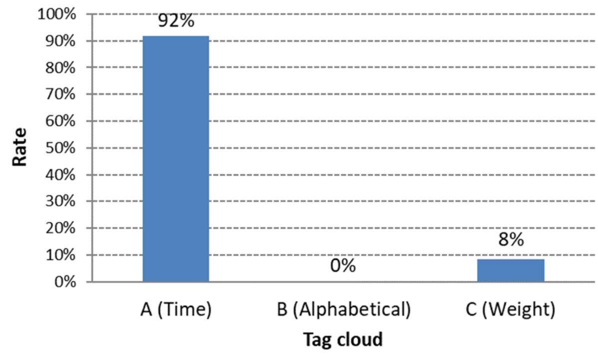

Figure 11: Most useful order of tags in Experiment 3 (Q2)

events"; "Tag cloud A was easy for recalling since the size and the color of the words were more diverse"; "The size and color of the words were different and easy to see"; and "I liked it because it had many words in large fonts." Tag cloud B only received one positive reason: "It had many keywords for impressive events, and they were displayed in a large font and precisely." Two negative comments (whose contents were basically the same) about tag cloud A were found in the reasons for other questions: "It had too many useless words." Tag cloud B received no negative comments, probably because the history structure of the calendar events included fewer useless terms than tweets.

Overall, the quality of our tag cloud A was the best among the tag clouds, although it contained some useless terms derived from Twitter.

5.4 Experiment 3

Next we evaluated the usefulness of the sorting algorithm to create tag clouds.

5.4.1 Method

We prepared three tag clouds that displayed 30 terms weighted by our proposed algorithm using calendar and Twitter for comparison: (a) sorting by time (tag cloud A, our tag cloud), (b) by character code (i.e., alphabetically, tag cloud B), and (c) weight (large to small, tag cloud C).

The following are the questions:

Q1: Rank the three tag clouds by the ease with which they helped you recalled your memories: A, B, or C.

Q2: Which is the most useful to recall your memories, A, B, or C?

5.4.2 Results and Analysis

We performed Friedman's test and assigned 1-3 ranks for Q1. For Q2, 1 was assigned to the best tag cloud and 2 was assigned to the rest of the tag clouds. We used Scheffe's method for multiple comparisons. For Q1, 75% selected A (9/12) and 25% (3/12) selected C for the first rank ($\chi^2(2) = 9.500, p < .01$). There was a significant difference between A and B ($p < .01$). For Q2, 92% (11/12) answered that A is best for recalling their memories, 8%

(1/12) selected C, and no subjects selected B ($\chi^2(2) = 18.500, p < .01$), (Figure 11). There was a significant difference between A and B or A and C (both $p < .01$).

For the Q2 result, the following are the reasons for choosing our tag cloud A: "The order fits the memory flow, and it's good that important things are displayed in a large font"; "I felt that I could recall various things just looking around the large words"; "A group of large words is not good because it catches the eye" (negative comments for C); "The order helps recall"; "It was easy to see and the words about daily activities were related." The only subject who selected tag cloud C made the following comment: "The system was good because the large words came first." Therefore, tag cloud A (sorting by date) is our first choice for this purpose.

From Experiments 1, 2, and 3, we found that our tag clouds, where terms were extracted from calendar events and tweets, weighted using the number of photos, and sorted by time, are basically useful for recalling memories. Since we also found that a few subjects deemed either calendar or Twitter only more beneficial than using both information sources, we investigated how the useful parameters varied by users in our next experiment.

5.5 Experiment 4

We evaluated our default parameter (1:1) for scoring calendars and Twitter in Eq. 3 and also how the useful parameters vary by users.

5.5.1 Method

The subjects can modify their tag clouds by selecting parameter α from 0.0 (i.e., information source is Twitter only) to 1.0 (i.e., calendar only) by 0.1. When the parameter was 0.5, the weight of calendars and Twitter is the same. After the subjects displayed the eleven tag clouds in whatever order they chose, they answered the following question:

Q: Choose and rank the three tag clouds which are more useful for recalling your memories.

