平成 27 年度

筑波大学情報学群情報科学類

卒業研究論文

題目　Understanding How Cues in Lifelog Affect Memory Recalls

主専攻　知能情報メディア

著者名　　　　松岡里咲

指導教員　　田中二郎・高橋伸・志築文太郎

Abstract

Lifelogging technologies can support human memory. Recently, due to their rapid development, these technologies became affordable to the public. For further spread of lifelogging devices, our goal is to design and develop an interface that enables users to find past moments in the most efficient way, while aiding memory recall at the same time. To achieve this goal, we conducted several surveys to psychologically understand how lifelog mediate human memory and to find the combination of cues that are most valuable in evoking memory recall. Based on these results, we developed a web application that displays lifelog images, enables keyword search, and suggests related keywords that are efficient in eliciting further recall.

Contents

1. Introduction 1

1.1 Fruition of Lifelog 1

1.2 Benefit of Lifelog 1

1.3 Capturing Method 2

1.4 Overview 2

2. Related Work 4

2.1 Recalling Memories 4

2.2 Memory Cues 4

2.3 Extracting Efficient Moments 5

3. Methodology 7

3.1 Preliminary Survey 7

3.1.1 Method 7

3.1.2 Results 8

3.2 Additional Survey 14

3.2.1 Method 14

3.2.2 Results 14

3.3 System 17

3.3.1 Design Decisions 17

3.3.2 Environment 18

3.3.3 Architecture. 18

4. Discussion and Future Work 24

Conclusion 25

Acknowledgements 26

References 27

List of Figures

1 Comparison of number of events recalled by independent recall versus lifelog-dependent recall 9
2 Cues used in independent recalls and lifelog-dependent recalls 10
3 Comparison of number of events recalled by independent-then-lifelog-dependent recall versus lifelog-dependent-only recall 11
4 Cues used in independent-then-lifelog-dependent recalls and lifelog-dependent-only recalls 13
5 Combination of cues the participants remembered for each event 15
6 Overview of our system 18
7 Visual image on what the users see under the View tab 20
8 Visual image on what the users see when editing tags 21
9 Visual image on what the users see when searching for tags 22
10 Visual image on what the users see after clicking the Delete Tags button under the View tab 23

Chapter 1.
Introduction

1.1 Fruition of Lifelog

Human memory is volatile, therefore technologies for storing every personal experiences have been anticipated for quite a while. In 1945, Bush proposed Memex as “a device in which an individual stores all his books, records, and communications, and which is mechanized so that it may be consulted with exceeding speed and flexibility” [4]. He did not foresee the exact technology to accomplish this at that time, but recently the development of “wearable computers” and massive increases in digital storage made Bush’s vision a reality. Mann created a prototype of a “wearable computer” in the early 1980s [18], which was “equipped with a head-mounted display, cameras, and wireless communications”. In the early 2000s, wearable light-weight devices with multiple sensors designed to capture the user’s experience, such as SenseCam [12], Vicon Revue [27], and Narrative Clip [22] became affordable to the public. As of storage, impressive increases in digital storage enabled systems such as MyLifeBits[11], Haystack[14], and Stuff I’ve Seen[8] to store and retrieve metadata.

1.2 Benefit of Lifelog

Since we became capable of recording and storing our everyday life, terms like “lifelog” and “lifelogging” came into use. Lifelog represents the act of recording one’s personal data continuously (such as photos and videos from one’s viewpoint, audio, location, activity log on computers, biometric signals and others), or the record itself. There are many potential benefits in lifelogging, but by far the most commonly invoked one is supporting memory recall and reminiscing about past personal experiences [24]. Bell et al. emphasize that “digital memories allow one to vividly relive an event with sounds and images, enhancing personal reflection in much the same way that the Internet has aided scientific investigations” [2].

1.3 Capturing Method

Two types of lifelog exist: total capture and situation-specific capture [24]. Total capture is a complete record of everyday life, whereas situation-specific focuses on events such as meetings or certain habits, aiming to capture rich data on a complex event. Sellen et al. emphasized the importance of capturing *everything*, not intentional selection of moments to capture. Sellen et al. also stated that passively captured images “cause people to remember more events than they would with their own actively-captured events” [23].

There are many ways to capture experiences, such as by video, audio, geographic data, heart rate, and even brain wave activity [1, 13, 9, 29, 16]. However, when it comes to recalling memories, many researchers observe that photos are the most promising and convenient method. Prior work stated that records with visual aid support a more effective recall of past events than paper diaries [9]. Another prior work stated that photos in general were better than videos for recalling details, because photos leave the users some room for inferring the context of the situation [5, 6]. Therefore, in this paper our capturing method will focus on total and automatic capture, via photos.

1.4 Overview

Since it has become possible to record and store our everyday events, researchers are focusing next on how to make use of the captured data. Neither recording technology nor data capability defines the productivity of lifelog; an interface that enables users to find what is most valuable to them is the key to a wider spread of lifelog. We strongly believe that a lifelogging system designed to provide users with cues that are effective in quick memory retrieval will be needed in the near future.

Therefore our goal is to first understand what cues are valuable in recalling memories, then discuss what designs would be appropriate in providing users with those effective cues. We believe our research will contribute to providing realistic explanations on how a lifelogging system support memory recall, and making them more familiar and worthy to the users.

More specifically, our approaches are:

(a) conduct a preliminary survey to understand the difference between independent memory recall and lifelog-dependent memory recall

(b) find memory cues that are valuable in memory recall when viewing captured lifelog

(c) design and develop a lifelog viewing system that enables users to efficiently recall memory, by providing the cues found in the preliminary survey

The rest of the paper is structured as follows: in chapter 2 we introduce related work and explain how our approach is different from them. In chapter 3, we describe the surveys we conducted and discuss the analyzed results. Then we describe the design decisions and architectural details of our system. We conclude with a discussion on our findings, limitations of our work, and a brief overview of the future work.

Chapter 2.
Related Work

2.1 Recalling Memories

Human memory is known to be a reconstructive process mediated by cues [5, 28]. Reconstruction is about "deducing that one must have participated in some event even if one can't actually recollect it" [16]. For example, a participant might think "I do not recall being told to send an e-mail to this client, but my boss must have given me the instruction to do so since I have a memo written on my notebook".

When it comes to reviewing lifelog, memory is elicited by either "true recall" or reconstruction. "True recall" is the act of recollecting or mentally reliving an event from the past, which Tulving names it as "episodic memory" [25]. Lifelogging systems seem promising because "lifelogging tools can both promote such reconstructive inferences, as well as support genuine recall" [23]. Since it is known that memory recall is shaped around memory cues as trigger, researchers of lifelogging systems focus on narrowing down to memory cues that are more effective.

2.2 Memory Cues

Since memories are reconstructed based on memory cues, recent approaches in lifelogging systems are designed and built upon statements saying that providing *where*, *what*, and *who* as memory cues cause effective memory recalls. Brewer stated that place, events, and people are stronger cues than time in recalling memories [3]. Thus, rather than asking "what were you doing on 2015.9.26?", asking "when was the last time you had lunch with your friend Naomi at the cafeteria?" would be better. Wagenaar discovered that the most successful order was when *what* was given first, followed by *when* as cues [28]. He also stated that *when* by itself was useless, and that the combination of cues *who* and *what* showed significant recalling accuracy as well. These observations with regard to memory cues were made in 1980s, long time before lifelogging systems became popular. Nevertheless, they constitute a strong basis for recent research.

2.3 Extracting Efficient Moments

Considering that lifelog data can be enormous (images taken every 30 seconds for 18 hours sum up to more than 2000 photos to review a single one-day event), researchers are now aiming to extract moments which are efficient in memory retrieval. In doing so, they focus on the cues mentioned above.

