

Supporting Human Recollection of the Impressive Events using the Number of Photos

Masaki Matsumoto¹, Sho Matsuura¹, Kenta Mitsuhashi² and Harumi Murakami¹

¹Graduate School for Creative Cities, Osaka City University, Osaka, Japan

²OGIS-RI Co., Ltd., Osaka, Japan

{07matsumoto.m, massyoukai, kenta.mitsuhashi}@gmail.com, harumi@media.osaka-cu.ac.jp

Keywords: Recollection, Photo Data, Tag Cloud, Twitter, Google Calendar.

Abstract: We present a system to support human recollection with tag clouds, which are created from keywords generated by our algorithms from the use information of Google Calendar and Twitter. The main feature of our research is to weight words using the number of photos taken by users to recall impressive events. We evaluated tag clouds by comparing our approach and a comparative approach, and our experiment results suggest the usefulness of our approach.

1 INTRODUCTION

The human memory is not certain. As time passes, our ability to recall past memories may weaken. Furthermore, the amount of information that we manage is increasing due to the spread of the internet. Managing our memories and information is very difficult. Against this background, lifelog studies are increasing that capture the lives of individuals as well as personal information management (PIM) schemes that categorize such personal information as schedules and address books.

Before computers and the internet's growth, people wrote plans on paper media (e.g., memo pads or calendars) to manage their schedules. As PDAs, cell phones, and smartphones proliferate, the number of people using digital media to manage personal information has increased. Online, web-based calendars are also being used. Google Calendar is one of the most popular web-based calendars.

Social Networking Services (SNS), through which users can send and share information with others, have also grown. The spread of cell phones and smartphones is one remarkable aspect of the popularity of SNSs. Users can easily send and view information even when they are away from their computers. Twitter became popular because users can post short messages anytime from anywhere and easily connect with others. Twitter users can immediately post their thoughts, activities, and feelings.

We believe that the schedules and messages

posted on SNSs are useful information sources to recall memories.

In this paper, we propose a memory recall support system that extracts useful keywords from the texts written by users on Google Calendar and Twitter. In addition, since we assume that we can identify memorable events on the days when a user takes many photos, we use the number of photos to weight the extracted keywords. We present weighted keywords using tag clouds to jog user recollections.

Below, in Section 2 we explain our approach's overview. Algorithms and preliminary experiments are described in Sections 3 and 4. Related work is shown in Section 5.

2 APPROACH

Our research supports human recollection by extracting keywords from Google Calendar and Twitter, weighting them using term frequency and the number of photos, and displaying them by tag clouds.

First, we obtain the data written by users on Google Calendar 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 (Murakami, 2010). Next, we generate tag clouds from history structures.

As a feature of this study, we support the recall of impressive memories often obtained on trips or at

special events because people tend to take photos on such occasions. To put it another way, impressive events can be found around days on which a user takes many photos. The above hypothesis is an important element in our study.

A tag cloud is a visual representation of text data, typically used to depict keyword metadata (tags) on websites or to visualize free form text. Tags are usually single words, and the importance of each tag is shown with font size or color. This format is useful for quickly perceiving the most prominent terms (Wikipedia, 2013). We exploit this advantage and adopt tag clouds to display keywords.

An overview of our approach is shown in Figure 1.

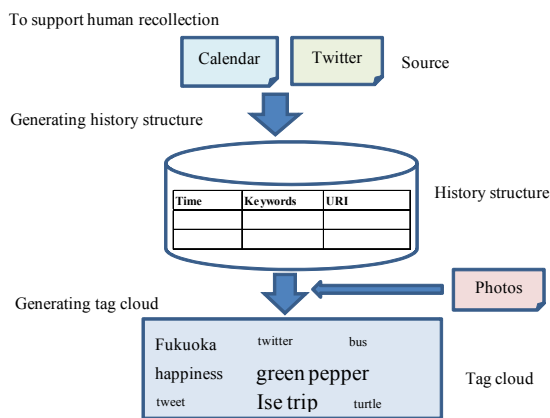


Figure 1: Overview of our approach.

