Retrieving Episodic Memory in Lifelog System Focusing on Daily Conversation

Yiwen Dong
44171611-9

Master of Engineering

Supervisor: Prof. Jiro TANAKA
The Graduate School of Information, Production and Systems
Waseda University

July 2019

Abstract

Lifelog can provide a detailed chronicle of people's life by using wearable technology, which involves a process of capturing, processing and recalling life experience passively. Episodic memory which is the collection of past personal experiences, is vital to human for maintaining past and anticipating future. Therefore, lifelog is usually used for retrieving episodic memory. With appropriate cues, images taken by wearable cameras can help lifeloggers recalling their past experiences. Most existing cues are focusing on contextual information which can be captured from lifelog images or wearable sensors. Also, conversation contains important information which can help lifeloggers retrieve their episodic memory. However, this important user-related memory trigger of conversation is not involved in current research. In order to improve the situation, we proposed other memory cue to extract important memory for lifeloggers. This memory cue is conversation cue, which includes keyword of conversation, speakers and emotions.

Our system consists of two parts: a web-based lifelog viewer for lifeloggers to conveniently retrieve memory, a conversation analyze system to analyze lifeloggers' conversation using voice recognition, speaker recognition and sentiment recognition. In order to retrieve our goal, we integrate two hardware devices into our system: Autographer hang on lifeloggers' neck for getting lifelog images and a smart phone to record lifeloggers' daliy conversation. By using our proposed system, lifeloggers can efficiently browse lifelog images taken during the whole day. Moreover, we also combine our user-related cues with traditional contextual cues in our system.

We have invited some participants to test the usability and efficiency of our system. The result have shown the positive potential of aiding episodic memory by using our approaches.

Keywords: Lifelog, Episodic Memory, Conversation, Autographer

Acknowledgements

Thanks to Professor Jiro TANAKA who gives me great advice and points me the right direction of my research. Thanks to him for opening my mind, finding my mistakes and helping me fix them. Without his patient guidance, I won't complete this. And also I want to thank my laboratory members for inspiring me, taking my experiments and encouraging me when I felt hard.

Table of contents

List of figures

List of tables

Chapter 1

Introduction

1.1 Introduction

1.1.1 Lifelog

The Life Log is a detailed chronicle of a person's life, containing a large amount of data that can be automatically captured by wearable technology. [1] This is a passive process of collecting, processing and recalling life experiences. Individuals defined as "Life Recorders" will implement a variety of sensors that can sense an individual's living environment and activities. [2]

Lifeog was first introduced to the world by Steve Mann[3] in the 1980s. Mann uses wearable cameras to continuously capture physiological data and real-time first-person videos. Since then, lifeog has gone through many other experiments that researchers have done. With the rapid development of wearable storage technology, life record devices have become more and more popular and easy to use. Representatives include SenseCam, Vicon Revue, Narrative Clip and Autographer. The study of life logs focuses on three areas: (1) using life logs and other wearable technologies to aid memory retrieval. (2) Use new interactions to share lifelines. (3) Discuss the privacy of the life log. Moreover, regardless of research, the life log requires viewers to see images taken with a wearable camera. A typical lifeline viewer is based on three platforms: Web, desktop, and smartphone. Among them, the network stands out for its superior performance. Lifelog Web Viewer is graphical and easy to navigate. Thanks to its platform independence, the Lifeline Recorder can easily interact with the Lifeline Viewer via a web browser on any platform.

Figure 1.1 (a) is the Autographer, a wearable camera widely used in biopsy. Developed by OMG Life, Autographer is a wearable technology that uses five different sensors to detect environmental changes and automatically take a photo every 30 seconds.[4] Figure 1.1(b) shows a partial sample image taken by the Autographer. We can clearly determine the environmental conditions of the lifeline recorder to help retrieve memory.

(a) Autographer

(b) Sample Images

Fig. 1.1 Overview of Autographer

1.1.2 Episodic Memory

Memory is the ability of our nervous system to encode, store and retrieve information. [5] It is the accumulation of a person's past activities, feelings and many other information. Memory can be divided into many categories according to different methods. As shown in Figure 1.2, the most common method is to classify memory into long-term memory, short-term memory, and sensory memory based on retention time. Episodic memory is part of long-term memory that, together with semantic memory, forms the entire memory system.

The concept of episodic memory first appeared around the 1970s. Episodic memory is the memory of autobiographical events, which is different from semantic memory, which focuses on the true facts of the world. These autobiographic events include background information (who, what, when, where and why) that can be clearly expressed and memorized.[6] In general, semantic memory helps us gain objective knowledge, and contextual memory helps us to recall past experiences and shape our character. Situational memory enables people to vividly trace back and recall events that occur at specific times and places.[7] For example, you can recall what happened during your first date and what happened when you last ate. These are the results of recalling the memory of the plot.

Episodic memory is very important to human beings and is considered to be a “true natural miracle”.[8] Also, according to Stanley B. Klein’s experiment,[9] the loss of episodic memory not only affects the individual’s memory of the past, but also expands the influence of people’s ability to predict the future. However, Cabeza R’s research suggests that brain regions involved in episodic memory recognition are much more active than brain regions involved in episodic memory. At the same time, due to brain size limitations and some serious memory diseases, it is necessary to find some external methods to help restore episodic memory.[10]

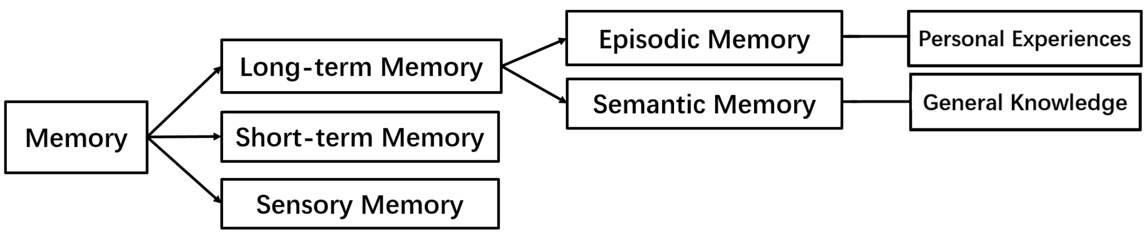


Fig. 1.2 Composition of memory

Ubiquitous lifelogging systems use wearable or embedded sensor technologies such as cameras, audio recorders, location trackers, and physiological sensors to passively and automatically record a user’s personal experiences. Lifelogging technologies allow people with EMI to automatically record, review, and thus regain an awareness of meaningful personal experiences in their lives to maintain their sense of self. Sellen et al. showed that episodic details from a visual "lifelog" can be presented to users as memory cues to assist them in remembering the details of the original experience. Other successful systems leverage other modalities such as the Audio Memory Prosthesis that records audio from personal experiences and presents it as searchable and browsable text. The Personal Life Log system uses a combination of location sensors, physiological sensors, and real-time voice annotation to identify potentially interesting scenes in a continuous video log. [11]

Memory Cue is a stimuli that triggers a memory in human mind. For example, if several central words of a conversation between you and a colleague would be shown to you, the memory of that conversation might be triggered in your mind. Schacter et. al. defines memory cue as "external information that is associated with stored information and helps bring it to mind" [12]. In this study, we utilize the notion of memory cue in our summarization system to help trigger personal memories of the past in one's mind. [2]

1.2 Organization of the Thesis

The rest of the thesis is organized as follows: In chapter 2, we describe the research purpose and the approaches we use. Chapter 3 will give the detailed introduction of how the target system is designed. Chapter 4 presents the implementation of how to utilize sentiment cue and movement cue in the system. Chapter 5 is about the evaluation which is used to prove the usability and efficiency of our system. Chapter 6 states 2 related work and some background on emotion detection and activity detection. Chapter 7 concludes the work and will discuss the future expectation.

Chapter 2

Goal and Approach

2.1 Problem

Episodic memory is vital to human for supporting life activities, including life management, social interactions and problem solving in daily life. Due to the limitation of brain function, aiding episodic memory using lifelog system has become more and more popular. This new approach not only can benefit those with significant memory impairment, but also has the potential of making normal individual's life more effective.

For aiding memory, we need to define memory triggers in lifelog system, which is defined as cue[13]. Most existing cues in lifelog research focus on contextual information like how, what, when, where and who. Representative is Lee ML [14] who use objects detected from lifelog image as memory cue and Chowdhury S [15] who involved GPS heat map as memory cue. We can find that these cues have nothing to do with users' own behavior and care little about users themselves.

Among all the interesting lifelogging applications are health care applications such as reviewing past memories for human memory enhancement and support of failing memories by being able to search these archives. In addition to memory enhancement and treating memory deficiencies a person reviewing summaries of his social interactions could be benefited of increased self-awareness, being able to plan the future better, etc.

Therefore, according to the investigation by Sellen A J [16], psychology as design framework can help define the types of memory. Moreover, in our daily life, important memory always happens with the following scenarios:

- Special keywords in conversation: Episodic memory always happens when there is a conversation. The keyword of the conversation may represent the episodic memory at that moment.
- Speaker: When have a conversation with others, the speakers in the conversation may help lifelogers retrieve the episodic memory at that moment.
- Sentiment: The sentiment extract from voice may contains information of memory when having a conversation.

These three scenarios can be summarized as user-relative and research in aiding episodic memory using lifelog discard user's conversation when considering memory cues. This is the problem we want to improve by using our proposed method.

2.2 Goal

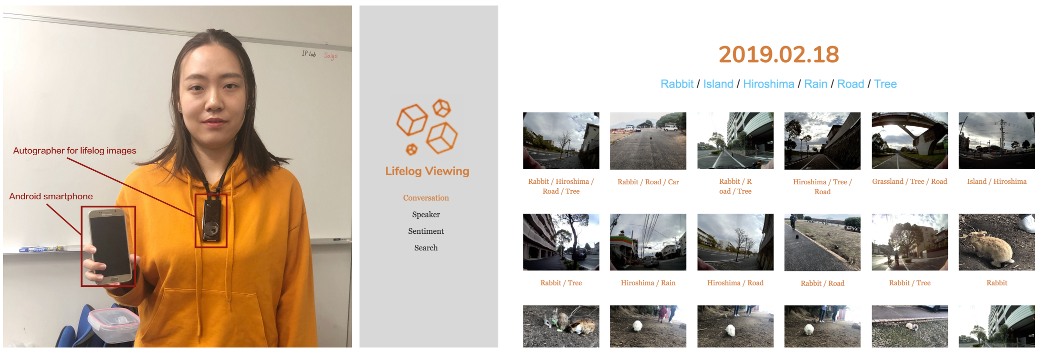
In this research, we aim to involve user's conversation into lifelog system, so as to help finding important memory for lifeloggers. Here, user's conversation including speaker, keyword of conversation and sentiment. For a clearer explanation of our research perspective, we corresponds the user status to three cues:

- Keyword of conversation
- Speaker
- Sentiment

2.3 Approach

Our proposed system mainly consists of two parts: a wearable sensor system and a web-based lifelog viewer. Unlike traditional contextual lifelog cues, we introduce personal daily converstion as new cues to enhance the effectiveness of memory recall. To achieve this goal, we set up a wearable sensor system(see Fig 2.1) . We use android smartphone to record user's daily conversation, which is used in system process module. Meanwhile, we use an Autographer hang on user's neck to capture lifelog image constantly.