Memon focused on a physical object (person or a thing) as key cue, and found that presenting archives that include specific objects is effective for memory recall [20]. In his research, he developed a prototype using Android smartphone that is capable of "autonomously identify(ing) the people and objects interacted by the user together with the usual logging activity". With his prototype, the "required lifelogs can be easily retrieved based on the people and objects currently present near the user. Likewise, the lifelogs related to the present or user defined location may also be directly accessed if the lifelog device possesses the records of any previous visits to that place". For example, whenever a user sees a person whom he had a meeting with before, he would use the prototype to recognize this person, then the prototype would automatically search and present the user with archive photos in which that person appears in. Memon states that, by doing so, the user can effectively remember what the situation was (where, when, the topic they were talking about) when he last saw that person.

Kalnikaite et al. developed lifelogging systems that provide the user with (a) visual record only, (b) geographical record only, and (c) both visual and geographical record. Then, they compared their effect on user's memory recall [16]. The results showed that each set of information resulted in a different type of memory recall; visual images evoked simple recall of past events, while geographical records allowed users to infer what must have happened at that time or how they must have moved from one place to another, according on where the user was before and after a certain length of time. Therefore, they concluded that different types of memory cues prompt different ways of remembering. Participants in this research preferred visual aid and locational information to be provided together, since recall and inference put together resulted in a more efficient memory recall.

Lee et al. developed a system which automatically selects the most efficient cue within the captured lifelog, with its aim to support people with memory impairment and their caregivers [17]. The system first "captures photos, ambient audio, and location information and leverages both automated content/context analysis and the expertise of family caregivers to facilitate the extraction and annotation of a salient summary consisting of good cues from the lifelog". "Good cues" are calculated by: identifying photos containing faces, identifying the movement of the user, recognizing an object within a photo, or breaking up experiences into distinct actions. Their system heuristically finds the most appropriate cue then presents the photos with red rectangles emphasizing the so-called "good cue" within the archive. The results showed that viewing the pictures automatically selected and presented by the proposed system was progressive in recalling more over time, and patients felt more confident about their past experiences.

Many researchers seek useful ways to utilize archived data by designing effective retrieval cues. However, they lack in psychological understanding of human memory and theoretical evidence on how their designs support memory recalls. The proposals introduced above are substantiated by retrofitted experimental data, therefore they lack in theoretical explanation and evidence on how their systems support human memory. Without revealing how people mediate memory from lifelogging systems, we can only take a best guess at designing a more effective system. Therefore, our goal is to conduct several surveys in order to psychologically understand how lifelog mediates human memory, then to design and develop a lifelogging system based on the surveys' results that supports the user's memory recall as much as possible.

Chapter 3.
Methodology

In order to psychologically understand how lifelog mediates human memory, and to determine the cues that are valuable in recalling memories, we conducted several surveys as described below. Then, based on our results, we developed a lifelog viewing system designed to effectively support memory recall.

3.1 Preliminary Survey

In our preliminary survey, we compared independent recall against lifelog-dependent recall in order to understand the difference between them. We also observed closely on how the participants recalled their memory, so that we can narrow down our focus on cues that are expected to be valuable in recalling memories.

3.1.1 Method

We gathered 8 participants (6 males and 2 females), with ages ranging from 21 to 52 years (average age 29.25 years), from inside and outside of the University of Tsukuba. Each of them put on a wearable camera called Narrative Clip on their upper chest facing the front. Narrative Clip is a clip-on device which automatically captures photos every 30 seconds and can be bought online. Narrative Clip comes with an official viewer system, so that the user can view the photos captured by the device. The participants wore the device for 4 continuous hours somewhere between 9 am and 7 pm. During that time, participants were free to do anything and go anywhere, allowing the device to capture a part of their daily life.

A week later, we conducted the recall stage where we asked the participants to recall what they were doing when they were wearing the Narrative Clip and compared independent recall against lifelog-dependent recall. Here, we define independent recall as recalling past experiences without seeing or referencing any information. Participants independently tried to recollect what has happened during their capture stage, until they reached the point where they could not remember any more. As opposed to independent recall, lifelog-dependent recall is where participants are able to review photos that Narrative Clip automatically took via its official viewer system. The photos taken within the 4 hours added up to approximately 500 photos, providing participants with visual aid of what was in front of them during the capture. By comparing these two methods, we aimed to understand how much they could remember and how they would elicit their memory for each method.

During the recall stage, we divided the participants into two groups, each consisting of 4 people. The first group first experienced independent recall, then, without any pause, experienced the lifelog-dependent recall. The second group experienced only the lifelog-dependent recall. After recalling, both groups answered the questions "What were the events that you recalled?" and "How did you remember those events (what cues did you trace)?".

3.1.2 Results

The average number of events recalled was 15.0 for independent recall by the first group, and 28.5 for lifelog-dependent recall by the second group (see Figure 1). This shows that even though the participants did not have much trouble remembering events that happened a week ago, lifelog-dependent recall supported more memory recall, by approximately twice the amount. We performed the f-test and the t-test in order to statistically evaluate the results. An f statistic result from the data was $p = 0.242529867292555$ ($p > 0.05$) meaning the data was normally distributed. Then we performed the t-test with no-correlation but normally distributed situation, and obtained $p = 0.089504762775408$ ($p < 0.1$). Therefore the results statistically show that there tends to be a significant difference on the number of events remembered between independent and lifelog-dependent recall.

This fact supports the research performed by Microsoft with their wearable camera, called SenseCam [21], and the research by Kalnikaite et al. [16]. Microsoft emphasizes on the effect lifelog photos have on alleviating memory loss by having their results open to the public. Kalnikaite et al. concluded that presenting users with visual aid caused better recall. Moreover, our preliminary survey provided concrete evidence of this statement as well.

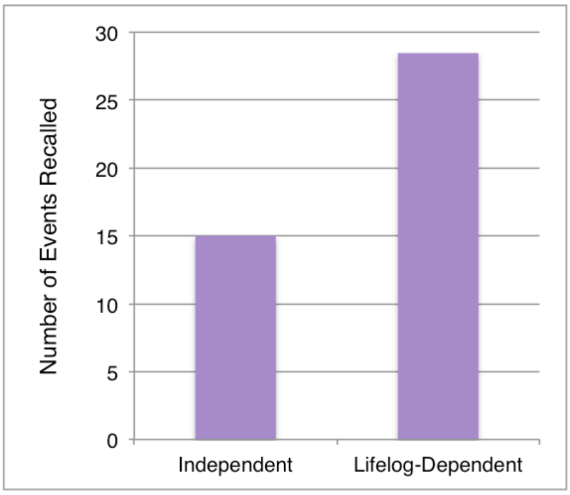


Figure 1. Comparison of number of events recalled by independent recall versus lifelog-dependent recall

The remembered events for independent recall consisted of: events unique to that date (events that do not happen often), everyday events or routine (such as having lunch or going back home), and events that users spent most of their time doing. In addition to those events, by viewing the lifelog the users were able to recall: more specific events, their feelings towards the event, background of how the event happened, unexpected trouble, small events that occurred, and exact names and positions of the persons that were present at the place. With these details added up, the number of events recalled doubled when viewing lifelog. Sellen et al.'s remark [23] that passively-captured photos cause more recall on past events can be backed up by our results as well. By capturing moments fairly and frequently, users were able to vividly remember the events that happened that day or feelings toward that day, because they could recall the details that slipped out of their mind. We believe people do not spend their days driven to select what moments they want to capture to recall later on. People are living by the moment, therefore it is only natural that lifelog should be passively recorded.