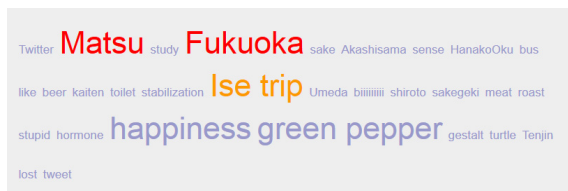


Figure 2: Tag cloud generated by our system.



Figure 3: Tag cloud generated by comparative system.

Figure 2 shows an example of a generated tag cloud created with our system with data from one month of the first author’s life (September 2012). Figure 3 shows an example of a created tag cloud with a comparative system that only calculates the

weights of tags by term frequency.

In this period, the author was diligently preparing for a national exam of the application information technology and also traveled to *Ise* with a friend named *Matsu* for a few days and had lots of fun.

In the comparative system (Figure 3), *study* and *application* appeared larger because he tweeted these terms many times.

In our proposed system (Figure 2), *Ise trip* and *Matsu* are displayed larger and *Matsu* is strengthened in red. The word *happiness* reflects his tweets during his trip, and *green pepper* reminds him that he ate incredibly hot green peppers for the first time at a unique Japanese style barbecue restaurant that brought tears to his eyes. It made a deep impression on him. He took many photos during the trip, and green pepper is displayed even though it only occurred on one day.

We believe that our proposed system effectively helps users recall impressive memories.

We show a complete image of our system in Figure 4.

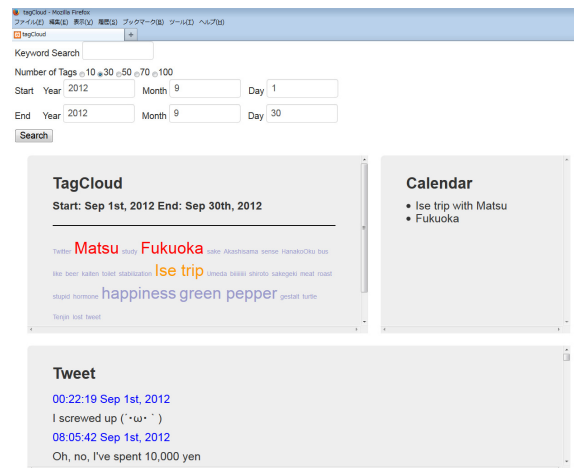


Figure 4: Complete image of system.

Our system, which works on a browser, has two search functions: *keyword* and *range*. Keyword search enables users to input keywords for output related words on tag clouds. Range search enables users to input periods (start and end dates) to display a tag cloud for that period. The functions can also be combined. The number of tags can be selected from five choices: 10, 30, 50, 70 or 100.

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 the event titles of a calendar, and the twitter part displays tweets. Furthermore, each tag can be clicked on, and the tag becomes a new search word.

The user can access detailed memories by clicking on tags.

3 ALGORITHM

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

3.1 Generating History Structure

3.1.1 Data Extraction from System Usage History

In this study, we use information from Google Calendar and Twitter to support human recollection.

(1) Google Calendar

Here, we obtain an event time (start time) and an event title. Keywords are extracted from the event title by the algorithm in the next section.

(2) Twitter

Since tweets generally express user thoughts or activities, we use all of them except those starting with @ because they are mainly discourse and official RTs (Retweets) because they are mainly the opinions of others. We extract a tweet's time, the tweet itself, and generate keywords from it using the algorithm described in the next section.

3.1.2 Generating Keyword Algorithm

We developed a generating keyword algorithm that creates a set of keywords from such texts as the event titles in calendars and tweets. Figure 5 shows its general outline.

First, it extracts noun phrases as keywords from the collected text data with MeCab, a Japanese morphological analysis tool. We removed 11 unnecessary words that often appear in tweets, such as *today*, *now*, and *tomorrow*.

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 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 become a keyword; an example of the latter is *jinko* (artificial) and *chino* (intelligence), which are concatenated into *jinkochino* (artificial intelligence) and become a keyword. 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. The detailed algorithm is described in (Mitsuhashi, 2011).