By using our wearable sensor system, we can get user's daily conversation. After that, these data will be input into the process system to analyze the keyword, speaker and sentiment when each lifelog image is taken. Because of the cross-platform compatibility and usability of web, then we build a web-based lifelog viewer with the output. The appearance of the viewer is shown in Fig. Lifeloggers can easily access important memory by using three proposed new cues in the viewer.



(a) Hardware Overview (b) Lifelog Viewer

Fig. 2.1 Proposed System

2.4 Novelty

1. To retrieving episodic memory needs both visible and audible information. In order to enhance the efficiency of retrieving user's memory, we innovatively put forward voice cues, which captured by android phone from daily conversation.

2. We use android smartphone to record user's daily conversation which is convenient and natural. Meanwhile, speech recognition and processing technology is mature enough for current research, so we can get accurate data from conversation.

Chapter 3

System Design

3.1 System Overview

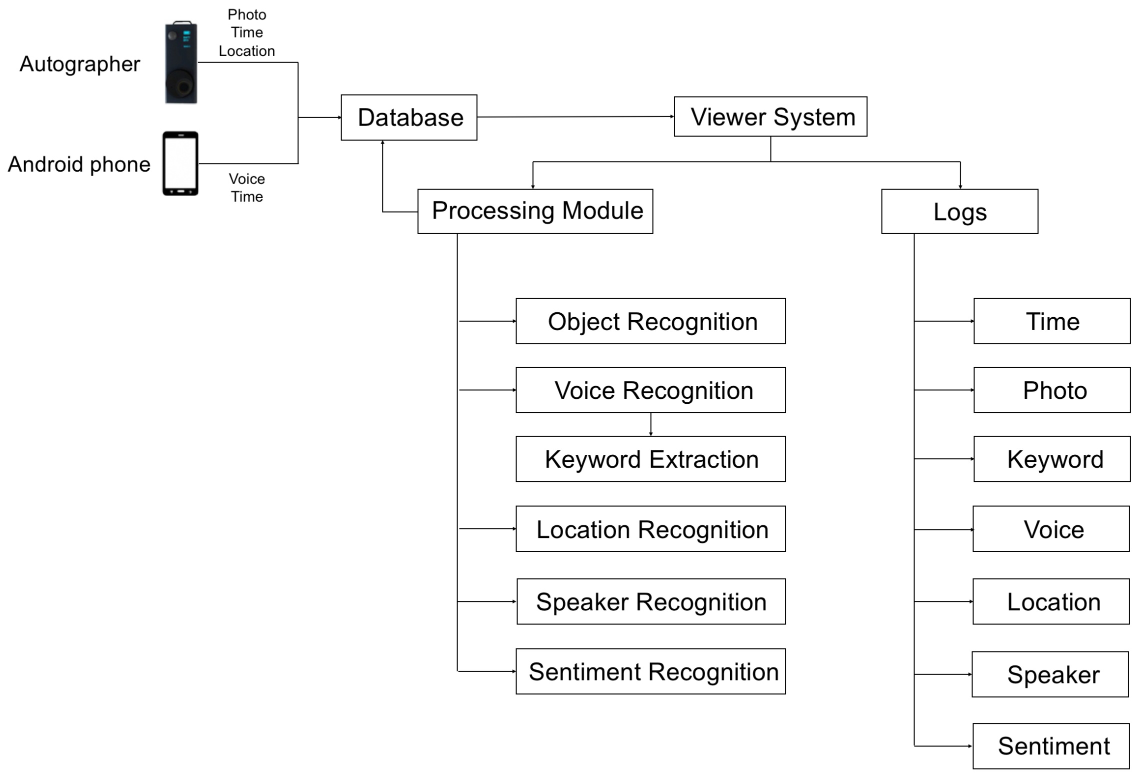


Fig. 3.1 System Overview

Our proposed method aimed to enhance the efficiency of aiding episodic memory. To achieve our goal, we come up with the idea of providing conversation cues for lifeloggers. The system overview is shown in Fig. 3.1. The system is mainly consisting of one web-based lifelog viewer and three accessory sub systems: a object recognition system used to recognize the object from the image, a conversation recognition system used to recognize the keyword, speaker and sentiment from conversation and a location recognition system to recognize the location where the image was taken. We use autographer and android phone to record photo, time, voice and location in users' daily life. Then the processing module process these data. These processed data combine into one log, and we built a web viewer using these logs to help user retrieve their episodic memory.

3.2 Use Case Diagram

Fig 3.2 shows the use case diagram of the system. The lifelog system has four function: record data, process data, view log and search log, in which the user could view the log and search the log. For search log, the user could search through time, conversation, object and location.

3.3 Data Record

3.3.1 Autographer

Autographer is a wearable technology that uses five different sensors to detect environmental changes and automatically take a photo every 30 seconds.[4] Also, each taken image will contain the information of geo-tagging and environmental data as figure3.3 shows. Using autographer we can record image, time, location data every 30 second. In the example, left side shows the image is taken from 2019/02/10 19:14:41. For the image detailed information showing in the right side, it record the latitude 33, longitude 130 and altitude 31. By the value of latitude, longitude and altitude, we can get the location information of the taken image.

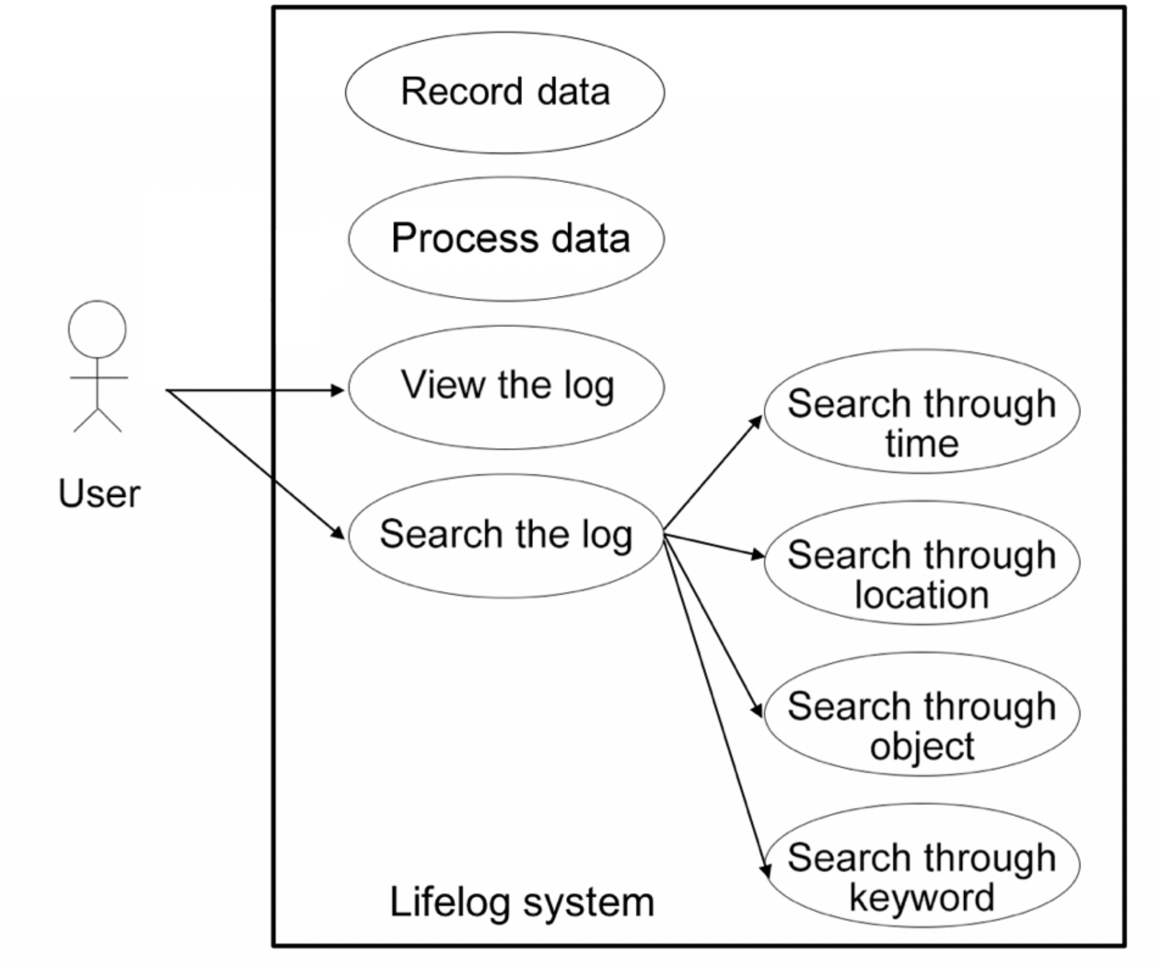


Fig. 3.2 Use Case

3.3.2 Smartphone

We use smartphone to record voice of user's daily conversation. To get the necessary data, we choose an application called Voice Recorder on android smartphone. In this application, it can convert the voice to text in the mean time when user is speaking. After recording, we can get both audible file and text file for next processing part as figure3.4 shows. For audible file, it is used to retreve user's episodic memory when user click the keyword of conversation. For text file, it is used to process the keyword of this conversation and present the keyword to users.