With regard to the answer to the question "How did you remember those events (what cues did you trace)?", the cues were very different according to their recall method (see Figure 2). For independent recall, we can see that the participants thought of the date and time provided, thought of the schedule a few days before and after the experiment date, calculated backwards what they were doing on the experiment date, and chronologically recalled what happened on that date using: unusual events, *who* they met, and *where* they have been as cues. However when it comes to lifelog-dependent recall, we can see that users directly recalled what they experienced from cues such as *who* is in the picture, *where* they were, *what* they had in their hands, and what they see on their computer or smartphone *screens*. These findings support the prior statements that memories are reconstructed by cues [5, 28]. Furthermore, our results indicate that independent recall evokes chronological recall on memories, and that lifelog-dependent recall evokes direct cue-oriented recall. To the best of our knowledge, no other researchers have emphasized the difference between recall processes and found the chronological versus cue-oriented elicitation. Mathur et al. designed a lifelog search system and made a comparison between date-ordered-chronological search and person-or-thing-oriented search. Their results implied that "linking of events is more beneficial when people are browsing their lifelogs with a specific goal in mind" [19]. Their research only applies in faster search of archived data, but our results broaden out to recalling methods. The characteristic in which people tend to directly recollect memory by provided cues within lifelog photos, highlights the importance of revealing what cues are valuable.

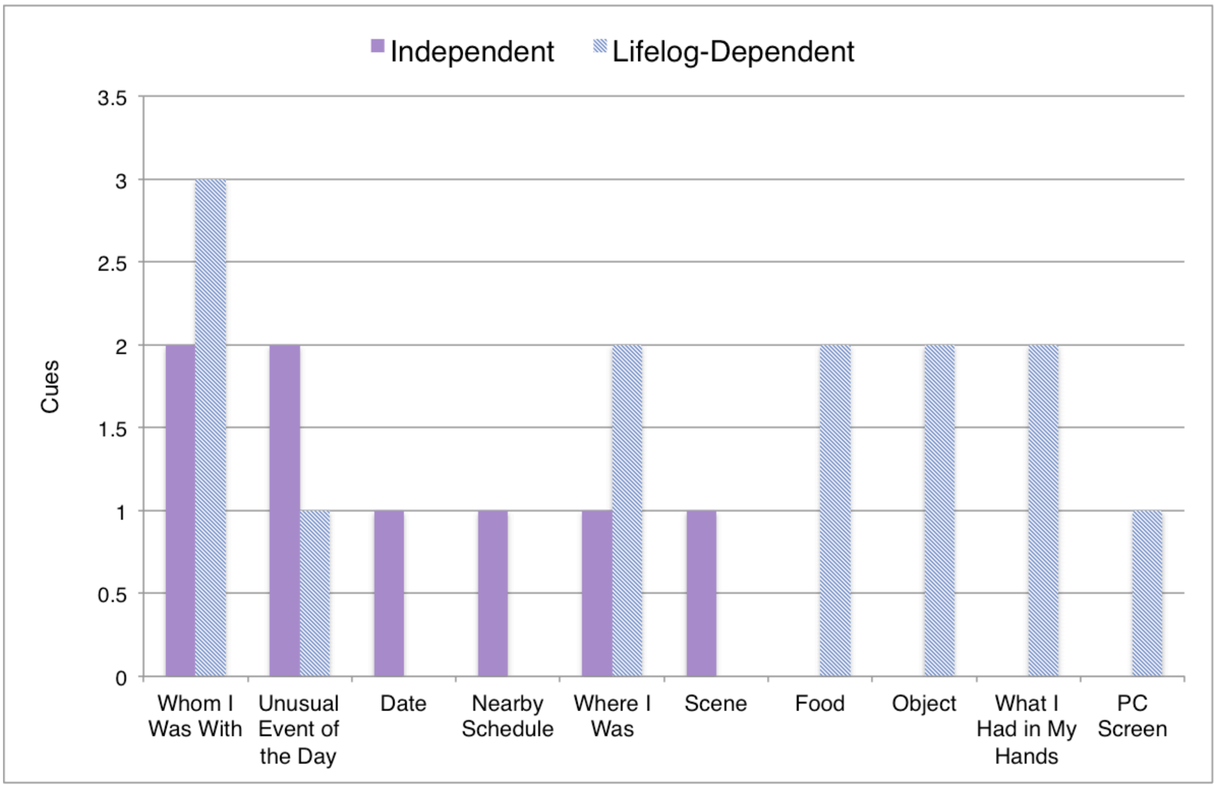


Figure 2. Cues used in independent recalls and lifelog-dependent recalls

In addition, seeing objects evoked caused users to recall emotional feelings or background information on that object. For example, when a participant was given an explanation by another person about a deck of cards (answering questions written on those cards will tell about oneself), seeing those cards in a picture caused the participant to remember what he talked about during that session and what results he got. In a different situation, an expensive machine was in the picture and a participant recalled his professor buying that machine a few weeks ago, since it was new and extremely precious to his laboratory. As we can see from these results, objects in a picture represent not merely an object, but an important cue in supporting emotional feeling and background information related to that object.

The same may apply to people's faces. Three participants answered that looking at faces in photos allowed them to recall the content of the conversations they had. However, one participant spent about an hour with a friend, but could not recall the topic of conversations by viewing her photos. So far, we can observe that presenting users with faces of people they have spent time with supports memory recall of what they did together, but further investigation is needed to confidently say that faces also enable recall on the content of the conversations they had.

Up to this point we have compared independent recall done by the first group versus lifelog-dependent recall done by the second group. Next, we compared how (a) viewing lifelog after independent recall differs from (b) viewing lifelog without independent recall. The method (a) was conducted by the first group, and the method (b) was operated by the second group. The goal of this task is to examine whether the method of viewing lifelog photos by itself has a promising impact on memory recalls.

The average number of events recalled by independent-then-lifelog-dependent method was 25.5, and 28.5 for lifelog-dependent-only recall (see Figure 3). The numerical values were not very different, thus we performed the f-test and the t-test in order to statistically evaluate the results. An f statistic result from the data was $p = 0.314618801$ ($p > 0.05$) meaning the data was normally distributed. Then we performed the t-test with no-correlation but normally distributed situation and obtained $p = 0.723942572311238$ ($p > 0.05$). Therefore, the results statistically show no significant difference on the number of events remembered between independent-then-lifelog-dependent recall and lifelog-dependent-only recall.

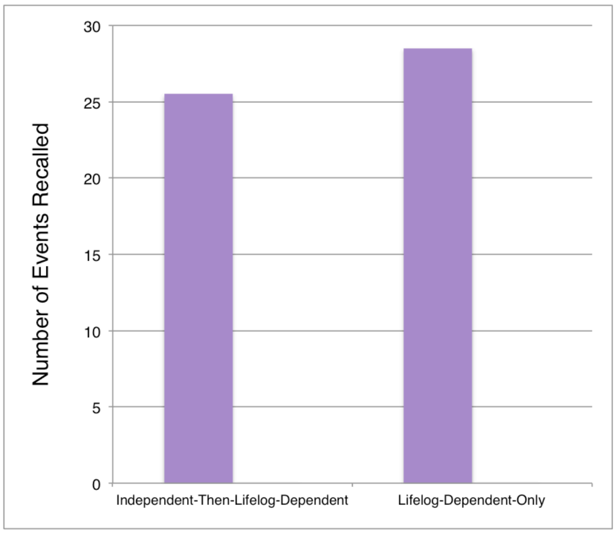


Figure 3. Comparison of number of events recalled by independent-then-lifelog-dependent recall versus lifelog-dependent-only recall

The content of remembered events for each recalling method did not show much of a difference either. For independent-then-lifelog-dependent recall, after viewing photos participants were able to recall: one whole event they could not recollect on their independent recall, the accurate sequence of events, the fact that a user chose a different route than usual, and generally more detailed information of events. We can see that lifelog-dependent recall filled the gaps that independent recall could not fill, but the characteristics of recalled matters were similar: more detailed information, specific circumstances of an event, and background information of an object within a picture.