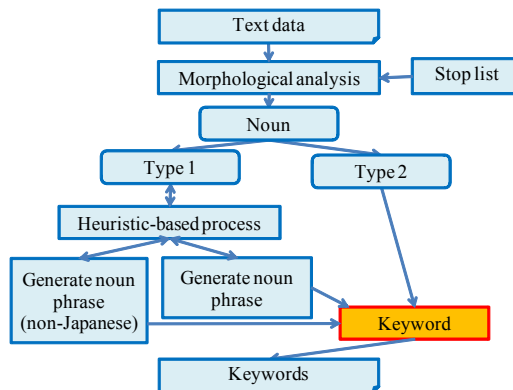


Figure 5: Generating keyword algorithms.

3.2 Generating Tag Clouds

3.2.1 Calculating Tag Weights

Calculating the weight of a word is different for the information source. In addition, our system considers the number of photos taken by users to calculate the weight, since we assume that on special days users take many photos.

Next we define how to weight the words that appeared in the calendars and those that appeared in the tweets and sum up the weights.

(1) Weighting calendar words

We define weighting function $CalW(t)$ to weight word t that appeared in the calendar:

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

$G_{date}(t)$ is a function that obtains the date of word t . $C_{photo}(R)$ is a function that obtains the number of photos of range R . HS means the history structure. The more photos taken by users, the more the weights of the words increase.

Figure 6 shows an example of calculating calendar words.

The data in Figure 6 are the author's usage on September 2012. We calculated the weight with this example data.

Fukuoka (place name) appears three times in this period. Each *Fukuoka* is weighted with the number of 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 these points are added.

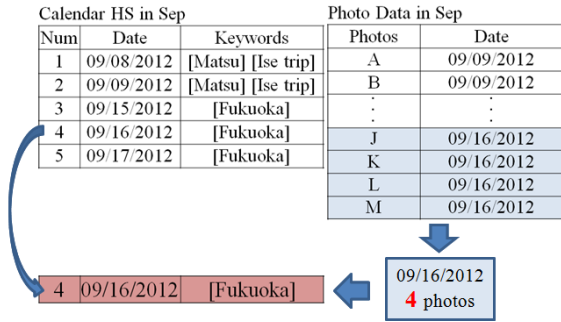


Figure 6: Example of calculating weight of calendar words.

(2) Weighting Twitter words

We define weighting function $TwiW(t)$ to weight word t that appear in tweets:

$$TwiA(t) = \frac{1 + C_{photo}(G_{date}(t))}{C_{tweet}(G_{date}(t))}, \quad (2)$$

$$TwiB(t) = C_{photo}(G_{time-ba}(t)), \quad (3)$$

$$TwiW(t) = \sum_{t \in HS} (TwiA(t) + TwiB(t)). \quad (4)$$

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

We assume that the more tweets in a day, the more both noisy and good words increase. We need to normalize the weights based on the number of tweets. The denomination of $TwiA(t)$ shows the normalization.

Additionally, Twitter disseminates information in real time. Since we believe that both a photo and a tweet made in a close time are highly related, the number of photos taken one hour before and after the tweet time is added to the weight (Eq.3).

Finally, the addition of $TwiA(t)$ and $TwiB(t)$ becomes $TwiW(t)$.

Figure 7 shows an example of the calculation of twitter words. The data of Figure 7 are again taken from the author's data on September 9. There are 71 tweets and five photos.

Next we calculated the weight of *green pepper* that was tweeted at 13:17 (hereinafter “gp1317”). First, we calculated $A(gp1317)$. Green pepper appears six times, and there are 71 tweets and five photos on September 9. The $A(gp1317)$ result becomes $(1+5) / 71 = 0.08$. Second, we calculated $B(gp1317)$. There are three photos one hour before

and after its tweet time. Thus, $B(gp1317)$ becomes 3. From the above results, the weight of $gp1317$ becomes $0.08 + 3 = 3.08$.

Other “*green peppers*” are calculated in the same way and these points are added.