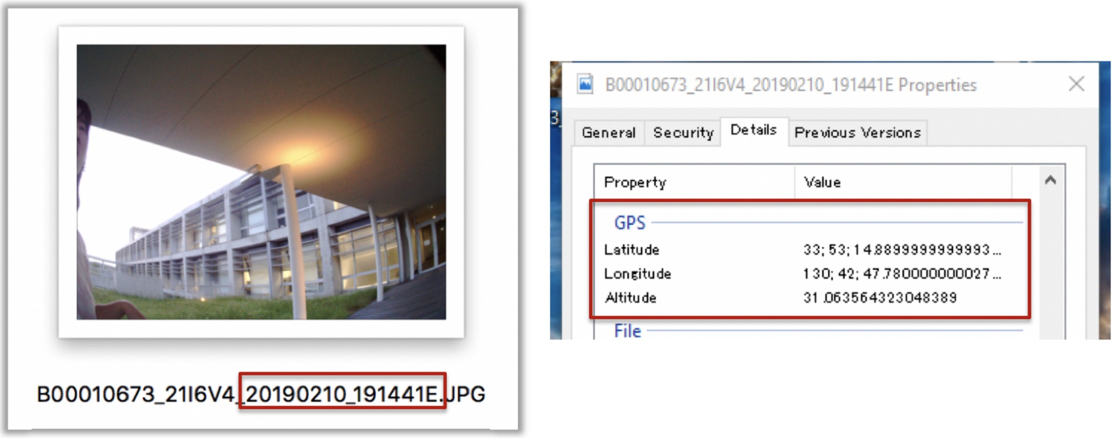


Fig. 3.3 Data Record

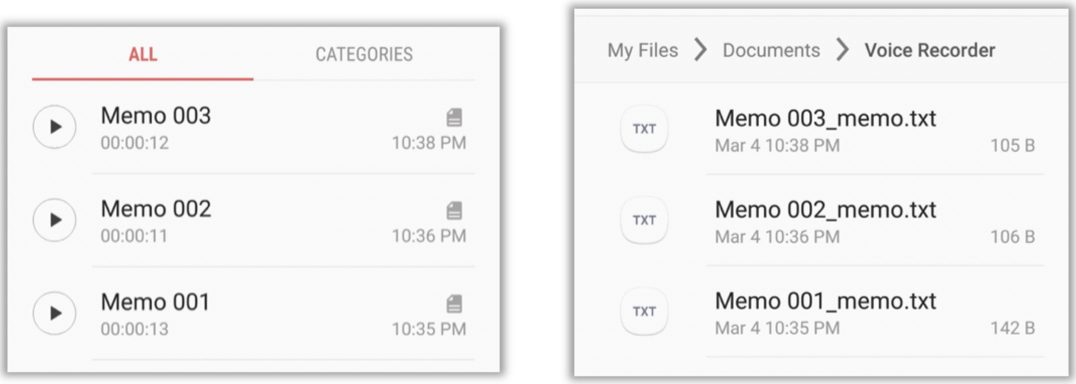


Fig. 3.4 Data Record

3.4 Conversation Cue

3.4.1 Keyword

Fig3.5 shows how we record conversation. Firstly, we record daily conversation by the application Voice Recorder in android smartphone, using the application could convert the voice into text in the meantime when user is speaking.

Fig3.6 shows how we extract keyword from conversation. After recording, we can get both audio and text file. The text file is used to extract keyword by Microsoft Azure. Text Analytics of Microsoft Azure can detect sentiment, key phrases, and language from your text. The API returns a list of strings denoting the key talking points in the input text.

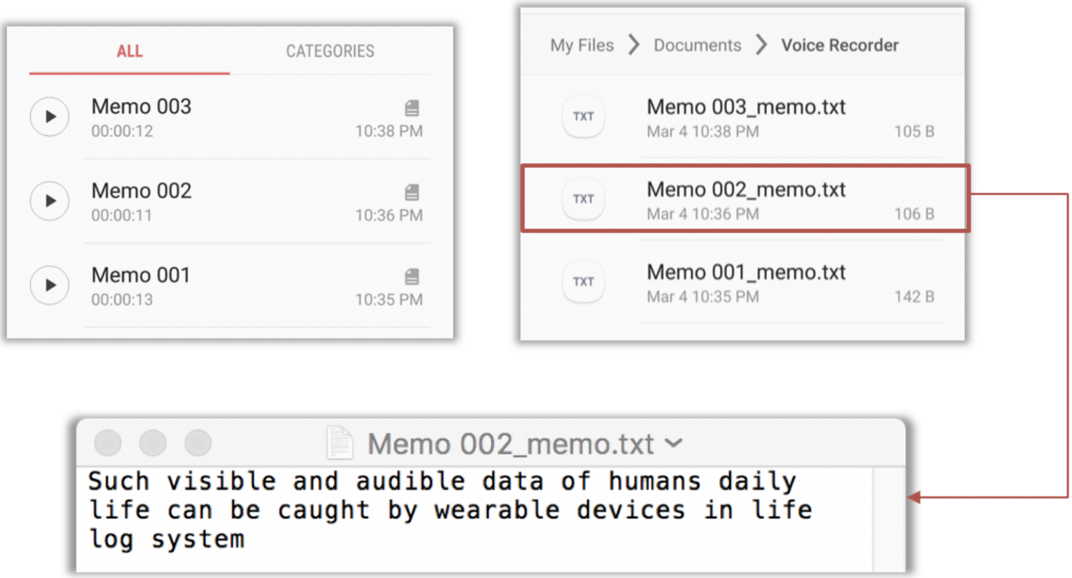


Fig. 3.5 Voice Recorder

They employ techniques from Microsoft Office's sophisticated Natural Language Processing toolkit. English, German, Spanish, and Japanese text are supported. The audio file could used to play when user click the keyword extract from the conversation in web viewer. By recording and processing, we can get the keywords of one conversation and the voice of the conversation, which can be used for user to retrieve their episodic memory.

Fig3.7 shows the system overview of conversation cue. The web page contains date, keword of conversation, image and keyword of each image. By looking through conversation cue, user can retrieve episodic memory when there was conversation happens. In the

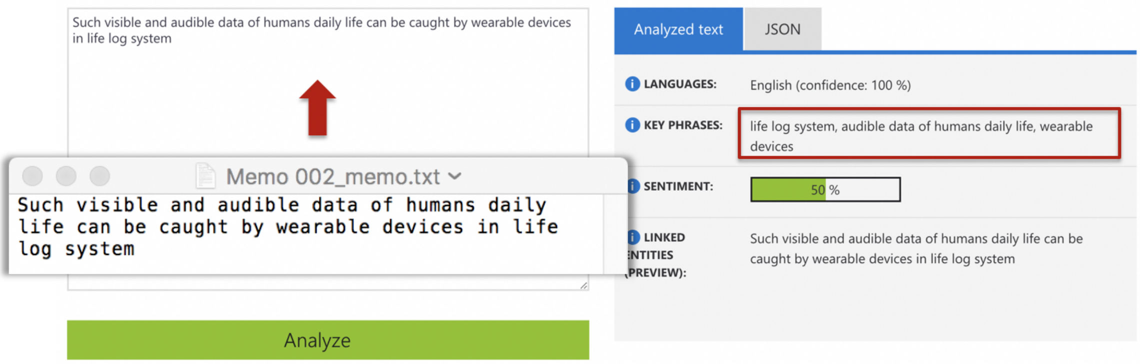


Fig. 3.6 Keyword Extration

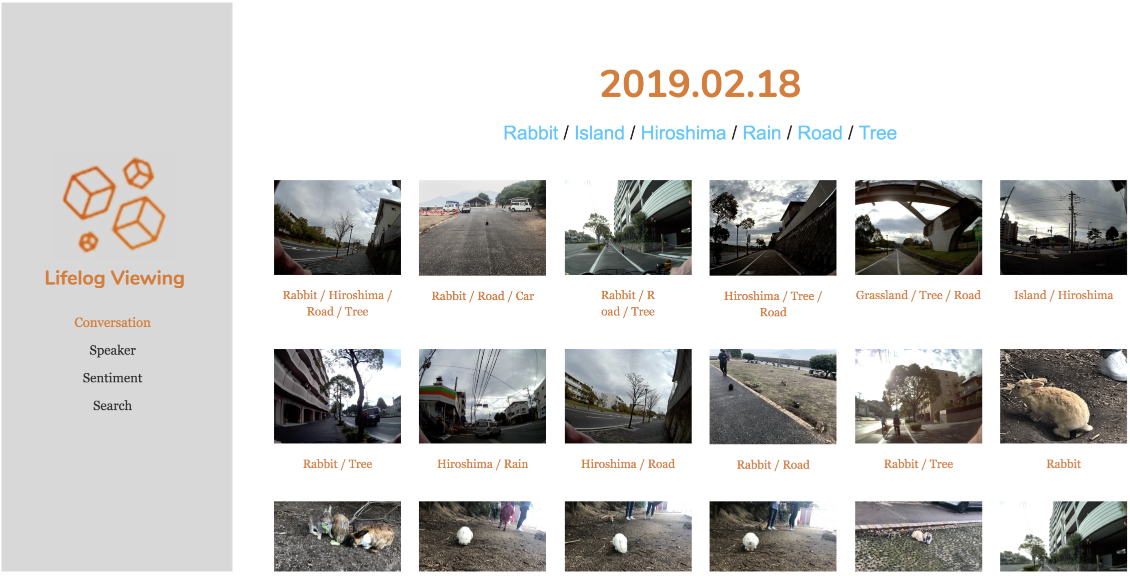


Fig. 3.7 Conversation Cue

example, it shows in the data 2019/02/18. For the whole day, the most exist keyword is Rabbit/Island/Hiroshima/Rain/Road/Tree. The image is filtered when there is conversation happened. Also, the word under the image shows the keyword of conversation when the image was taken.

Fig. 3.8 Keyword Rabbit

Fig3.8 shows the result when click the keyword rabbit, it contain the keyword, rabbit. Also, it shows all images contains the keyword rabbit. The keyword is extract from the conversation when the image was taken.

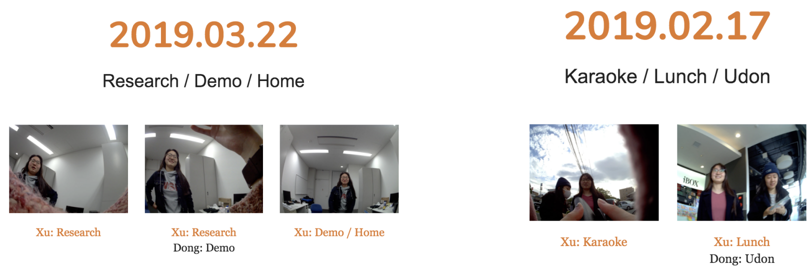
3.4.2 Speaker

Fig. 3.9 Speaker

We can recognize different speakers of the conversation by Microsoft Azure Speaker Identification. Microsoft Azure Speaker Identification Identify individual speakers or use speech as a means of authentication with Speaker Recognition. It can identify who is speaking. The API can be used to determine the identity of an unknown speaker. Input audio of the unknown speaker is paired against a group of selected speakers, and in the case there is a match found, the speaker's identity is returned. By using Speaker Identification we can get different speaker from the conversation. Also, the system summarize the speakers all conversation for user to browse. In the example as figure3.9 shows, there are three speakers recognized by Microsoft Azure Speaker Identification. The system shows each person's image taken by the autographer and their name under each image. When we cilck each person, it will show the detailed information when user talked with the specific person like figure3.10 shows.