Since the differences among each tag cloud is subtle and we thought that choosing just one tag cloud was difficult, we asked them to rank three tag clouds.

5.5.2 Results and Analysis

Table 4 shows the results of Experiment 4. The average values were around 0.5 to 0.6. The results suggest the basic appropriateness of default parameter 0.5 and that the best parameter varies by user from 0.0 to 0.9.

We classified the answers (rank 1) and (all) into three categories: 0.0-0.3, 0.4-0.6, and 0.7-1.0 (Table 5) and performed a chi-square goodness of fit test. For the rank 1 parameters, there was no significant difference ($\chi^2(2) = 3.800, p = .15, ns$), and for all the answers, there was a significant difference ($\chi^2(2) = 8.600, p < .05$).

We therefore found that the default parameter is appropriate for our tag clouds. In addition, since the best parameter varies by user, providing a facility for changing parameters based on user desires is good.

Experiments 1-4 show the usefulness of the tag cloud created by the proposed method. Next we evaluated the usefulness of our complete tag browser system.

Table 4: Selected parameters in Experiment 4

Subject	1st	2nd	3rd
Subject 1	0.5	0.8	0.1
Subject 2	0.6	0.5	0.4
Subject 3	0.5	0.6	0.7
Subject 4	0.9	0.5	0.6
Subject 5	0.9	0.8	0.7
Subject 6	0.5	0.4	0.6
Subject 7	0.5	0.6	0.4
Subject 8	0.6	0.7	0.8
Subject 9	0.2	0.3	0.0
Subject 10	0.9	0.5	0.4
Mean	0.6	0.6	0.5

Table 5: Classified selected parameters in Experiment 4

	0.0-0.3	0.4-0.6	0.7-1.0
1st	1 (10%)	6 (60%)	3 (30%)
All	4 (13%)	17 (57%)	9 (30%)

5.6 Experiment 5

We evaluated the usefulness of our complete system.

5.6.1 Method

We developed three prototype systems: A, B, and C. System B is our proposed system with a tag cloud created by our method with lists of calendar events and tweets (see Tag browser in Figure 6). System A has lists of calendar events and tweets (with no tag cloud). System C has a tag cloud created by the comparative method and lists the calendar events and tweets. Our aim is to evaluate System B. For this purpose, first we compared systems A and B, and next we compared Systems B and C to evaluate the basic usefulness of System B. In the final step, we asked the subjects to evaluate system B in detail.

The following are the questions for comparison:

Q1: Which system is easier to explain your memories, A or B?

Q2: Which system is more useful to recall your memories, A or B?

Q3: Which system is more useful to recall your memories, B or C?

Table 6 shows questions 4-9 and the results of a five-point evaluation (1-5) of our proposed system. Q4 to Q6 address the usefulness for each part, and Q7 to Q9 address the whole system.

Finally, Q10 asks this question: What are the positive and negative aspects of the system?

Table 6: Questions for prototype systems in Experiment 5

Question		Mean	SD
Q4	Was tag cloud useful for recalling your memories?	4.3	0.9
Q5	Was the calendar display useful for recalling your memories?	3.6	1.3
Q6	Was the twitter display useful for recalling your memories?	4.3	1.1
Q7	Was this system useful for recalling your memories?	4.4	1.0
Q8	How did you feel after using the system?	4.2	0.6
Q9	Do you want to use this system in the future to recall your memories?	3.9	0.9

Table 7: Comments for prototype systems in Experiment 5 (Q10)

(a) Subjects selected our system B in Q3

Positive comments
"I liked being able to adjust the tag cloud by the slider. The tag clouds are easy to see."
"The color of the tag cloud is easy to see."
"Since tags with many photos are displayed in a very large font, pleasant memories can be recalled more clearly."
"The tag clouds are easy to see due to emphasizing words, with no unnatural connections."
"Terms about good memory are displayed."
Negative comments
"The display is somewhat difficult to see."
"It was difficult to set the slider to 0 or 1."
"Although negative words seem hidden, by clicking tags, they appeared. When clicking on the tags, the scrolls of the calendar and tweets displays are reset."
"The tags are limited to the actual words that appeared. The number of tags was too small."
"I did not look at the tweet display part."