With regard to the answer to the question "How did you remember those events (what cues did you trace)?", the cues were very similar between independent-then-lifelog-dependent recall and lifelog-dependent-only recall (see Figure 4). Even though the number of answered cues was not the same, the trend of cues was similar. They can be narrowed down to *who* they were with, *what* was in their hands or in the picture, *where* they were, and whether they were looking at a computer or smartphone *screen*. This is similar to Brewer and Wagenaar's statements on cues [28, 3]. Our result implies that an object (what they had in their hands, including food, and what physical object was in the picture) facilitates memory recall the most, followed by the information on *who* is in the picture. These results are practically equivalent to Wagenaar's conclusion, and we can agree that *when* as a cue by itself does not seem to have a place in recalling memories. Cues on *where* the picture was taken seem to trigger memory as well, as Brewer claimed. However, the most significant outcome of our survey is that computer or smartphone screens captured in a picture seem to tell many things to the user. Computers and smartphones are now very popular and commonly used in our everyday life, and seeing what was on the screen through the photos helps in recalling what the users were doing with it, such as texting to a friend, searching for a keyword on the internet, reading the news, or working on their school tasks. Personalized computer devices were not as popular in 1980s when Brewer and Wagenaar conducted their research. The change in our life styles throughout the years seems to have made information on *screen* become an important cue.

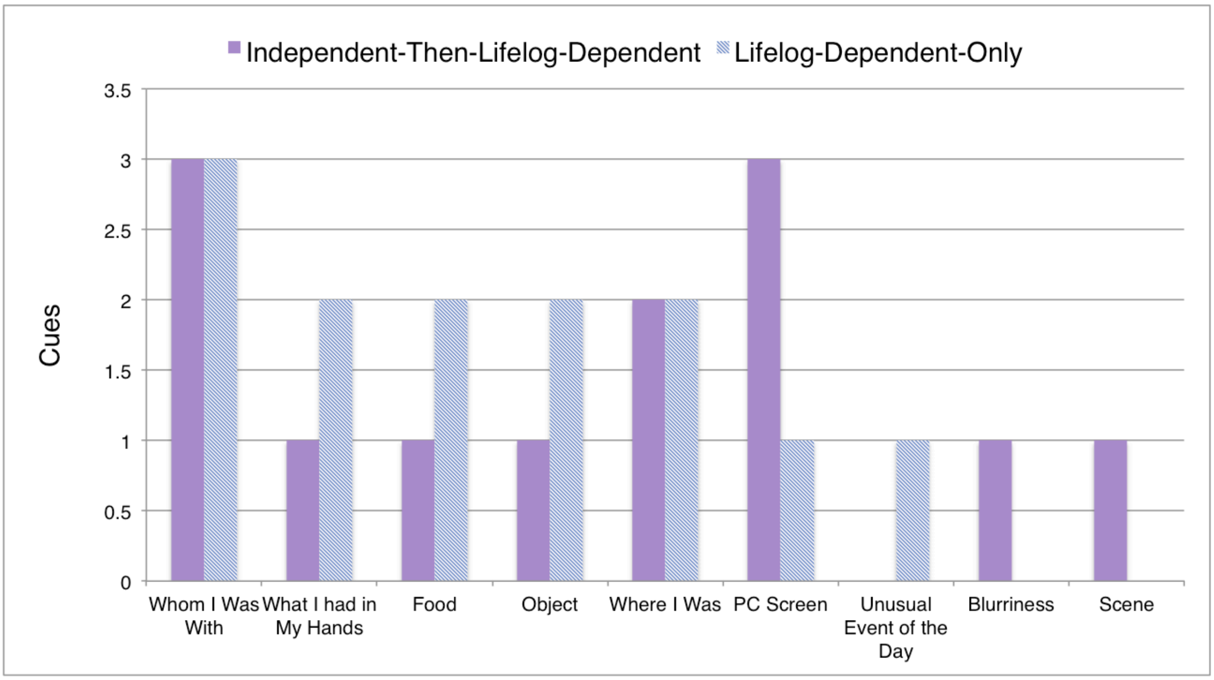


Figure 4. Cues used in independent-then-lifelog-dependent recalls and lifelog-dependent-only recalls

In summary, the results from our preliminary survey showed that

(a) presenting the users with lifelog photos supported better memory recall than independent recall, by twice the number of events

(b) starting right away with viewing lifelog photos supported a sufficient amount and content of memories to be recalled

(c) by reviewing the captured photos, users were able to remember: more detailed information on events and situations, accurate sequence of events, and background information on objects within the picture

More importantly, the highlight of the results are as follows:

(d) independent recall caused users to chronologically recall the past, whereas lifelog-dependent recall caused direct and cue-oriented recall

(e) cues on *what* was in the picture or what the users were holding in their hands, *who* they were with, *where* they were, and, last but not least, what was on the user's *screen* seem to have the potential of being the most valuable cues in recalling memories

3.2 Additional Survey

Next, we conducted an additional survey to specifically determine which cue or what combination of cues are most valuable in recalling memories.

3.2.1 Method

We gathered 4 participants (all different from the 8 participants from the preliminary survey), all male, aged 22 or 23 years (average age 22.5), within the University of Tsukuba. Each participant wore a Narrative Clip for 4 continuous hours, somewhere between 9am and 7pm. Similarly to the preliminary survey, they wore the Narrative Clip on their upper chest facing front, and the device automatically took photos every 30 seconds during the 4 hours. A week later, we first asked the participants to view their lifelog photos via Narrative Clip's official viewer system and recollect what they were doing a week ago. Then, for each event that they recalled, participants were asked which one or which combination of cues out of the 5 cues they could remember. The 5 cues we prepared were "*who, what, screen, where,* and *when*". For example, if a participant remembered having lunch with her friend at the cafeteria, cues on her remembered event would be "*who, when,* and *where*". If she remembered having sandwich for lunch, "*what*" would also be included. In another example, if a participant recalled working on an assignment at his desk in his laboratory, the recalled cues would be "*screen* and *where*". The participants were free to choose any number of cues, therefore the number of possible combinations added up to a total of 31 (5 (only one cue) + 10 (2 cues) + 10 (3 cues) + 5 (4 cues) + 1 (all 5 cues) = 31).

3.2.2 Results

Out of the 31 patterns, the combination of cues the participants chose for each event is shown in Figure 5. The total number of events the participants recalled was 57, and by far the most chosen combination was "*what* and *where*", followed by "*who* and *where*", "*screen* and *where*", "*who, what,* and *where*". An example of cases where "*what* and *where*" were chosen would be: "*I took out a bottle of coke out of the refrigerator*", or "*I was reading a comic book at the sofa*". For cues "*who* and *where*", an example would be "*I was talking to a friend on our way back to the laboratory*" or "*I was discussing a project with a friend near the table*", and so on.