3.4.3 Sentiment

Figure 3.11 shows, by using Microsoft Azure Text Analyze, we can recognize different sentiment of the conversation. The API returns a numeric score between 0 and 1. Scores close



(a) Date1 (b) Date2

Fig. 3.10 Logs of Speaker

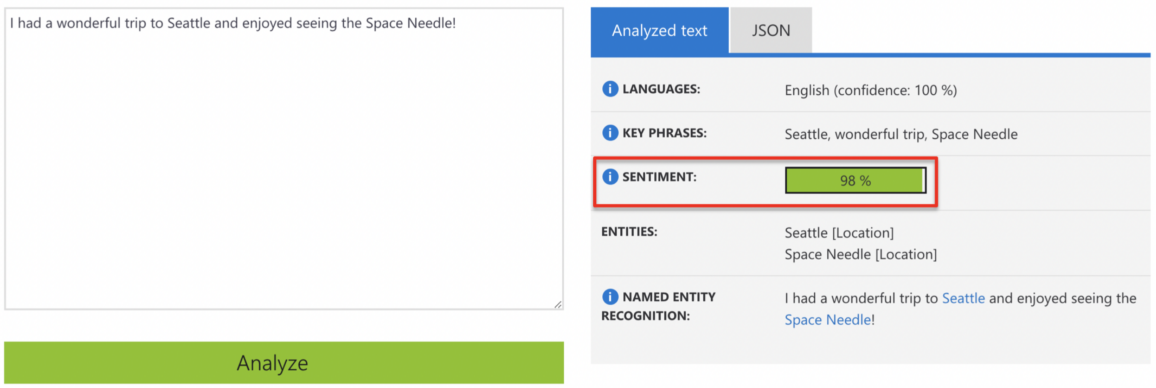


Fig. 3.11 Sentiment API

to 1 indicate positive sentiment, and scores close to 0 indicate negative sentiment. Sentiment score is generated using classification techniques. The input features of the classifier include n-grams, features generated from part-of-speech tags, and word embeddings. It is supported in a variety of languages. User can use the web viewer to browse logs of different sentiment. We sort the sentiment into three kinds. Happiness, calm and sad as figure3.12 shows. For the value from 0%-33%, we detect the sentiment as sad. For the value from 34%-67%, we detect the sentiment as calm. For the value from 68%-100%, we detect the sentiment as Happiness.

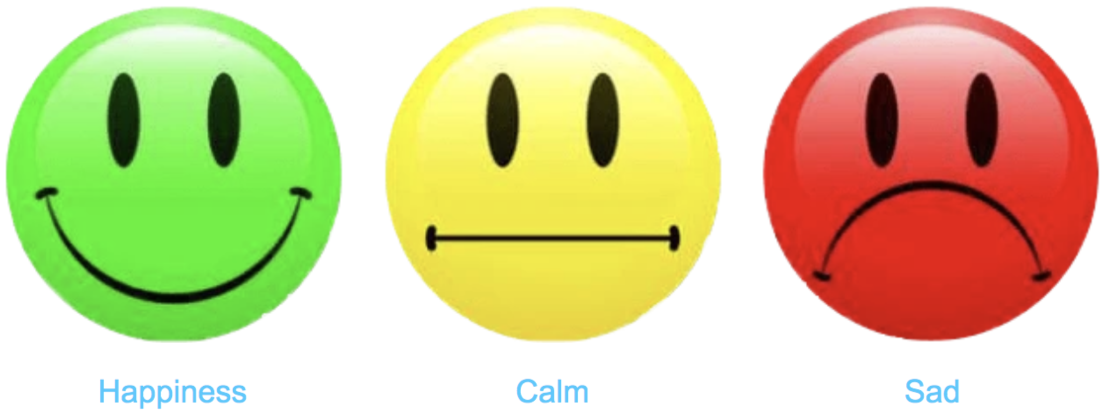


Fig. 3.12 Sentiment

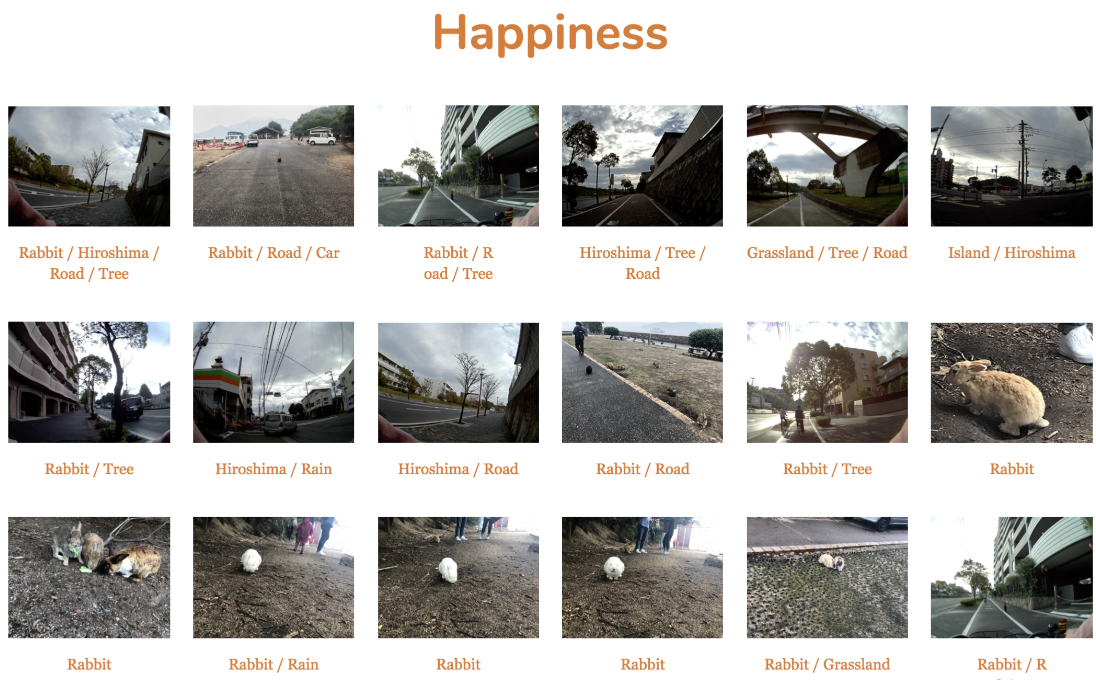


Fig. 3.13 Happiness

3.5 User stories

When using the lifelog system, different user will choose different function to achieve their goal. Here, we give four kinds of user stories to show how to use our system with different goal.

3.5.1 Story1

User: The user who wants to retrieve the memory of specific day.

Task: The user can look through the conversation cue, the conversation cue shows the specfic days logs. It not only contains data, but also contains image when having the conversation, keyword extract from the conversation and could here the recorded voice when the user click the specific keywords.

Goal: Retrieve the memories of that day. The memories include what kinds of event the user has done, who did the user talk with and what kinds of emotion when the user is having a conversation.

3.5.2 Story2

User: The user who wants to retrieve memory when talking with the specific person.

Task: The user can look through the speaker cue, the speaker cue shows the image and name of each person. When user click the specific person, the system will show the result of logs when the user talking with that person. The logs include the image, keyword of the conversation when the image was taken and recorded voice when user click the keyword blow the image.

Goal: Retrieve the memories which happens when the user talked with one specific person. The memories include what kinds of event was happen when user talking with that person, what kinds of topic is it when having the conversation.

3.5.3 Story3

User: The user who wants to retrieve their sentiment of past.

Task: The user can look through the sentiment cue, the sentiment cue shows three kinds of sentiment: happiness, calm and sad. The user could click any sentiment to look through the logs of each sentiment. Each sentiment includes image, keyword extract from conversation when the image was taken and the recorded voice when user click the keyword blow the image.

Goal: Retrieve the memories when the user has different sentiment. The memories include what kinds of event happened when the user is happniess or calm or sad, what conversation did he talked when the user is happniness or calm or sad.

3.5.4 Story4

User: The user who wants to retrieve the specific memory of past.

Task: The user can look through the search function, search function allows user choosing specific conditions of time, location, object, keyword of conversation, speaker and sentiment. By analysing the choosen conditions, the system give the result which satisfy the user's choose. The final result includes date, image and keyword of conversation.

Goal: Search the specific memory of past by time, location, object, keyword of conversation, speaker and sentiment.

3.6 Usage Scenario

Lifeloggers need to wear wearable devices to enable the data acquisition of our system. The devices include an Android smartphone and Autographer on neck. After recording lifelog information, lifeloggers need to import the data in these 2 devices into specific file path. After that, a web-based lifelog viewer will be generated. Lifeloggers can use conversation cue and special movement cue to recalling their episodic memory efficiently by using the lifelog viewer.

3.6.1 Recording Scenario

Lifeloggers using wearable devices to record daily logs. The log include image, time, location record by autographer and daily conversation record by android smartphone. User can upload the data to the webside for the system to process.

3.6.2 Viewing Scenario

After data record and process, the logs will show in the lifelog viewer for lifeloggers to retrieve episodic memory. The catelogs devide into conversation, speaker and sentiment. For conversation cue, lifeloggers can view the logs with the keyword extracted from the conversation when the image was taken. For speaker cue, lifeloggers can view all the speakers who have had a conversation before. For sentiment cue, lifeloggers can view the logs by three kinds of sentiment, happiness, calm and sad. Also, we present a search function for user to search specific conversation, time, sentiment, speaker, location and object to help lifeloggers retrieve episodic memory efficiently.

Chapter 4

System Implementation

4.1 System Hardware

4.1.1 Hardware Overview

4.1.2 Autographer

With the development of lifelog, there are more and more wearable devices that are developed for lifelogging. Wearable cameras are one of them. These cameras can passively and constantly capturing lifelog images. Autographer is a hands-free wearable camera that is used in current research. It is 90x36mm in size and weighs approximately 58 grams which makes it easy to clip on any clothes or hang on a neck. The camera is 5 megapixels and has a 136-degree wide-angle lens which imitates human eyes. Without external interference, the capturing frequency of Autographer is 30 seconds per image. However, unlike other wearable cameras, the capturing frequency can change due to changes in external environment. Autographer is embedded with some sensors, including color sensor, magnetometer, PIR, temperature GPS, etc. When there is a sudden change of sensor value, Autographer will take an extra image which make lifelog images more valuable. Therefore, Autographer is favored by lifelog researchers for its superior performance and relatively intelligent shooting.

Fig. 4.1 Hardware Overview

4.1.3 Android Smartphone

Smartphone is a mobile phone that has a mobile operating system and can expand functions by installing applications, games and other programs. Android is a free and open source based operating system based on Linux, mainly used in mobile devices like smartphones. It is led and developed by Goggle and the open mobile alliance. As one of the most popular mobile operating systems, Android smartphone have covered more than half in the market according to the survey by Kantar. Nowadays, with the development of mobile technique, smartphones are always embedded with variety of sensors like accelerometer, gyroscope, gravity sensor, magnetic field sensor, etc. These sensors can be read by developer using Android SDK and is useful in detecting smartphone holder's movement condition.