(b) Subjects selected comparative system C in Q3

Positive comments
"It is good that common tags are displayed in red."
Negative comments
"The tweet displays are too small for people who often tweet."
"For me, the timing of the tweeting and taking photos does not overlap."

5.6.2 Results and Analysis

For Q1 to Q3, we performed a chi-square goodness of fit test. For Q1, eight subjects answered System B (80%, $\chi^2(1) = 3.600, p = .06, ns$). For Q2, all subjects answered System B (100%, $\chi^2(1) = 10.000, p < .01$). For Q3, eight subjects answered System B (80%, $\chi^2(1) = 3.600, p = .06, ns$). Tag clouds are basically useful to recall user memories, and our system is more useful than any other system.

For Q4-Q9, the average values exceeded 3.6. For Q7-Q9, only one subject answered less than 3 (actually 2), and other subjects answered equal or more than 3 for all of the questions. If we omit the above subject who answered 2 for Q7 and Q9, the average values for Q7 become 4.8 and 4.1 for Q9.

Table 7 shows all the Q10 results. Most of the negative comments concerned the system's user-interface and not the tag cloud itself.

Overall, the above results support the basic usefulness of our proposed system. Our tag cloud system is useful for recalling memories and the subjects felt good after using it.

6 Related Work and Discussion

6.1 Related Work

This research is a part of a project that helps users construct "externalized-memory [9]," which is a concept that virtually externalizes and stores the contents of human working memory. Murakami [2] presented a concept of information structure called history structure, which is a subset of externalized-memory, constructed from time, keywords, and URI sets. The history structure integrates various kinds of information usage. Murakami and Hirata [10] created an interest-space browser from web browsing history by generating a two-dimensional term space called interest space. Murakami et al. [11] created a knowledge-space browser from five information usages (web browsing, calendars, Twitter, e-mail, and book purchases) by generating a term network called knowledge space. The following are the two main differences between our previous research and this research: (1) we selected two important information sources (calendars and Twitter) for memory recall and (2) we presented new algorithms for weighting terms using the number of photos and the displayed tag clouds. We presented our idea and an initial prototype as a position paper in a conference [7]. In our paper, we presented a complete system that has the ability to change the weight of tags and described in detail the effectiveness of the proposed method and system through five experiments. After the initial prototype, Murakami and Murakami [12] developed a different version of a tag-cloud-based approach using multiple SNSs such as LINE and Twitter. The algorithms of generating tag clouds are different and their paper includes limited experiments.

The idea of "externalized-memory" itself has been investigated by other researchers. For example, Huang et al. [13] proposed personal image repositories that capture user's memories as externalized memory spaces. They placed personal photos collections in a semantically meaningful layout for image information retrieval. We use the number of photos to weight the keywords for memory recall.

Our research uses tag clouds for human memory recall, even though little research uses them for this purpose. Here are some exceptions. Chen and Jones [14] developed a prototype system called iCLIPS that searches through personal lifelogs for memory support. Their lifelog data include 20 months of data, including visual capture of the physical world events with Microsoft SenseCams, full indexing of accessed information on computers and mobile phones, and context data, including location via GPS and people with Bluetooth. In iCLIPS, computer activities and the names of locations and people are displayed in term clouds. No detailed algorithms for generating term clouds and user studies of the prototype have been reported. We focus on two information sources (calendars and Twitter) to create better tag clouds rather than accumulating all human activities. We also conducted user studies of our prototype and showed our system's usefulness. Mathur et al. [15] 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 is displayed. Our approach automatically generates tags from calendars and Twitter and generates tag clouds from a mixture of generated tags. Aiordachioae and Vatavu [16] introduced a wearable, smartglasses-based system for abstracting life in the form of clouds of tags and concepts called "Life-Tags," which summarize users' life experiences using word clouds and highlight an "executive summary." They simply used third-party services to extract tags from images. We extract tags from calendars and Twitter.