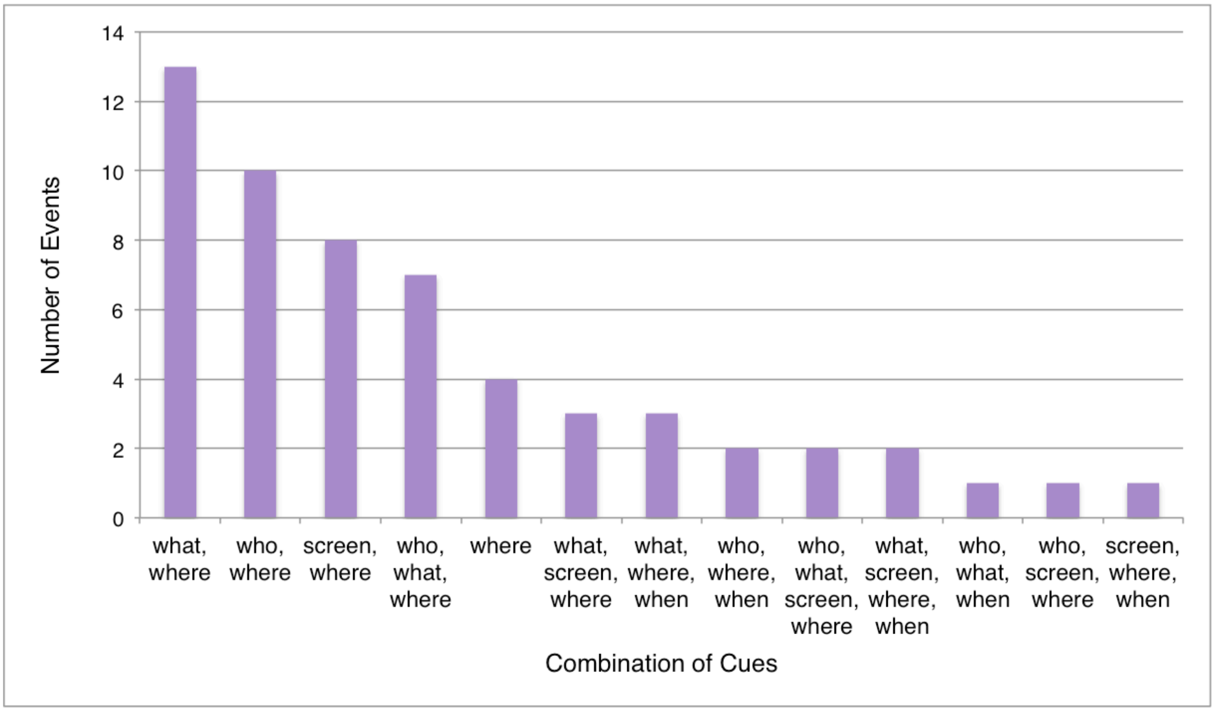


Figure 5. Combination of cues the participants remembered for each event

As we can see from Figure 5, the cue *where* was chosen in every single combination except one, which ranked second from the last. We believe the results point out that information on *where* a person was at the time is well remembered, even after a week, and is a valuable cue in representing an event. Even more, *where* was the only cue that had an effect in recalling memories by itself. The top three combinations each involved *what, who* or *screen,* together with *where*. In this regard, an object or a person is not necessarily involved in daily events, but users are always somewhere, whether the place is in the laboratory, at the cafeteria or moving from one place to another. This might be one of the reasons why *where* as a cue by itself has an important value. Altogether, it can be said that *where* was the most common cue a person remembers about past events.

When it comes to the combination of cues, two cues seems to be enough in representing event. Wagenaar found that "the probability of correctly answering questions increased when more cues were given" [28], and our results confirm his finding. A similar result was claimed by Memon, who said that more than 20 days after the event, "the users were more interested to retrieve lifelogs specific to their current situation by combining various key element combinations" [20]. Therefore, providing users with multiple cues makes it more likely for them to recall the correct and specific moment of their past.

These results conform with Brewer's statement that place, events, and people are strong cues, as they appear on 4th place in Figure 5. We also showed that even stronger combinations of cues exit. However, our results conflict with what Wagenaar stated, i.e. that the combination of "*what* and *when*" are the strongest. This combination did not appear in our results. We will discuss further, in the *Discussion and Future Work* section, what could have caused this difference.

From the preliminary survey we found that *screen* as a cue has revealed its presence in our daily life. As we can see from our additional survey results, participants spent a lot of time in front of screens, especially computer screens. Some might say that computers or smartphones represent objects, therefore *screen* as an independent cue is not approvable, and that it should be counted in as *what*. However, we strongly believe in the presence of *screen* as an individual cue. These objects are not merely objects, because people often recall the interactions they had with them, as explained earlier in *3.1.1 Results*. What the participants did with their computer or smartphone might be: interacting with someone far away; gathering information on some topic using the Internet; studying or reading; and working on task, such as developing a software. From this perspective, we believe *screen* as a cue is valuable in representing a participant's activities.

In summary, we found that the combination of cues "*what* and *where*" was most commonly remembered, followed by "*who* and *where*", "*screen* and *where*", "*who, what*, and *where*". The strongest cue as an individual was *where*, and it was the only cue that participants used when representing their events with a single cue. Compared to previous research, we found that *screen* as a cue is valuable in representing our everyday life.

3.3 System

In this section, we will first discuss our design decisions based on what we have achieved from our preliminary survey and our additional survey. Then, we will describe the environment our system was developed under. Finally, we will describe the architecture of our system by first explaining the overview, then by listing up the detailed mechanism of our system.

3.3.1 Design Decisions

According to Bush, the ultimate version of Memex would be a system which could do the following: "with one item in its grasp, it snaps instantly to the next that is suggested by the association of thoughts, in accordance with some intricate web of trails carried by the cells of the brain" [4]. Mathur et al. found that "linking of lifelog events promotes faster recall in goal-oriented browsing" [19]. Memon clarified that users tend to retrieve a specific moment of their lifelog by combining various key elements [20]. From these works, we can assume that when searching through lifelog for a particular moment, having information hyper-linked would cause a faster recall. There are many situations in daily life where we remember a part of an event but cannot recall the whole story. By having key elements hyper-linked, searching through the archive would be much more helpful since eliciting a cue from another cue would be possible.

From our surveys we found that lifelog allows users to recall their past in a cue-oriented way, and discovered the combination of cues that represent user's events. These cues are "*where, what, who,* and *screen*", and they are equivalent to what Bush referred to as "intricate web of trails". Hence, having these cues hyper-linked would drastically make lifelog search faster and meaningful to the users. For example, when a user is looking for a moment where she went shopping with her friend Kate at the Central Shopping Mall, but cannot remember what she bought, she can first search for the moments that include Kate and Central Shopping Mall as tags. Then our system would provide the user with all the merchandise they bought that day (e.g. a jacket, skirt, and a pair of earrings) as cues on *what*. If the two spent time watching a screen (e.g. a digital map of the mall displayed on a large screen), the cue *screen* would also be provided. By seeing the suggestions made by our system, the user could recall what they bought together. In addition, by entering a name of a specific item or a screen tag as a keyword, the system will reveal photos that include the item or the screen. This would make memory recall much faster than going through photos taken during the 2 hours they spent at the mall, which would add up to approximately 240 photos.

Therefore, we will design our system to save information on photos by *who* they were with, *where* they went, *what* they were interacting with, and what *screens* were involved. When searching for a moment with a keyword, the system will automatically (a) show moments that include the entered keyword and (b) list up cues that are linked with the keyword as a suggestion.

3.3.2 Environment

We created a web application that operates under our local MySQL server. We developed our system on a laptop computer, with a display size of 1366 by 768 pixels. Coding for the frontend was performed in html and JavaScript, and coding for the backend was performed in PHP (see Figure 6). A web browser such as Chrome, Internet Explorer, and Safari is needed to activate the system.

3.3.3 Architecture

In this section, we will describe the architecture of our system in three steps. First, we will explain how the lifelog data is managed in our system. Second, we will we will list the functions our system preforms, from the user's point of view. Third, we will describe in details how the system manages the data in the backend.