4.2 Software Implementation

4.2.1 Development environment

Hardware Environment

Personal computer from IPLab: Lenovo Thinkpad x260

Software Environment

In order to develop our system, we mainly use Java/HTML/CSS/JavaScript language and IDE MyEclipse. As we involve Android devices in our system, we also use Android SDK. For the web server part, we use Apache Tomcat 7.0 x. Meanwhile, we use MySQL for database. The running environment of our web viewer is Chrome and Firefox.

4.2.2 Structure and Framework

Browser/Server (B/S) Structure

Along with the rise of Internet technology, B/S structure is a change and improvement structure of Client/Server (C/S). Under this structure, user interface is implemented through a WWW browser. Very little transaction logic is implemented in the front end (browser), and the main transaction logic is implemented on the server side (server). The advantages of B/S structure are that it greatly simplifies the load scale of client computer, reduces the cost of system maintenance and upgrades and reduces the total cost of user. The B/S structure of our system is shown in Fig 4.5. The structure is consisting of three layers: presentation layer, application layer and data layer. There will be request and reply transaction between each layer. User's behavior is mainly happening at presentation layer. User can use our system at any platform with a web browser.

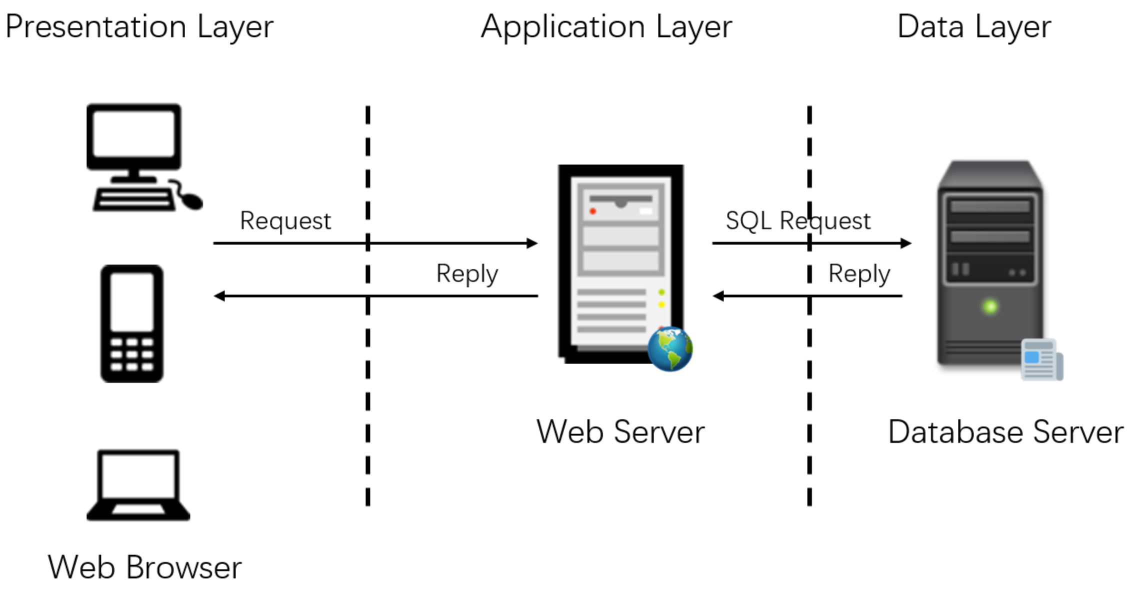


Fig. 4.2 B/S Structure

Struts+Spring+Hibernate (SSH) Framework

SSH is an integrated framework of Struts+Spring+Hibernate. It is a popular open source framework for web application. SSH framework is divided into four layers to help developers build web application with clear structure: presentation layer, business logic layer, data persistence layer and domain module layer. Therefore, SSH framework have the advantage of good reusability and convenient maintenance. In SSH framework, Struts is the overall infrastructure of the system and control business jump. Hibernate framework is used to support the persistent layer. Spring is in charge of managing the Struts and Hibernate. The overall framework is shown in Fig 4.6. Persistence layer uses hibernate framework and SQL to control operation to database. Business layer use Spring framework to control business code. DAO classes are used to communicate between business layer and persistence layer. Struts framework is used in Presentation Layer. This layer involve some .jsp files which show user interface and some .xml files in charge of jumping between each .jsp files.

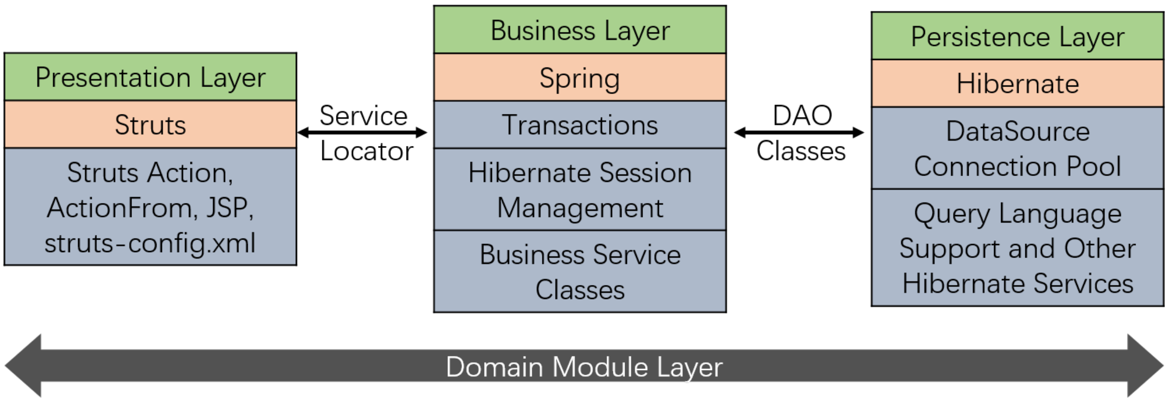


Fig. 4.3 SSH framework

4.3 API Introduction

4.3.1 Microsoft Azure Text Analytics-Keyword

Introduction:

Text Analytics of Microsoft Azure can detect sentiment, key phrases, and language from your text. The API returns a list of strings denoting the key talking points in the input text. They employ techniques from Microsoft Office's sophisticated Natural Language Processing toolkit. English, German, Spanish, and Japanese text are supported.

Request URL:

https://westcentralus.api.cognitive.microsoft.com/text/analytics/v2.1/keyPhrases[?showStats]

Request parameters:

Table 4.1 Request Parameters of Keyword Analytics

Field	Type	Description
showStats	boolean	if set to true, response will contain input and document level statistics
Content-Type	string	Media type of the body sent to the API
OAS-Key	string	Subscription key which provides access to this API

Core Code:

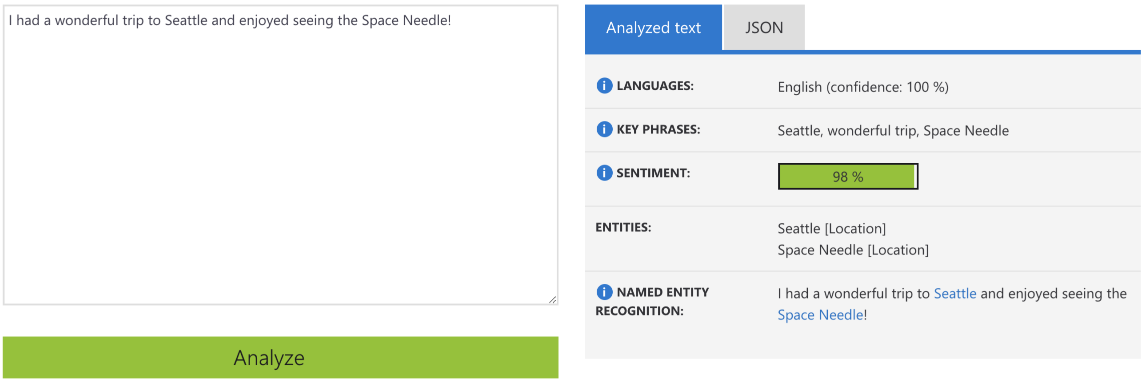


Fig. 4.4 Example

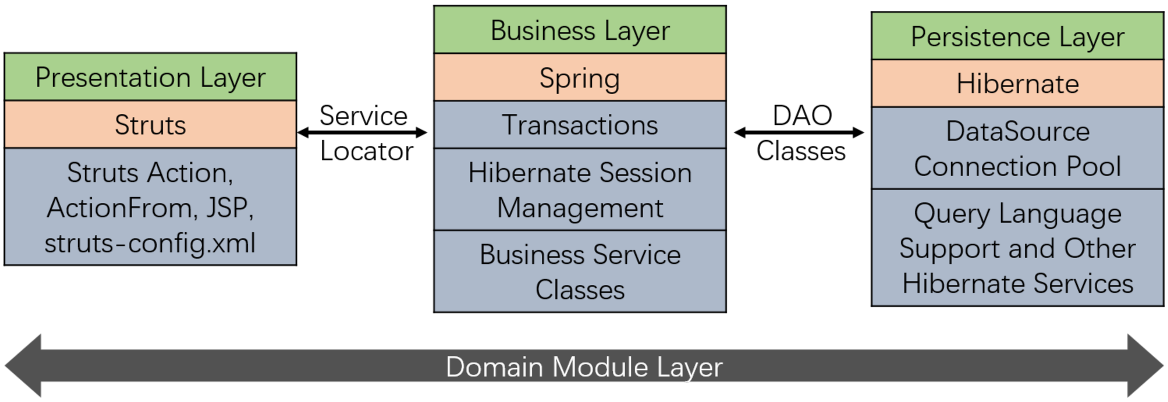


Fig. 4.5 SSH framework

4.3.2 Microsoft Azure Speaker Identification

Introduction:

Microsoft Azure Speaker Identification Identify individual speakers or use speech as a means of authentication with Speaker Recognition. It can identify who is speaking. The API can be used to determine the identity of an unknown speaker. Input audio of the unknown speaker is paired against a group of selected speakers, and in the case there is a match found, the speaker's identity is returned.

Request URL:

https://westus.api.cognitive.microsoft.com/spid/v1.0/identify?identificationProfileIds= identificationProfileIds[shortAudio]

Request parameters:

Table 4.2 Request Parameters of Speaker Identification

Field	Type	Description
identificationProfileIds	string	Comma-delimited identificationProfileIds
shortAudio	boolean	Instruct the service for identification
Content-Type	string	Media type of the body sent to the API
OAS-Key	string	Subscription key which provides access to this API

Request body:

You should include the enrollment audio file in the request body. The minimum recommended amount of accumulated speech for identification, after removing silence, is 30 seconds. In the case you wish to start identification using any amount of speech time, you should include the "shortAudio" parameter. It instructs the service to waive the recommended 30 seconds of audio needed to do identification. When doing so you can send audio files starting 1-second-long but no longer than 5 minutes.