The lifelog idea can be traced back at least 75 years to a famous hypertext system called Memex [17]. Its vision is that technology will allow us to capture everything that ever happened to us, to record every event we ever experienced, and to save every bit of information we have ever touched [18][19]. MyLifeBits [20] is one of the most well-known lifelog research projects. Recently, lifelog research on multimedia data has become active in the information retrieval community. For example, the NTCIR Lifelog task aims to advance the state-of-the-art in lifelogging as an application of information retrieval [21] and to encourage research into the organization and retrieval of data from multimodal lifelogs [22]. The Lifelog Search Challenge (LSC) is an annual comparative benchmarking activity for comparing approaches to interactive retrieval from multimodal lifelogs [23]. In typical lifelog research, lifeloggers wear computers that capture what they have seen and heard, such as SenseCam. The majority of lifelog research has been conducted in the context of multimedia or multimodal processing. Our approach uses calendar and Twitter texts without any special devices to capture information from the real world. Our work can be viewed as text-based lifelog research. Since manual tagging of lifelogs is time-consuming, we automatically extract keywords from calendar events and tweets.

If we explore research for searching personal calendar events or tweets, our research is related to personal information management (PIM) [24][25]. Aires and Gonçalves [26] 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. Since we previously investigated that just mixing everything (various kinds of information sources) causes "junk" keywords, we concentrated on calendars and Twitter to produce good tag clouds for memory support. In PIM research, since the primary target is information rather than memory, evaluation tends to determine whether the information can be retrieved. Our research examined whether memory can be recalled instead of information (e.g., calendar events or tweets).

There is growing attention in the HCI community on how technology could be designed to support experiences of reminiscence on past life experiences [27]. People use a number of media and methods to support reminiscing, especially photos, which both record memories and share remembered experiences [28]. Since photos are one of the most important tools for reminiscence strength, our idea is to use the number of photos to find impressive events. A number of systems have been developed to support reminiscence. Pensieve [29][28] is one such system that supports everyday reminiscence by emailing memory triggers to people that contain either social media content they previously created on third-party websites or text prompts about common life experiences. Like Pensieve, most systems for reminiscence use written journals as triggers. Our research uses calendar events and tweets without prompting users to create tags or memory triggers to avoid special effort by the system.

Many systems and researches have created tag clouds. The two main purposes are navigating or summarizing certain content. Basically, tag selection algorithms are based on the frequency of objects assigned to terms in tag clouds and the frequency of terms included in the documents in word clouds.

Torres-Parejo et al. [30] described general tag selection algorithms in tag clouds. Their survey introduced a study by Skoutas and Alrifai [31] that compared different tag selection strategies: (1) frequency-based, (2) diversity-based, and (3) ranking of aggregation based. The simplest strategy (frequency-based: number of objects assigned to tags) works well and diversity-based strategies (that use tags to cover many objects) achieved the best performance. Venetis et al. [8] evaluated existing algorithms for exploring and understanding a set of objects against tf-idf-based algorithms and concluded that a maximum covering algorithm (COV) seems to be a very good choice for most scenarios. Another good choice is the popularity algorithm (POP), which is easier to implement and performs well in specific contexts. Both works [31][8] emphasize the usefulness of frequency (popularity) and diversity (coverage) strategies for tag selection algorithms.

Our research presented a unique algorithm based on the frequency of photos and terms. Since our tag cloud resembles a word cloud rather than a typical tag cloud in the sense that we treat a certain amount of calendar texts and tweets as a document, we compared our algorithms with term frequency algorithms. Our algorithm has the following benefits compared to simple term frequency methods. The number of photos very effectively helps users recall their memories since they emphasize impressive events. Our system can be used even when calendar events or tweets do not exist. Even if users do not take pictures, our algorithm works with raw tf for calendars plus normalized tf for Twitter. In addition, users can change the calendar and Twitter ratios to adapt user behaviors.

Since Twitter has a retweet function that usually reflects a tweet's interestingness, the terms in retweeted tweets can be weighted. Zhao et al. [32] proposed a topical keyphrase extraction method that considers such interestingness using retweets to summarize Twitter. Because our purpose is not to summarize general Twitter content, retweets are not useful to support personal memory since this behavior informs others of interesting tweets; we ignored retweets.