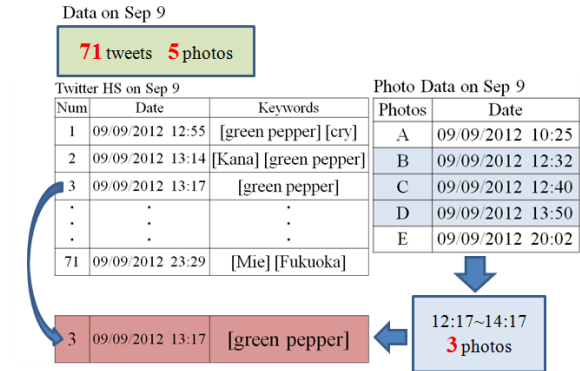


Figure 7: Example of calculating weight of twitter words.

Finally, we define weighting function $Weight(t)$ of word t :

$$Weight(t) = CalW(t) + TwiW(t). \quad (5)$$

This is its final weight.

3.2.2 Font Color

We designed font colors for tags based on the information sources in which they appear.

(1) Calendar only

If word t only appears in the Calendar, it is orange (#FF9900).

(2) Twitter only

If word t appears in Twitter, it is blue (#9999cc).

(3) Both Calendar and Twitter

If word t appears both in Calendar and Twitter, it is red (#FF0000).

The above color selection is based on our previous work (Mitsuhashi, 2011).

3.2.3 Sorting Tags

Tags are sorted by the time of the word. When word t appears more than once, the oldest time at which t appeared is used for sorting.

4 EXPERIMENT

4.1 Overview

Our subjects were five males, aged 23-25. We used one month-long data from August 2013. All subjects

used Twitter and took photos. Since three did not use Google Calendar, they copied their schedules onto it.

4.2 Experiment 1

We evaluated the usefulness of the algorithms that weight the extracted keywords.

4.2.1 Method

We extracted the top 30 keywords from two systems: our method and a comparative method. Our method is described in Section 3.2. Weighting function $Com(t)$ for word t in the comparative method is defined as follows:

$$Com(t) = \sum_{t \in HS} tf(t). \quad (6)$$

$Com(t)$ is calculated by the frequency of the word's occurrence. That is, the more it appears, the more its weight increase.

From the two systems, we merged the extracted 60 keywords and sorted their appearances alphabetically to form a list so that the subjects cannot guess the weighting algorithms.

Our subjects evaluated whether the extracted keywords on the list helped them recall their memories at five levels (5: very useful; 4: useful; 3: neutral; 2: not very useful; 1: not useful).

4.2.2 Results and Analysis

The average values of each algorithm are shown in Table 1, and the data obtained by each subject are shown in Table 2.

Table 1: Results of experiment 1.

	P5	C5	P10	C10	P30	C30
Subject 1	4.0	3.3	3.2	3.8	3.4	3.3
Subject 2	4.2	3.2	3.1	2.7	3.0	2.5
Subject 3	4.2	4.2	4.2	2.7	2.4	2.3
Subject 4	2.6	3.8	2.6	3.2	2.6	2.6
Subject 5	4.0	4.0	4.2	2.9	3.9	2.4
Mean	3.8	3.7	3.5	3.1	3.1	2.6

Notes: P: Proposed method; C: Comparative method; 5: Top 5; 10: Top 10; 30: Top 30

Overall, our method outperformed the comparative method.

However, we found little difference between the proposed and comparison methods with respect to Subjects 1, 3, and 4, because their number of keywords is smaller than those of other subjects. The same keywords were extracted from the two different approaches. For example, for Subject 3, 27

words are identical in the two approaches. On the other hand, for Subject 2, only ten words are the same. If Subjects 1, 3, and 4 used calendars and/or Twitter more often, the results might improve.

These results suggest the usefulness of our algorithm for weighting keywords.

Table 2: Data set.

	All	Cal	Twi	Photo
Subject 1	69	13	57	36
Subject 2	467	8	462	73
Subject 3	33	7	28	19
Subject 4	65	25	50	41
Subject 5	748	15	735	32
Mean	276.4	13.6	266.4	40.2

Notes: All: extracted keywords from calendar and Twitter; Cal: extracted keywords from calendar; Twi: extracted keywords from Twitter

4.3 Experiment 2

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

4.3.1 Method

We prepared six tag clouds that display 30 keywords for comparison: (a) three information sources (Calendar and Twitter, Calendar only, Twitter only) \times (b) two weighting methods (our method and comparative method).