The audio file format must meet the following requirements.

Table 4.3 Requirement of audio file format

Container	WAV
Encoding	PCM
Rate	16KI
Sample Format	16 bit
Channels	Mono

Core Code:

4.3.3 Microsoft Azure Text Analytics-Sentiment

Introduction:

The API returns a numeric score between 0 and 1. Scores close to 1 indicate positive sentiment, and scores close to 0 indicate negative sentiment. Sentiment score is generated

```
<script type="text/javascript">
    $(function() {
        var params = {
            // Request parameters
            "shortAudio": "{boolean}",
        };

        $.ajax({
            url: "https://westus.api.cognitive.microsoft.com/spid/v1.0/identify?identificatio
nProfileIds={identificationProfileIds}&" + $.param(params),
            beforeSend: function(xhrObj){
                // Request headers
                xhrObj.setRequestHeader("Content-Type","application/octet-stream");
                xhrObj.setRequestHeader("Ocp-Apim-Subscription-Key","{subscription key}");
            },
            type: "POST",
            // Request body
            data: "{body}",
        })
        .done(function(data) {
            alert("success");
        })
        .fail(function() {
            alert("error");
        });
    });
</script>
```

Fig. 4.6 Core Code of Speaker Identification

using classification techniques. The input features of the classifier include n-grams, features generated from part-of-speech tags, and word embeddings. It is supported in a variety of languages.

Request URL:

https://westcentralus.api.cognitive.microsoft.com/text/analytics/v2.1/sentiment[?showStats]

Request parameters:

Table 4.4 Request Parameters of Sentiment Analytics

Field	Type	Description
showStats	boolean	if set to true, response will contain input and document level statistics
Content-Type	string	Media type of the body sent to the API
OAS-Key	string	Subscription key which provides access to this API

CoreCode:

```
<script type="text/javascript">
    $(function() {
        var params = {
            // Request parameters
            "showStats": "{boolean}",
        };

        $.ajax({
            url: "https://westcentralus.api.cognitive.microsoft.com/text/analytics/v2.1/senti
ment?" + $.param(params),
            beforeSend: function(xhrObj){
                // Request headers
                xhrObj.setRequestHeader("Content-Type","application/json");
                xhrObj.setRequestHeader("Ocp-Apim-Subscription-Key","{subscription key}");
            },
            type: "POST",
            // Request body
            data: "{body}",
        })
        .done(function(data) {
            alert("success");
        })
        .fail(function() {
            alert("error");
        });
    });
</script>
```

Fig. 4.7 Core Code of Sentiment Text Analytics

4.4 Conversation Cue

To get keyword from conversation, we firstly record daily conversation by the application Voice Recorder in android smartphone, using the application could convert the voice into text in the meantime when user is speaking. After recording, we can get both audio and text file. The text file is used to extract keyword by Microsoft Azure. Text Analytics of Microsoft Azure can detect sentiment, key phrases, and language from your text. The API returns a list of strings denoting the key talking points in the input text. They employ techniques from Microsoft Office's sophisticated Natural Language Processing toolkit. English, German, Spanish, and Japanese text are supported. The audio file could used to play when user click the keyword extract from the conversation in web viewer. By recording and processing, we can get the keywords of one conversation and the voice of the conversation, which can be used for user to retrieve their episodic memory.

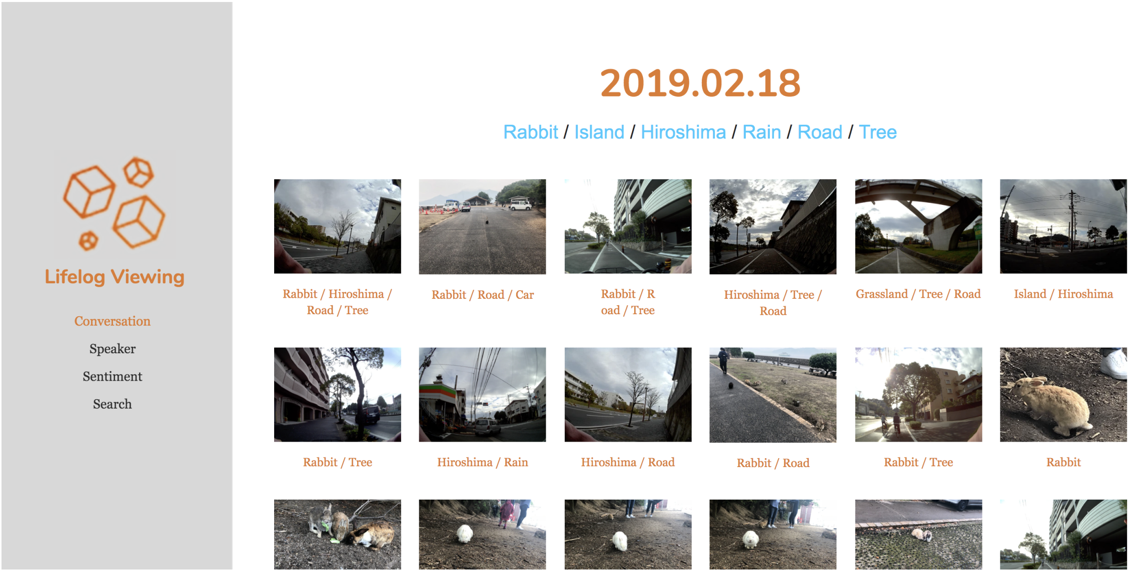


Fig. 4.8 Conversation Cue

Fig4.8 shows the system overview of conversation cue. The web page contains date, keword of conversation, image and keyword of each image. By looking through conversation cue, user can retrieve episodic memory when there was conversation happens. In the example, it shows in the data 2019/02/18. For the whole day, the most exist keyword is Rabbit/Island/Hiroshima/Rain/Road/Tree. The image is filtered when there is conversation happened. Also, the word under the image shows the keyword of conversation when the image was taken.

Fig4.9 shows the result when click the keyword rabbit, it contain the keyword, rabbit. Also, it shows all images contains the keyword rabbit. The keyword is extract from the conversation when the image was taken.

4.5 Speaker Cue

We can recognize different speakers of the conversation by Microsoft Azure Speaker Identification. Microsoft Azure Speaker Identification Identify individual speakers or use

Fig. 4.9 Keyword Rabbit

Fig. 4.10 Speaker

speech as a means of authentication with Speaker Recognition. It can identify who is speaking. The API can be used to determine the identity of an unknown speaker. Input audio of the unknown speaker is paired against a group of selected speakers, and in the case there is a match found, the speaker's identity is returned. By using Speaker Identification we can get different speaker from the conversation. Also, the system summarize the speakers all conversation for user to browse. In the example as figure4.10 shows, there are three speakers recognized by Microsoft Azure Speaker Identification. The system shows each person's image taken by the autographer and their name under each image. When we cilck each person, it will show the detailed information when user talked with the specific person like figure4.11 shows.

(a) Date1

(b) Date2

Fig. 4.11 Logs of Speaker

4.6 Sentiment Cue

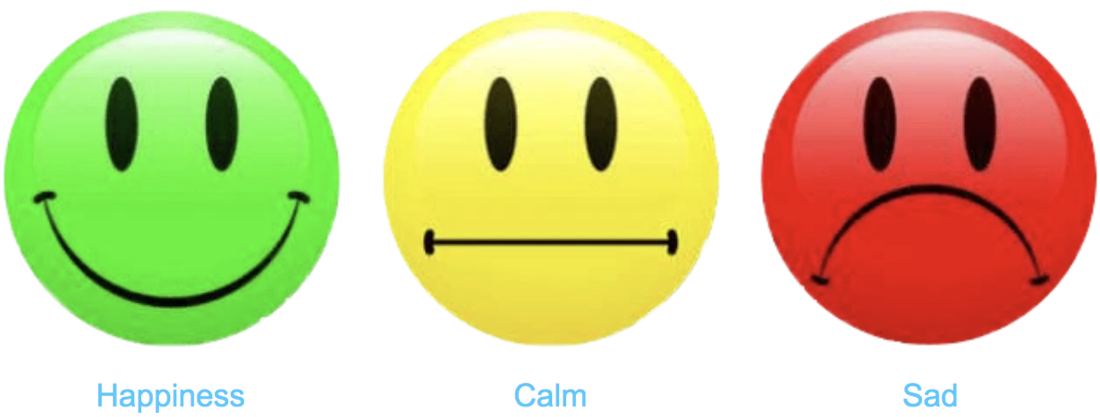


Fig. 4.12 Sentiment

By using Microsoft Azure Text Analyze, we can recognize different sentiment of the conversation. The API returns a numeric score between 0 and 1. Scores close to 1 indicate positive sentiment, and scores close to 0 indicate negative sentiment. Sentiment score is generated using classification techniques. The input features of the classifier include n-grams, features generated from part-of-speech tags, and word embeddings. It is supported in a variety of languages. User can use the web viewer to browse logs of different sentiment. We sort the sentiment into three kinds. Happiness, calm and sad as figure4.12 shows. For the value from 0%-33%, we detect the sentiment as sad. For the value from 34%-67%, we detect the sentiment as calm. For the value from 68%-100%, we detect the sentiment as Happiness.

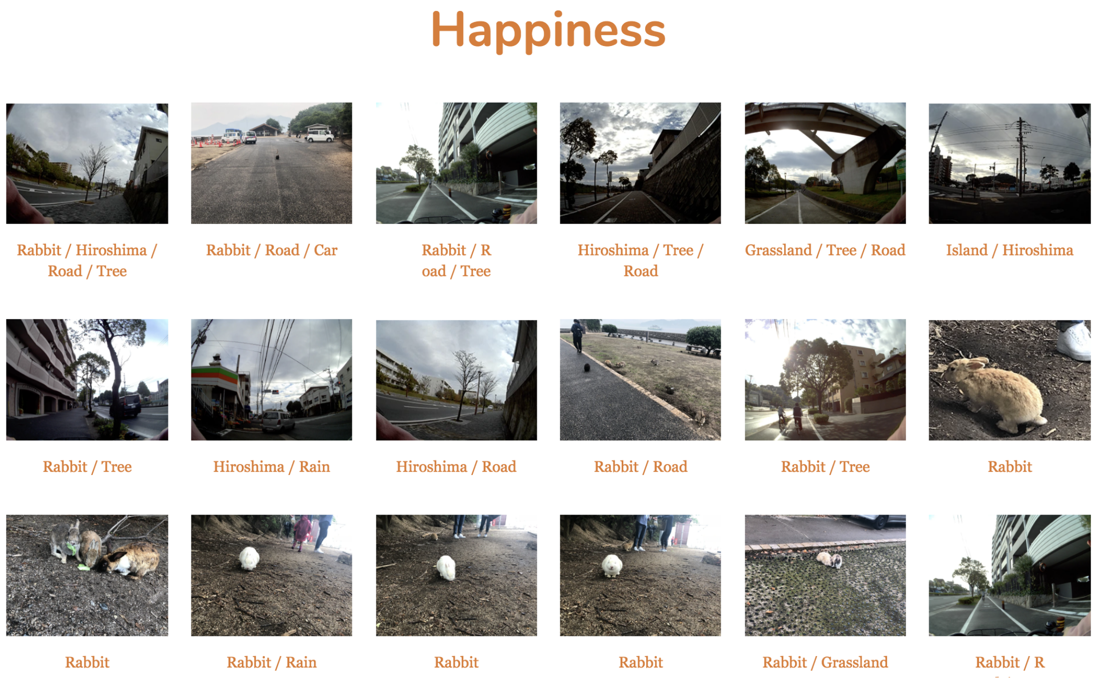


Fig. 4.13 Happiness

Chapter 5

Related Work

In this section we present a brief overview of related work Seyed Ali Bahrainian et al. [17] has comprehensively reviewed research works in the context of lifelogging. Here we briefly describe some of the works that are mostly relevant to augmenting human memory and conversations.