Tag clouds can be visualized by different designs: by the order of the tags (alphabetically, semantically, by frequency, etc.) or the shape of the tag cloud (cubes, circles, tags in sequential lines, etc. [30]. The most popular is the 'classic' rectangular tag arrangement with alphabetical sorting in a sequential line-by-line layout [33]. Our research shows that the classical tag cloud that sorts by time is useful for memory recall support.

6.2 Discussion

The following are our paper's main contributions. First, we presented algorithms that extracted keywords from the use of two information sources, calendars and Twitter, and generated tag clouds using the frequency of the keywords and the number of photos related to those keywords to recall impressive memories. Second, we developed a tag browser to support human recall based on this approach. Third, we evaluated whether our system helps users recall a particular month. Experimental results revealed the usefulness of our approach and implemented system.

Our experiments revealed the following: (1) We examined a combination of weighting methods and information sources and found that our tag cloud (using our weighting method with calendar and Twitter data) is superior to other tag clouds. (2) We examined three tag sorting methods (time, alphabetical, and weight) and concluded that sorting by time outperforms the others. (3) Our implemented prototype system is useful for recalling user memories.

From an academic point of view, our research is positioned, within the multimedia and multimodal mainstream, as text-based lifelog research. Even if the source of information is only text, annotating a lifelog is a time-consuming and labor-intensive task. We proposed a method to automatically extract keywords from the text generated in the user's daily activities. In addition, we proposed a method of using the number of photos taken by users as a way to find important keywords in the text. We believe this is a promising approach to overcome the limitations of extracting important terms in natural language processing.

To enlarge our approach's applicability, we only gathered essential information for both calendars and Twitter: time and text contents. Basically, any schedules on any media and any SNS messages such as Facebook or Instagram can be the source for our system. In this research, we implemented two tools to generate history structures from the schedule entries of Google Calendar and tweets. Those who do not use Google Calendar can copy the times and events of their schedules to it.

Future work will investigate the following three aspects. First, we need to improve our algorithms for generating keywords and tag clouds. In particular, when the number of photos is large, tweets sometimes contain useless keywords. Offering a facility for choosing an algorithm with or without photos might be useful for users whose photo-taking activities are not always related to their memorable events. Second, we need to improve our user-interface so that it is easy to use, for example, by enlarging the tweet display area. Third, we must examine the system for different periods, such as a week or a year.

Finally, let us explain the contributions of our work to future research from a social perspective. The purpose of this study is to support human memory in everyday life, including both work and private life. In the text, the examples used were biased toward those of private life, since the subjects of the experiment were mainly students. However, our system can be used by anyone who uses a calendar and SNS platforms, regardless of age or occupation. In addition, it can be used for a variety of work purposes that require memory of the past. Although the sources of information were limited to calendars and Twitter in this paper, basically any text with time information can be added to the system, and the system can be applied to a wide variety of work situations. These are issues to be further explored in the future.

7 Conclusions

We presented a system that supports users' memory recall by creating tag clouds from calendar and Twitter data. The main feature of this research is that we use the number of photos taken by the user to recall impressive events. We evaluated the tag cloud generation algorithms and our implemented prototype system and found the following. (1) Our tag cloud (using our weighting method with calendar and Twitter data) is superior to the other tag clouds. (2) Sorting by time outperforms other sorting methods. (3) The implemented prototype system is useful for recalling user memories.

Future work will investigate improving our algorithms and the user-interface and evaluate the system for different periods. Adding other text with time information and using the system in a wider variety of work situations are also issues to be further explored.

References

[1] M. Harvey, M. Langheinrich, and G. Ward, "Remembering through lifelogging: A survey of human memory augmentation," *Pervasive and Mobile Computing*, vol. 27, pp. 14–26, 2016, doi: 10.1016/j.pmcj.2015.12.002.