Figure 6 illustrates the overview of our system. The wearable lifelogging device, Narrative Clip, first captures photos as lifelog data. Then, our system loads the lifelog data and displays them. The user can select photos then add tags to the selected photos, and search for photos that include the entered keyword. File names of the lifelog data are saved in the database as paths, and the tags the user added are also saved in the database. When the user searches for a keyword, our system retrieves data from the database on tags that include the entered keyword, the paths of photos that include the keyword as tags, and tags linked to the entered keyword. Then the system displays the tags that are linked to the entered keyword, and photos that include the keyword as tags.

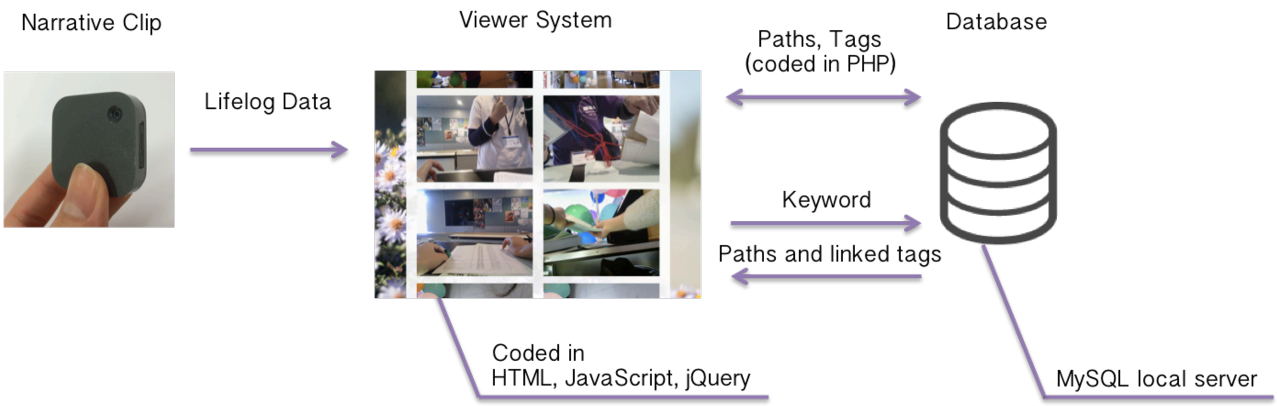


Figure 6: Overview of our system

The user can facilitate our system by performing the following:

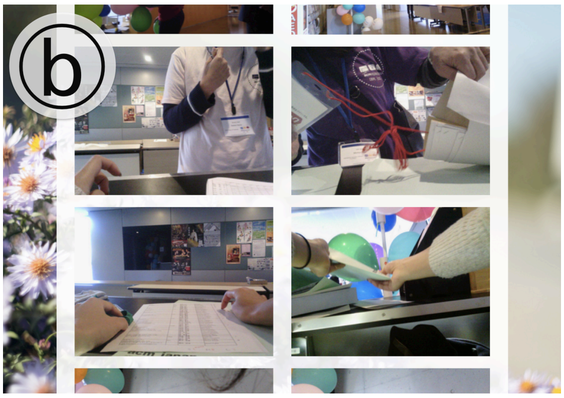
1. Take photos using Narrative Clip. Set up Narrative Clip's official uploader so that copies of pictures will be saved on the local drive when Narrative Clip is plugged into the computer. Move copies of photos to the designated directory.
2. Open the application. The application will automatically load photos from the designated directory and display them in a grid layout.
3. Select photos (see Figure 7). Select single photo by clicking on the image, or select multiple images by pressing the shift key at the same time. A red box will surround the selected images to distinguish it from the others.
4. Edit tags (see Figure 8). When one or more images are selected, click the Edit Tag button located in the upper area of the page. A form will appear with entry spots for the cues *where, what, who,* and *screen*. Enter at least one tag information (multiple tags can be saved for each type of cue by splitting keywords with a space or a comma), then click the Save Tags button.
   Tags can be deleted as well. If tags had already been added to the selected photos, those tags will appear in the tag editing form. Changes will be saved by clearing the existing tags and overwriting them with new ones.
5. Search moments (see Figure 9). Click the Search tab on the navigator then enter one or more keywords in the box. By clicking the Search Keyword button or hitting the enter key, photos that include all the entered keywords will appear in the lower section of the page. In addition, suggestions on tags related to the entered keywords will appear in the middle section.
6. Delete all tags (see Figure 10). Click the View tab on the navigator. A Delete Tags button is located above the photos shown in the gird layout. Clicking this button will erase all tags attached to photos.
7. Add photos. If new photos are taken by Narrative Clip, add copies of photos to the designated directory and reload the View page. New photos will appear in the page and all previous tags will still be attached.

Next, we will explain how the system works in the background. The system consists of 4 modules: View, Edit Tags, Search, and the Delete module.

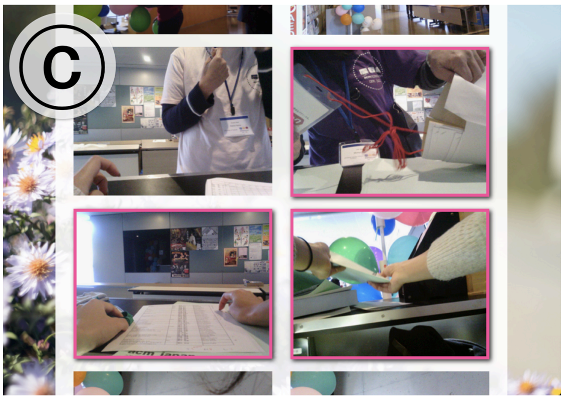
The Viewer module first loads file names with JPEG extensions from the designated directory. File names will be saved for each image div element as an ID for later usage. Photos will be displayed under the View tab (see Figure 7-a) in a grid view, and the user can view the photos by scrolling the page (see Figure 7-b). When the mouse is clicked over an image, a function written in JavaScript will hold the selected file names in a hash with its file name as key and true as value. If the user clicks the image once more, the value with the file name as its key will be switched to false. By clicking while holding down the shift key, the user can select or unselect multiple images. Each time the user performs a click action, values in this hash will be updated and a red box will appear or disappear around the image (see Figure 7-c). With regard of the red box, jQuery is embedded in the Viewer module and each time the user performs a click action, it will first delete all the existing red boxes then show the boxes on images that have their value as true in the hash. Therefore, the selected images can be visually emphasized.

The Edit Tags button is consistently shown in the upper part of the page (see Figure 7-a). If no photos are selected before clicking the button, it will alert to select one or more photos. By clicking this button with one or more photos selected, this module passes the file names of the selected photos to the Edit Tags module.

7-a: Screen right after clicking the View tab on the navigator (Delete Tags and Edit Tags buttons are always shown at the top of the View page).



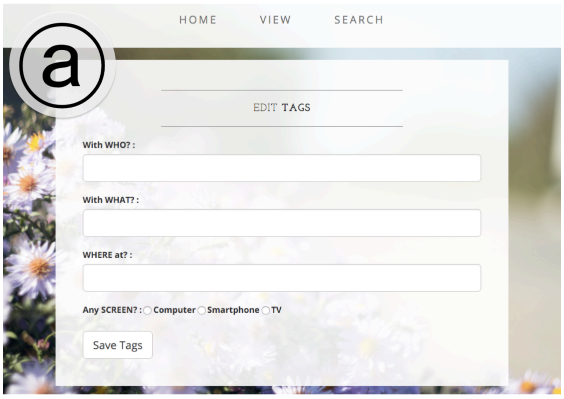
7-b: Lower area of the View page (photos are shown in a grid layout). The users can see more photos by scrolling the page.