5.1 Augmentation of Human Memory

Sellen et. al.[18], tested the effect of reviewing events captured by SenseCam images on the memory of a number of undergraduate students who presumably had strong memories. The test subjects were deliberately chosen to be undergraduate students in order to prove that “lifelogging technologies can benefit all of us, not just the elderly or memory-impaired”. The findings of this study suggested that, with regard to recollection of past events, automatically captured images may help people remember more than if they had manually taken images on their own. They showed through a case study that images were effective memory cues for recalling memories better after a short time period. Moreover, the study of Lee et al.[19] showed the effect of replaying a person’s significant outing recorded by a SenseCam camera, audio recorder, and GPS logger on elderly people with normal memories. The study showed that the test subjects experienced improvement in recalling the reviewed events. In a subsequent study[11] , the same authors showed that a person with episodic memory impairment, who reviewed memories of a significant event using automatically captured content, could better remember the details of the event after 28 days, as compared with another person with memory impairment who was helped to remember the details of an event by a caregiver who also attended the event. Analogously, the study of Hallberg et. al.[20] reports a memory support tool to increase autonomy for people with mild dementia. The tool supports replaying images, video, audio, etc. to it's user in order to facilitate reminiscence of past memories.

All above-mentioned studies support the claim that a person reviewing her past activities, regardless of whether they had intact memory, mild dementia, or even memory impairment, would benefit and would experience improvements in recalling past memories. However, the problem with the methods presented in these studies is that they mostly present the raw recorded video, audio, or image to their users, and therefore reviewing an audio or video recording of a past memory would require users to spend about the same amount of time as the original experience. On the contrary, in this paper we develop a summarization system that enables users to review their past conversations in a shorter amount of time. Additionally, we extract the most significant topics from transcribed audio recordings to help users better recall past events.

5.2 Related Work on Conversation

One related work is focusing on tracking one's social interactions with the purpose of augmenting his memory.[21] They present Novel Social Interactions Log Analysis System (SILAS) which summarizes one's daily social interactions, connects the conversations with similar topics over time and replays them back. Such system[22] could benefit people with perfectlly working memory as well as people with memory deficiencies by assisting them to recall events or to improve their episodic memories.

In the study[23], they record the conversation of participants' social interactions. Then summary ten words to help participants retrieve their memories. Their research[24] mainly focus on the conversation which includes many contents. For example, taking classes, consulting the doctor or having the meeting.

Despite whether these people are young or old, their memories are intact or not, they are living independently or are assisted by nurses at a nursing home, studies have shown that replaying their lives to them have significant effect in helping them better recall past events of their personal lives[19][25][18]. Therefore, developing technologies that help people to live a more independent and autonomous life with respect to their memory is of crucial importance.

Chapter 6

Evaluation

6.1 Experiments

6.1.1 Participants

As Table 6.1 shows, in order to evaluate usability and efficiency of using our system, we recruited 6 participants, aging from 22 to 26. All participants are students who have general knowledge of computer and have the experience of using web browser. Before our experiment, there is a pre-description of how to use our devices and system.

Table 6.1 Subject Demographic Information

Participants	3 males. 3 females
Age	22-26; Mean: 22.3
Profession	Students

6.1.2 Method

To start the experiment, we give a hypothetical scenario for each participant to use our system:

1. Lifeloggers need to wear wearable devices to enable the data acquisition of our system. The devices include an Android smartphone and Autographer on neck. After

recording lifelog information, lifeloggers need to import the data in these 2 devices into specific file path. After that, a web-based lifelog viewer will be generated. Lifeloggers can use conversation cue and special movement cue to recalling their episodic memory efficiently by using the lifelog viewer.

2. Lifeloggers using wearable devices to record daily logs. The log include image, time, location record by autographer and daily conversation record by android smartphone. User can upload the data to the webside for the system to process.

3. After data record and process, the logs will show in the lifelog viewer for lifeloggers to retrieve episodic memory. The catelogs devide into conversation, speaker and sentiment. For conversation cue, lifeloggers can view the logs with the keyword extracted from the conversation when the image was taken. For speaker cue, lifeloggers can view all the speakers who have had a conversation before. For sentiment cue, lifeloggers can view the logs by three kinds of sentiment, happiness, calm and sad. Also, we present a search function for user to search specific conversation, time, sentiment, speaker, location and object to help lifeloggers retrieve episodic memory efficiently.

The method we used for evaluation is described in the following 3 steps:

1. Each participant is asked to use our wearable devices for successive 5 hours in 1 day. There are no strict regulations on when to start. However, we strongly recommend choosing 10:00am to 3:00pm as recording period.

2. After recording, hardware devices can generate two data: conversation (voice), lifelog data (image, location, time). These data are imported into our proposed lifelog system and a related lifelog viewer is generated.

3. Participant then tried to use our web viewer by their own. Our viewer contains three parts: Conversation keyword, Speaker, Sentiment. Conversation keyword contains the date and several keyword extract from conversation of that date. Speaker cue shows all speakers who have had a conversation before and click each speaker we can view the lifelog image when they had a conversation. Sentiment cue shows the different sentiment extract form the conversation. Participants are required to record how many events they can reminisce by using each viewer (shown in Table 6.2).

Table 6.2 Questionnaire for Efficiency

Question	Answer
How many events you can remember before using our system?	
How many events you can recall by viewing lifelog images?	
How many events you can recall by using our system?	

4. After using our system, a personal interview is given to each participant. The interview is mainly based on following questionnaire. The answer is grading from 1 to 5 (1 = very negative, 5 = very positive).

Table 6.3 Questionnaire for Usability

Question	1	2	3	4	5
Do you think it is comfortable to wear our devices?					
Do you think our viewer is easy to use?					
Do you feel extracted images useful?					
Do you think our system help in aiding memory intuitively?					

6.1.3 Questionnaire

6.2 Evaluation

In our evaluation, all participants have successfully completed the plan and have given effective feedback. In order to get the result of our evaluation, we collected 6 questionnaires from 6 participants (3 males and 3 females) and analyzed their feedback. Since our evaluation is divided into two aspects: efficiency and usability, our results analysis will also be carried out separately in these two aspects. Q1-Q3 reflects the efficiency of our proposed system. To get the result, we calculate the average amount of each question from 6 participants. As shown in Fig 6.2, the result of Q1, Q2 and Q3 is 6.3, 12.5 and 18.5. From the result, we can see that compare to recalling memory in default manner (not using any assistant), using any method in our proposed system can increase the number of recalling events. With the image data in its interface, we can find the result of using conversation cue shows better

Questionnaire

Ladies/gentleman,

Thanks for helping us to complete this questionnaire. All the information will be only used in this research. Your information will be greatly valued and of course will be strictly confidential. Thanks again for your cooperation!

About you

Name ____________ Age ______ Male / Female

Record (Record how many events you can reminisce by using each viewer.)

Question	Answer
How many events you can remember before using our system?	
How many events you can recall by viewing lifelog images?	
How many events you can recall by using our system?	

Questions (1 = very negative, 5 = very positive)

Question	1	2	3	4	5
Do you think it is comfortable to wear our devices?					
Do you think our viewer is easy to use?					
Do you feel extracted images useful?					
Do you think our system help in aiding memory intuitively?					

Fig. 6.1 Questionnaire

performance. It is surprise to find that conversation cue have a positive feedback. In fact, during the interview after evaluation, participants mentioned that when using search function to recall events, at least one trigger need to be memorized. Therefore, it is not so easy to recall memory by using search function.

Another important aspect of our evaluation is the usability of proposed wearable device system and web-based viewing system. The answer of Q4-Q7 can show the result. Q4-Q7 are based on Likert Scale, which describe the answer in five-level: 1. Strongly disagree 2. Disagree 3. Neither agree nor disagree 4. Agree 5. Strongly agree. Therefore, the average score of each question can describe the agreement of each question. And the higher the score is, more positive the result will be. The average score result of Q4-Q7 is shown in Fig 6.3. We can see that the score of all question is above average level, which shows positive result in our system's usability. Moreover, we can see that the usability of hardware system gets lowest score. This reminds us we need to improve our hardware devices to make it easier to be carried out.