[2] H. Murakami, "History structure for exploring desktop data," in *Proc. SIGIR 2010 Workshop on Desktop Search (Understanding, Supporting and Evaluating Personal Data Search)*, 2010, pp. 25–26.

[3] A. W. Rivadeneira, D. M. Gruen, M. J. Muller, and D. R. Millen, "Getting our head in the clouds: Toward evaluation studies of tagclouds," in *Proc. SIGCHI Conf. Human Factors in Computing Systems (CHI '07)*, 2007, pp. 995–998, doi: 10.1145/1240624.1240775.

[4] M. J. Halvey and M. T. Keane, "An assessment of tag presentation techniques," in *Proc. 16th Int'l Conf. World Wide Web (WWW '07)*, 2007, pp. 1313–1314, doi: 10.1145/1242572.1242826.

[5] D. Helic, C. Trattner, M. Strohmaier, and K. Andrews, "Are tag clouds useful for navigation? A network-theoretic analysis," *Int'l Journal of Social Computing and Cyber-Pyhsical Systems*, vol. 1, no. 1, pp. 33–55, 2011, doi: 10.1504/IJSCCPS.2011.043603.

[6] M. Matsumoto, "Supporting human recollection of the impressive days using the number of photos," Master's Thesis, Graduate School for Creative Cities, Osaka City University (in Japanese), 2014.

[7] M. Matsumoto, S. Matsuura, K. Mitsuhashi, and H. Murakami, "Supporting human recollection of the impressive events using the number of photos," in *Proc. 6th Int'l Conf. Agents and Artificial Intelligence - Volume 1:ICAART*, 2014, pp. 538–543, doi: 10.5220/0004902805380543.

[8] P. Venetis, G. Koutrika, and H. Garcia-Molina, "On the selection of tags for tag clouds," in *Proc. 4th ACM Int'l Conf. Web Search and Data Mining (WSDM '11)*, 2011, pp. 835–844, doi: 10.1145/1935826.1935855.

[9] H. Murakami and T. 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.

[10] H. Murakami and T. Hirata, "A system for generating user's chronological interest space from web browsing history," *Int'l Journal of Knowledge-based and Intelligent Engineering Systems*, vol. 8, no. 3, pp. 149–160, 2004, doi: 10.3233/KES-2004-8303.

[11] H. Murakami, K. Mitsuhashi, and K. Senba, "Creating user's knowledge space from various information usages to support human recollection," in *Advanced Research in Applied Artificial Intelligence (IEA/AIE 2012)*, H. Jiang, W. Ding, M. Ali, and X. Wu, Eds., 2012, pp. 596–605, doi: 10.1007/978-3-642-31087-4_61.

[12] H. Murakami and R. Murakami, "A system using tag cloud for recalling personal memories," in *Advances and Trends in Artificial Intelligence. From Theory to Practice (IEA/AIE 2019)*, F. Wotawa, G. Friedrich, I. Pill, R. Koitz-Hristov, and M. Ali, Eds., 2019, pp. 398–405, doi: 10.1007/978-3-030-22999-3_35.

[13] H.-H. Huang, Y. Sumi, and T. Nishida, "Personal image repositories as externalized memory spaces," *Int'l Journal of Knowledge-based and Intelligent Engineering Systems*, vol. 10, no. 2, pp. 169–180, 2006, doi: 10.3233/KES-2006-10206.

[14] Y. Chen and G. J. F. Jones, "Augmenting human memory using personal lifelogs," in *Proc. 1st Augmented Human Int'l Conf. (AH '10)*, 2010, paper 24, doi: 10.1145/1785455.1785479.

[15] A. Mathur *et al.*, "Lifeview: A lifelog visualization tool for supporting sentimental recall and sharing," in *Proc. 24th Australian Computer-Human Int'l Conf. (OzCHI '12)*, 2012, pp. 371–380, doi: 10.1145/2414536.2414596.

[16] A. Aiordachioae and R.-D. Vatavu, "Life-tags: A smartglasses-based system for recording and abstracting life with tag clouds," in *Proc. ACM Hum.-Comput. Interact.*, vol. 3, no. EICS, 2019, paper 15, doi: 10.1145/3331157.