7-c: Red boxes appear around the image when selected. In this photo, an image on the upper right corner and two images on the lower row are selected.

Figure 7: Visual image on what the users see under the View tab

The Edit Tags module, first obtains the file names of selected photos from the Viewer module. It then displays a form for the user to input tags (see Figure 8-a). The form has three blank spaces and three-option-radio-button with the questions; "with *who?*", "with *what?*", "*where* at?", and "any *screen?*" written besides them. For the first three questions, users are free to input more than one tag by splitting them with a space or a comma. The user can also choose between computer, smartphone, and television for the question “any *screen?*”. If there are tags already attached to the selected files they will be shown in the form from the beginning. Users must input at least one of the four cues, or else an alert will instruct to do so. Then, when inputs are valid, clicking the Save Tags button will send information to the MySQL database (see Figure 8-b). The database has three individual tables prepared. File names for each photo will be added to a table which holds only the file name of images. The entered tags and their cue (e.g. tag for who: “Kate”, tag for where: “Central Shopping Mall”) will be inserted into a different table. Each row within tables will have a unique ID assigned. An ID for a particular photo and an ID for a particular tag will be added to another table, which contains the links for all the photos and their tags. Every time the user edits tag information, MySQL will search for duplicates and each combination will be kept unique.



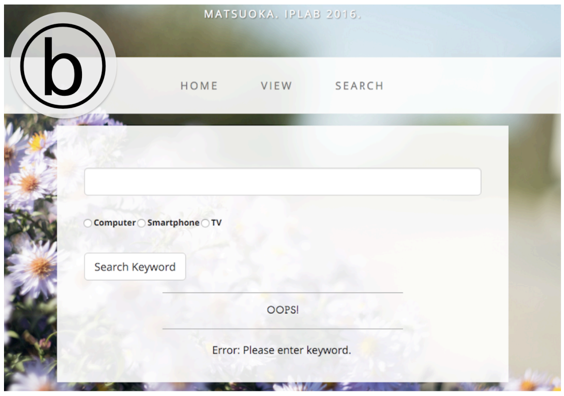
8-a: Screen right after selecting photos and clicking Edit Tags button.

8-b: Screen after clicking Save Tags and saving changes in database.

Figure 8: Visual image on what the users see when editing tags

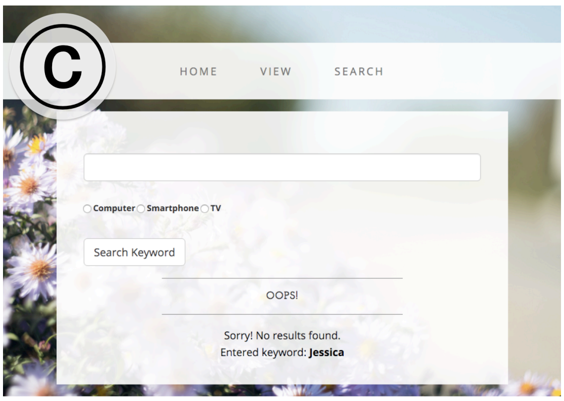
The Search module receives the keyword from the input area under the Search tab on the navigator (see Figure 9-a) then searches through the database for the keyword. If the Search Keyword button was pressed with the box blank, it will alert the user to enter a keyword (see Figure 9-b). If the keyword was not found in the database, it will display a relevant message which also shows the entered keyword (see Figure 9-c). Other than those cases, the system will (1) search for a tag that matches entered keyword, (2) search for photos that have the entered keyword as a tag, (3) if there are multiple keywords, do the same for the next keyword and display only the photos that have all the previous keywords (see Figure 9-d).

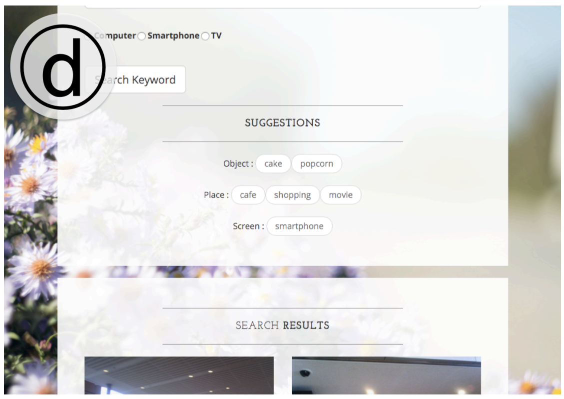
With regard to related keyword suggestion, the Search module will do the following: (1) search for a tag that matches the entered keyword; (2) search for photos that have the entered keyword as a tag; (3) for each photo found, list up all tags and its cues, and calculate how many times it appears; (4) if there are multiple keywords, do the same for the next keyword; (5) for each cue type, sort tags so that they are in descending order of occurrences; (6) for each cue type, display related tags so that the system suggests only the related tags that are not entered as keyword As a result, the system will automatically suggest related tags while avoiding to suggest information the user has already input.



9-a: Screen right after clicking the Search tab on the navigator (search box, screen options, and Search Keyword button is always shown on the top of this page).

9-b: If inputs are blank, error screen shows up telling the user to enter at least one keyword or choose from one of three options on screen tag.





9-c: If the entered keyword is not tagged to any of the photos, an error screen shows up. It also repeats what keyword that was entered.

9-d: If keyword search is successful, the system displays photos with entered keyword linked as tags. At the same time, suggestions on related tag shows up in the middle section of the page.

Figure 9: Visual image on what users see when searching for tags

The last module is the Delete module. When the Delete Tags button under the View tab is clicked, this module will delete data and reset ID column for each table (see Figure 7-a and 10). It will not delete the images saved in the designated directory, hence it will only reset the database.

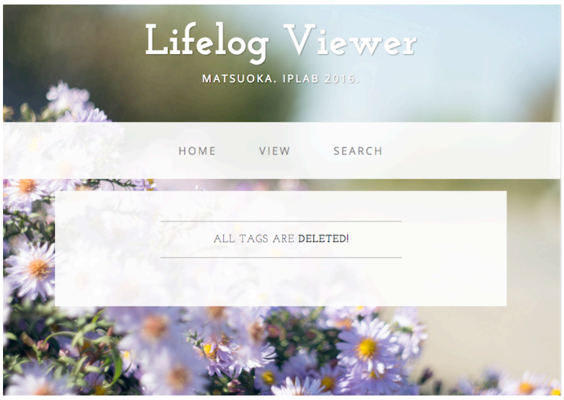


Figure 10: Visual image on what the users see after clicking the Delete Tags button under the View tab

Chapter 4
Discussion and Future Work

In this section, we will discuss how our results conflicted with previous work and what might have caused the difference. We will also discuss how our system draws the line with existing photo management applications, and consider what our future work will be.

The results of our additional survey confirmed Brewer's statements but conflicted with Wagenaar's statements, as we mentioned earlier. We assume that the difference was due to the fact that Wagenaar provided cues first then tested if the participant (which was himself) could recall the event correctly. However, in our method, the participants first thought of a particular moment, then to answered which cues they can relate to the event. In cases where the user does not have a specific moment in mind that he would like to recall, our results on strongest combination of cues that cause quick memory recall might be similar to his conclusion. The methods of finding valuable cues were different from those employed by Wagenaar. We believe that when participants are capturing their daily activities as lifelog, and are interested in searching for a moment from the archive, they should have a specific moment in mind they would like to view information on. Therefore, we confidently valued the combination of cues we discovered, which were several patterns of combination of *where, what, who*, and *screen* cues.