The following are the six tag clouds:

- Tag cloud A is composed of Google Calendar and Twitter and uses our weighting method.
- Tag cloud B is only composed of Google Calendar and uses our weighting method.
- Tag cloud C is only composed of Twitter and uses our weighting method.
- Tag cloud D is composed of Google Calendar and Twitter and uses the comparative weighting method.
- Tag cloud E is only composed of Google Calendar only and uses the comparative weighting method.
- Tag cloud F is only composed of Twitter and uses the comparative weighting method.

The following are the questions:

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 is the most useful to recall your memories among the six tag clouds?

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

Every tag cloud displays up to 30 tags.

4.3.2 Results and Analysis

The results of experiment 2 are shown in Table 3.

Table 3: Results of experiment 2.

	Q1	Q2	Q3	Q4	Q5	Q6
Subject 1	A	B	C	A	D	A
Subject 2	A	B	C	A	D	A
Subject 3	A	B	C	A	E	A
Subject 4	A	B	C	A	D	A
Subject 5	A	B	C	B	E	B

Our method is better than the comparative method regardless of the information source, based on the results of Qs 1-3 (5/5). From the result of Q6, one subject said that “Tag cloud A greatly displayed important words.” Other opinions included, “Changing the color for each source was effective.” On the other hand, Subject 5 selected tag cloud B and complained that “Tag cloud A has much more noise than tag cloud B.” There were probably many useless tweets on that day when many photos were taken. These results suggest the overall usefulness of our method.

5 RELATED WORK

This research is a part of a system called Memory-Organizer that helps users construct “externalized-memory” (Murakami, 2002). Murakami et al. (Murakami, 2012) created Knowledge-space browser from five information usages to support human recollection. The differences between previous research and this paper are that we selected two important information sources for memory support: new algorithms for weighting words using the number of photos, and displaying tag clouds.

Much research integrated such information in the light of personal information management (PIM)(William, 2007).

There are many researches on tag clouds. Kuo et al. presented PugCloud, whose tools use tag clouds to summarize the results from queries over the PubMed database of biomedical literature (Kuo,

2007). Eda et al. created a novel tag cloud (Eda, 2009) that uses tag entropy values to determine font size. Many of these studies summarize the results of searches or web pages. Our research uses tag clouds for the recall of memory support. To the best of our knowledge, no research uses tag clouds for human memory recollection.

6 SUMMARY

We presented a system that supports human recollection with tag clouds created from the use of Google Calendar and Twitter. This research’s main feature is to weight words using the number of photos taken by users to recall memorable events. Preliminary experiments suggest the usefulness of our approach. Since this is merely its first step, we need to improve our algorithms and conduct further experiments with more subjects. We believe our approach is useful to recall the memories of impressive events and should be investigated in the future.

REFERENCES

- Murakami, H., 2010. 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.
- Wikipedia, 2013. http://en.wikipedia.org/wiki/Tag_cloud. (accessed 02-January-2014).
- Mitsuhashi, K., 2011. A Human Recollection Support System by Integrating Diverse Information and Creating Knowledge Space. *Master’s Thesis, Graduate School for Creative Cities, Osaka City University*. (in Japanese).
- Murakami, H., Hirata, T., 2002. Information Acquisition and Reorganization from the WWW by using Memory-Organizer”, *Bulletin of Osaka City University Media Center*, vol. 3, pp. 9-14.
- Murakami, H., Mitsuhashi, K., 2012. A System for Creating User’s Knowledge Space from Various Information Usages to Support Human Recollection. *International Journal of Advancements in Computing Technology*, vol. 4, no. 22, pp. 496-508.
- William, J., 2007. Personal Information Management, *ARIST*, vol. 41, pp. 453-504.
- Eda, T., Uchiyama, T., Uchiyama, T., Yoshikawa, M., 2009. Signaling Emotion in Tagclouds. In *WWW 2009*, pp. 1199-1200, ACM Press.
- Kuo, B. Y-L., Hentrich, T., Good, B. M., Wilkinson, M. D., 2007. Tag Clouds for Summarizing Web Search Results. In *WWW 2007*, pp. 1203-1204, ACM Press.