[17] V. Bush, "As we may think," *The Atlantic Monthly*, vol. 176, no. 1, pp. 101–108, 1945.

[18] A. J. Sellen *et al.*, "Do life-logging technologies support memory for the past? an experimental study using sensecam," in *Proc. SIGCHI Conf. Human Factors in Computing Systems (CHI '07)*, 2007, pp. 81–90, doi: 10.1145/1240624.1240636.

[19] A. J. Sellen and S. Whittaker, "Beyond total capture: A constructive critique of lifelogging," *Commun. ACM*, vol. 53, no. 5, pp. 70–77, 2010, doi: 10.1145/1735223.1735243.

[20] J. Gemmell, G. Bell, R. Lueder, S. Drucker, and C. Wong, "Mylifebits: Fulfilling the memex vision," in *Proc. 10th ACM Int'l Conf. Multimedia (MULTIMEDIA '02)*, 2002, pp. 235–238, doi: 10.1145/641007.641053.

[21] "NTCIR16 lifelog task." [Online]. Available: http://ntcir-lifelog.computing.dcu.ie/

[22] C. Gurrin *et al.*, "Experiments in lifelog organisation and retrieval at NTCIR," in *Evaluating Information Retrieval and Access Tasks*, T. Sakai, D. Oard, and N. Kando, Eds., 2021, pp. 187–204, doi: 10.1007/978-981-15-5554-1_13.

[23] C. Gurrin *et al.*, "Introduction to the third annual lifelog search challenge (LSC'20)," in *ICMR '20: Proc. 2020 Int'l Conf. Multimedia Retrieval*, 2020, pp. 584–585, doi: 10.1145/3372278.3388043.

[24] W. Jones, "Personal information management," *Annual Review of Information Science and Technology*, vol. 41, no. 1, pp. 453–504, 2007, doi: 10.1002/aris.2007.1440410117.

[25] W. Jones, *The future of personal information management, part I: our information, always and forever*. Morgan & Claypool Publishers, 2012, doi: 10.2200/S00411ED1V01Y201203ICR021.

[26] J. Aires and D. Gonçalves, "Personal information dashboard - me, at a glance," in *PIM workshop 2012*, 2012.

[27] M. Yoo, W. Odom, and A. Berger, "Understanding everyday experiences of reminiscence for people with blindness: Practices, tensions and probing new design possibilities," in *Proc. 2021 CHI Conf. Human Factors in Computing Systems (CHI '21)*, 2021, paper 167, doi: 10.1145/3411764.3445212.

[28] S. T. Peesapati *et al.*, "Pensieve: Supporting everyday reminiscence," in *Proc. SIGCHI Conf. Human Factors in Computing Systems (CHI '10)*, 2010, pp. 2027–036, doi: 10.1145/1753326.1753635.

[29] D. Cosley *et al.*, "Using technologies to support reminiscence," in *Proc. 23rd British HCI Group Annual Conf. People and Computers: Celebrating People and Technology (BCS-HCI '09)*, 2009, pp. 480–484.

[30] U. Torres-Parejo, J. R. Campaña, M. A. Vila, and M. Delgado, "A survey of tag clouds as tools for information retrieval and content representation," *Information Visualization*, vol. 20, no. 1, pp. 83–97, 2021, doi: 10.1177/1473871620966638.

[31] D. Skoutas and M. Alrifai, "Tag clouds revisited," in *Proc. 20th ACM Int'l Conf. Information and Knowledge Management (CIKM '11)*, 2011, pp. 221–230, doi: 10.1145/2063576.2063613.

[32] W. Zhao *et al.*, "Topical keyphrase extraction from Twitter," in *Proc. 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies (HLT '11)*, 2011, pp. 379–388.

[33] S. Lohmann, J. Ziegler, and L. Tetzlaff, "Comparison of tag cloud layouts: Task-related performance and visual exploration," in *Human-Computer Interaction - INTERACT 2009*, T. Gross *et al.*, Eds., 2009, doi: 10.1007/978-3-642-03655-2_43.