There are existing services that manage photo display, adding tags, and tag search such as flickr [10], Tumblr [26], and Instagram [15]. The difference between these systems and our system is that while the existing services aim to socially share photos and interact with others via sharing, our system aims to support memory recall of one single user. Our system assists editing tags by having a spot ready for the valuable cues (*who, what, where* and *screen*). And because our surveys proved that people rely on those cues to represent their experiences, saving information on those cues and suggesting tags that are tagged along the input keyword will definitely contribute to memory recall.

The limitation in our system is that it requires the users to manually select photos and enter tag information. If these steps could be done automatically it would relieve a lot of the user's burden. Doherty et al. presented a method of combining and organizing similar images into moments [7]. Lee et al. developed a system which automatically, selects the most efficient cue within captured lifelog [17]. If combined with these technologies, our system could organize photos into moments and add tags on *where, what, who*, and *screen* without the need for user involvement. In our future work we intend to consider a similar process, eliminating this burden and contributing to further spread of lifelogging systems.

Conclusion

From the preliminary survey that we conducted, we highlighted the importance of providing the users with cues that directly evoke memory recall. Independent recalls are fulfilled by chronological recalls on memories, whereas lifelog-dependent recalls are elicited directly by seeing cues within lifelog. Therefore, designing an interface that provides the user directly with information that are efficient in memory recall is desired.

The additional survey revealed that *where* was the strongest single cue and that the combination of cues "*what* and *where*", followed by "*who* and *where*", "*screen* and *where*", "*who, what,* and *where*" were also strong. Moreover, our surveys found that *screen* as a cue has a large impact in our daily lives and that it can represent many events a user has experienced.

With our theoretical explanation on which cues to focus on, we designed our system to efficiently evoke memory recall by linking not only photos with tags but also tags with other tags. The system suggests tags linked with the entered keyword, with the aim to support memory recall in cases where the user is not confident about entering the proper keyword and finding the exact moment in his mind.

Our contribution to this field is that we provided theoretical explanations on how a lifelogging system supports memory recall; furthermore, we pointed out which cues are effective in evoking memory recall, especially in our modern lives where technology is rapidly evolving.

Acknowledgements

The author would like to thank Jiro Tanaka, Shin Takahashi, and Buntaro Shizuki for their help on running and analyzing the study. We are also thankful to our participants for sharing their lifelog data with us and participating in our surveys, Itsuki Tada and Daiki Yamaji for their advice on programming, and Simona M. Vasilache and Hani Karam for their review and comments on this paper.

References

[1] Kiyoharu Aizawa, Datchakorn Tancharoen, Shinya Kawasaki, and Toshihiko Yamasaki. Efficient retrieval of life log based on context and content. In CARPE 2004, p22-31.

[2] Gordon Bell and Jim Gemmell. A Digital Life. In Scientific American, March 2007, p58-65.

[3] William F. Brewer. Memory for randomly sampled autobiographical events. In Remembering Reconsidered: Ecological and Traditional Approaches to the Study of Memory, p21-90, 1988.

[4] Vannevar Bush. As We May Think. In The Atlantic Monthly, July 1945, p101-108.

[5] Scott Carter and Jennifer Mankoff. When Participants Do the Capturing: The Role of Media in Diary Studies. In CHI 2005, p899-908.

[6] Felicia Cordeiro, Elizabeth Bales, Erin Cherry, and James Fogarty. Rethinking the Mobile Food Journal: Exploring Opportunities for Lightweight Photo-Based Capture. In CHI 2015, p3207-3216.

[7] Aiden R. Doherty, Ciarán Ó Conaire, Michael Blighe, Alan F. Smeaton, and Noel E. O'Connor. Combining Image Descriptors to Effectively Retrieve Events from Visual Lielogs. In MIR 2008. p10-17.

[8] Susan Dumias, Edward Cutrell, JJ Cadiz, Gavin Jancke, Raman Sarin, Daniel C. Robbins. Stuff I've Seen: A System for Personal Information Retrieval and Re-Use. Microsoft Research.

[9] Margery Eldridge, Micheal Lamming, and Mike Flynn. Does a Video Diary Help Recall? EuroPARC Technical Rep. No. EPC-1991-124.

[10] flickr. Retrieved January 28, 2016, from https://www.flickr.com/

[11] Jim Gemmell, Gordon Bell and Roger Lueder. MyLifeBits: A Personal Database for Everything. Microsoft Bay Area Research Center, February 2006, p1-18.

[12] Jim Gemmell, Lyndsay Williams, Ken Wood, Roger Lueder, and Gordon Bell. Passive Capture and Ensuing Issues for a Personal Lifetime Store. In ACM CARPE 2004, October 2004, p48-55.

[13] Testuro Hori and Kiyoharu Aizawa. Content-based video retrieval system for the life-log applications. In MIR 2003 Proceedings of the 5th ACM SIGMM international workshop on Multimedia information retrieval, p31-38.

[14] David Huynh, David Karger, and Dennis Quan. Haystack: A Platform for Creating, Organizing and Visualizing Information Using FDR. Semantic Web Workshop, 2002.

[15] Instagram. Retrieved on January 28, 2016, from https://www.instagram.com/

[16] Vaiva Kalnikaite, Abigail Sellen, Steve Whittaker, and David Kirk. Now Let Me See Where I Was: Understanding How Lifelogs Mediate Memory. In CHI 2010, p2045-2054.

[17] Matthew L. Lee and Anind K. Dey. Lifelogging Memory Appliance for People with Episodic Memory Impairment. In UbiComp 2008. pp44-53.

[18] Steve Mann. Wearable Computing: A First Step Toward Personal Imaging. In IEEE Cybersquare, February 1997, p25-32.

[19] Akhil Mathur, Anirban Majumder, Samik Datta, Sreedal Menon, Shipra Malhorta, and Ankur Dahiya. LifeView: A Lifelog Visualization Tool for Supporting Sentimental Recall and Sharing. In OzCHI 2012, p371-380.

[20] Mohsin Ali Memon. Ensuring Privacy and Sharing Framework for Pervasive Logging. University of Tsukuba, July 2014.
[21] Microsoft Research SenseCam analysis. Retrieved on January 28, 2016, from http://research.microsoft.com/en-us/um/cambridge/projects/sensecam/memory.htm
[22] Narrative Clip website. Retrieved on January 28, 2016, from http://getnarrative.com/
[23] Abigail Sellen, Andrew Fogg, Mike Aitken, Steve Hodges, Carsten Rother and Ken Wood. Do Life-Logging Technologies Support Memory for the Past? An Experimental Study Using SenseCam. In CHI 2007, p81-90.
[24] Abigail Sellen and Steve Whittaker. Beyond Total Capture: A Constructive Critique of Lifelogging. In Communications of The ACM, May 2010, Vol. 53, No. 5, p70-77.
[25] Endel Tulving. Episodic memory: from mind to brain, Episodic Memory: from Mind to Brain, 53 (2002), p. 1-25.
[26] Tumblr website. Retrieved January on 28, 2016, from https://www.tumblr.com/
[27] Information on Vicon Revue. Retrieved on January 28, 2016, from https://www.viconrevue.com/product.html
[28] Willem A. Wagenaar. My Memory: A Study of Autobiographical Memory over Six Years. In COGNITIVE PSYCHOLOGY 18, p225-252, 1986.
[29] Rui Wang, Fanglin Chen, Zhenyu Chen, Tianxing Li, Gabriella Harari, Stefanie Tignor, Xia Zhou, Dror Ben-Zeev, and Andrew T. Campbell. StudentLife: Assessing Mental Health, Academic Performance and Behavioral Trends of College Students using Smartphones. In UbiComp 2014, p3-14.