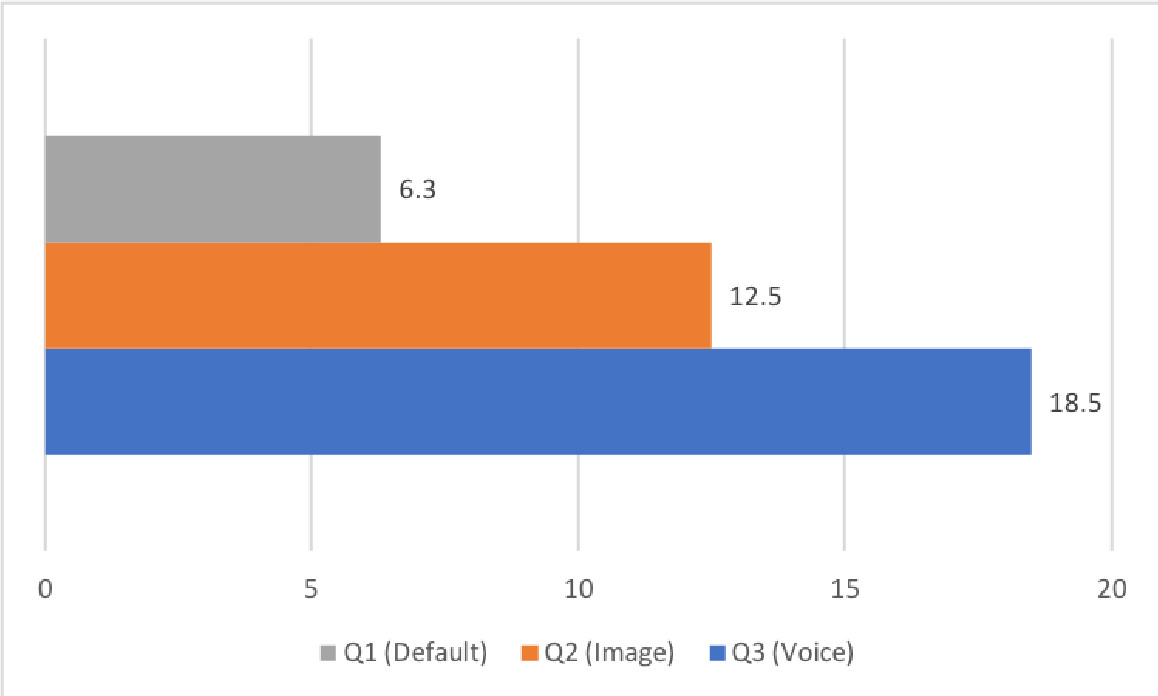


Fig. 6.2 Analyze Result of Q1-Q3

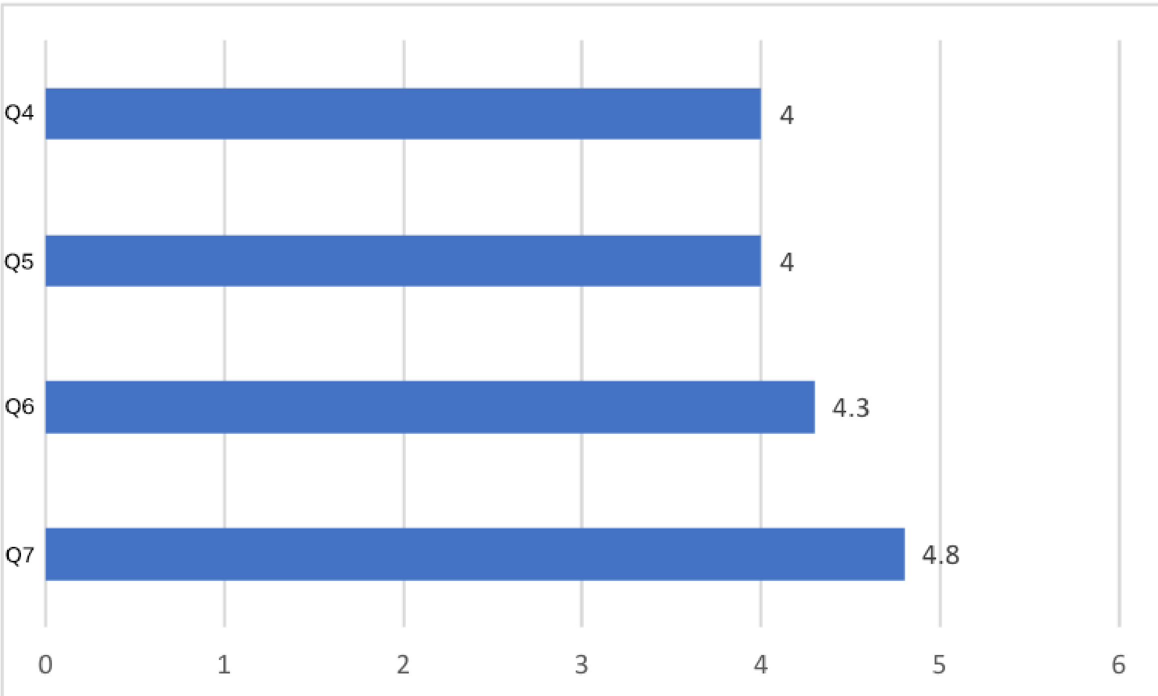


Fig. 6.3 Analyze Result of Q4-Q7

Chapter 7

Conclusion

7.1 Summary

In general, our research explains the limitation of memory cue in lifelog system for aiding episodic memory. In order to improve the situation and enhance the efficiency of aiding episodic memory, we propose three user-related cues in this research and has implemented them into a web-based lifelog viewer.

In the proposed system, we mainly implement conversation cue, speaker cue and sentiment cue as two new user-related cues for lifeloggers to retrieve memory. We use two Autographers and an Android smartphone to capture conversation data and lifelog images. After processing the obtained data, we generate a web viewer for lifeloggers to use conversation cue, speaker cue and sentiment cue to view lifelog images and retrieve memory. Moreover, we combine our user-related cues with traditional contextual cues (what, where, who, when) in "search" part of our system.

We assume that lifeloggers can wear our proposed hardware devices in their daily life. After recording for the whole day, lifeloggers can upload their data onto our system and generate their own web viewer. They can retrieve their memory efficiently and conveniently by using our proposed system.

In order to test the usability and efficiency of our proposed system, we have included some participates in our evaluation. The feedback is quite positive.

7.2 Future Work

Although we have proposed a prototype of utilizing three new cues in lifelog system, there are still some limitations and future possibilities to improve the efficiency of aiding episodic memory. Because of the large scale of images and the restriction of APIs, processing speed in our system is a limitation. We hope this can be solved by optimizing image processing algorithm in the near future. Also, we are thinking of the possibility of involving more user-related cues. This can be done by monitoring other physiological signals like skin conductance.

References

[1] Cathal Gurrin, Alan F. Smeaton, and Aiden R. Doherty. *Lifelogging: Personal Big Data*. Now Publishers Inc., Hanover, MA, USA, 2014.

[2] Cathal Gurrin, Alan F. Smeaton, and Aiden R. Doherty. Lifelogging: Personal big data. *Foundations and Trends in Information Retrieval*, 8(1):1–125, 2014.

[3] Steve Mann. Wearable computing: A first step toward personal imaging. *Computer*, 30(2):25–32, 1997.

[4] Autographer, March 2017.

[5] Lauralee Sherwood. Human physiology: From cells to systems. *Cengage Learning*, pages 157–162, 2015.

[6] Lu Wang and Claire Cardie. Summarizing decisions in spoken meetings. In *Proceedings of the Workshop on Automatic Summarization for Different Genres, Media, and Languages*, WASDGML '11, pages 16–24, Stroudsburg, PA, USA, 2011. Association for Computational Linguistics.

[7] Larry R. Squire and Stuart M. Zola. Episodic memory, semantic memory, and amnesia. *Psychology*, 8(3):205–211, 1998.

[8] Endel Tulving. Episodic memory: from mind to brain. *Annual Review of Psychology*, 53(1):1–25, 2002.

[9] Stanley B. Klein, Judith Loftus, and John F. Kihlstrom. Memory and temporal experience: the effects of episodic memory loss on an amnesic patient's ability to remember the past and imagine the future. *Social Cognition*, 20(5):353–379, 2002.

[10] Douglas W. Oard and Joseph Malionek. The apollo archive explorer. In *Proceedings of the 13th ACM/IEEE-CS Joint Conference on Digital Libraries*, JCDL '13, pages 453–454, New York, NY, USA, 2013. ACM.

[11] Matthew L. Lee and Anind K. Dey. Lifelogging memory appliance for people with episodic memory impairment. *Proceedings of the 10th International Conference on Ubiquitous Computing*, (10):44–53, 2008.

[12] Foteini Agrafioti, Dimitris Hatzinakos, and Adam K Anderson. Ecg pattern analysis for emotion detection. *IEEE Transactions on Affective Computing*, 3(1):102–115, 2012.

[13] Matthew L. Lee and Anind K. Dey. Providing good memory cues for people with episodic memory impairment. In *Proceedings of the 9th international ACM SIGACCESS conference on Computers and accessibility*, pages 131–138. ACM, 2007.

[14] Matthew L. Lee and Anind K. Dey. Using lifelogging to support recollection for people with episodic memory impairment and their caregivers. In *Proceedings of the 2nd International Workshop on Systems and Networking Support for Health Care and Assisted Living Environments*, page 14. ACM, 2008.

[15] Soumyadeb Chowdhury, Md Sadek Ferdous, and Joemon M Jose. A user-study examining visualization of lifelogs. In *Content-Based Multimedia Indexing (CBMI), 2016 14th International Workshop on*, pages 1–6. IEEE, 2016.

[16] Abigail J. Sellen and Steve Whittaker. Beyond total capture: a constructive critique of lifelogging. *Communications of the ACM*, 53(5):70–77, 2010.

[17] David M. Blei and John D. Lafferty. Dynamic topic models. In *Proceedings of the 23rd International Conference on Machine Learning*, number 8, pages 113–120, New York, NY, USA, 2006. ACM.

[18] Abigail J. 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 *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, number 10, pages 81–90, New York, NY, USA, 2007. ACM.

[19] Matthew L. Lee and Anind K. Dey. Wearable experience capture for episodic memory support. In *Proceedings of the 2008 12th IEEE International Symposium on Wearable Computers*, number 2, pages 107–108, Washington, DC, USA, 2008. IEEE Computer Society.

[20] Josef Hallberg, Basel Kikhia, Johan Bengtsson, Stefan Sävenstedt, and Kåre Synnes. Reminiscence processes using life-log entities for persons with mild dementia. In *1st International Workshop on Reminiscence Systems*, number 6, pages 16–21, 2009.

[21] Dong Chen, Xudong Cao, Fang Wen, and Jian Sun. Blessing of dimensionality: High-dimensional feature and its efficient compression for face verification. In *Proceedings of the 2013 IEEE Conference on Computer Vision and Pattern Recognition*, CVPR '13, pages 3025–3032, Washington, DC, USA, 2013. IEEE Computer Society.

[22] Nebojsa Jojic, Alessandro Perina, and Vittorio Murino. Structural epitome: A way to summarize one's visual experience. In *Proceedings of the 23rd International Conference on Neural Information Processing Systems - Volume 1*, NIPS'10, pages 1027–1035, USA, 2010. Curran Associates Inc.

[23] Seyed Ali Bahrainian and Fabio Crestani. Cued retrieval of personal memories of social interactions. In *Proceedings of the First Workshop on Lifelogging Tools and Applications*, LTA '16, pages 3–12, New York, NY, USA, 2016. ACM.

[24] Yaniv Taigman, Ming Yang, Marc'Aurelio Ranzato, and Lior Wolf. Deepface: Closing the gap to human-level performance in face verification. In *Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition*, CVPR '14, pages 1701–1708, Washington, DC, USA, 2014. IEEE Computer Society.

[25] Basel Kikhia, Josef Hallberg, Kåre Synnes, and Zaheer Ul Hussain Sani. Context-aware life-logging for persons with mild dementia. In *In Engineering in Medicine and Biology Society*, number 4, pages 6183–6186. Annual International Conference of the IEEE, 2009.