I-044

Semi-automatic Annotation of Personal Digital Photographs with W6H2 Metadata

Supheakmungkol SARIN[†] Toshinori NAGAHASHI[‡] Tadashi MIYOSAWA[‡] and Wataru KAMEYAMA[†]

[†]GITS, WASEDA University
mungkol@fuji.waseda.jp
wataru@waseda.jp

[‡]Corporate R & D Div., SEIKO EPSON Corporation
nagahashi.toshinori@exc.epson.co.jp
miyosawa.tadashi@exc.epson.co.jp

1 Introduction

As the number of personal digital photographs continues to grow in an exponential rate, there is a real need in developing a system which is capable of managing such tremendous amount of photographs for users. The key solution to this is obviously to provide a better annotation to each of the photographs. However, until now, photo annotation still suffers from the trade-off between time-consuming and semantic-gap issues.

How are we going to deal with this concern?

Personal digital photographs have very different characteristics as compared with other types of image such as those from the museum or web image collections. Usually, users' personal digital photos reflect their daily activities and what happens in their ambient environment. This doubtless implies that all the information from their daily life would be the ideal resources that can be used to extract semantic information to describe their photos taken on the same day or within an interval of time. For instance, *personal information* such as emails, schedules, chats, document files, etc. might contain more or less some direct interpretation of users' point of view about the photos they have taken, while *public information* such as articles from wikipedia, tourism websites, online news about the location of the photos or events that happened on the day or within an interval of time during which the photo was captured contain more or less the facts or related information about the photos. Later when trying to look for their photos, users will be very likely to use the same keywords that they once used in their personal documents or what they have seen or heard of from their experiences. In addition, recently, all digital cameras provide time information and most camera-phones can infer rough location from GPS or Cell ID information. It is promising that all cameras will eventually be equipped with location information capturing system.

Based on the above facts and hypothesis, we propose a system using the exact time and location where the photos were taken to match against users' readily-available contextual resources from both their *personal* and *public information*. With those related contents, we extract by using a number of natural language techniques, some potential keywords of different categories to describe the photo semantically.

2 System Design & Implementation

2.1 System Architectures

We have produced the prototype of the system. Since information extraction (IE) is a time-consuming task, we want to perform it in advance in order to increase the interactiveness during the run-time photo annotation task. Thus, we separate the system into two phases: *pre-processing* and *run-time*.

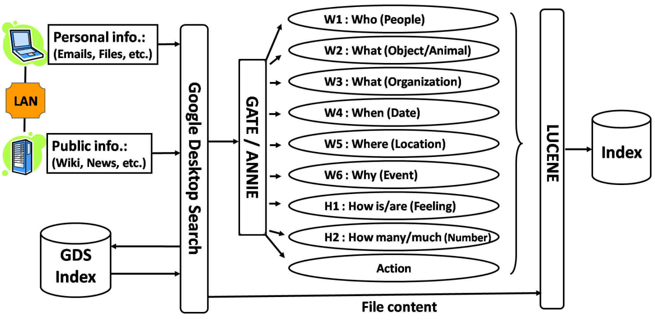


Fig. 1: Pre-processing Architecture

Pre-processing phase is to acquire all kinds of resources concerning users' daily life including what happens in their ambient environment; integrate them and then perform the IE tasks to extract metadata. We download an archive of Wikipedia and online news from different news websites and store them locally. We use Google Desktop Search [2] and configure it as a tool to integrate and index the downloaded documents as well as all kinds of file from users' computers. The latter also includes users' emails, chats and daily web browsing histories. The integrated resouces are then passed to GATE/ANNIE [1] module which has the function of information extraction. 9 categories of metadata keywords from each source are extracted and sent to Lucene [4] module along with the original file content for indexing. Those categories of metadata are people's names (Who), names of animate and inanimate objects (What Object/Animal), names of organization (What Organization), date and time (When), names of locations (Where), events (Why), feeling or emotion (How is/are), number (How much/many). We group these categories together and name it as *W6H2 metadata*. Beside these metadata, action keywords are also extracted. Figure 1 shows the steps in the pre-processing phase.

During the run-time annotation, *date* and *location* information about the photo are used to search the established Lucene index. We merge all the keywords of related sources and sort them by category. From the

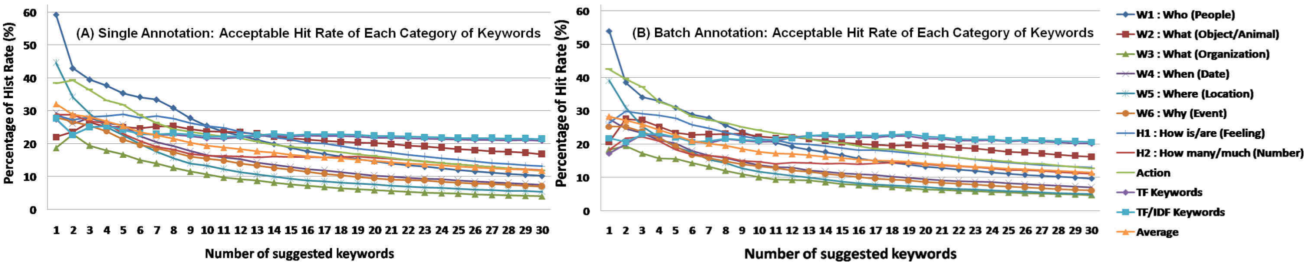


Fig. 2: Accept. Hit Rate of keywords of each category: (A) Single Annotation & (B) Batch Annotation

index, we can establish two other categories of keyword which are TF (keywords by term frequency) and TF/IDF (keywords by term frequency and the inverse document frequency). Then, we rank keywords of each category according to their frequency of appearances from all the related documents. The top keywords of each categories are presented to users via our annotation interface for their validation before being sent to the XML metadata database. Figure 3 depicts our steps in the run-time phase.

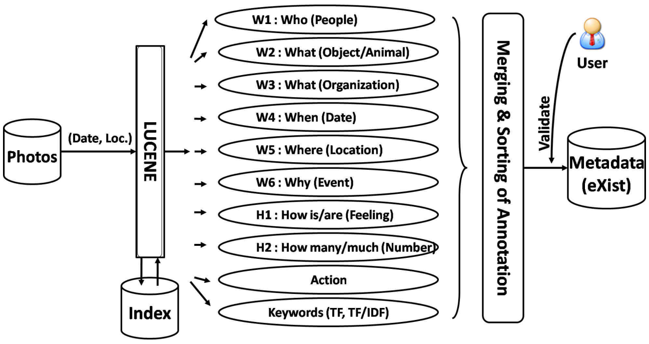
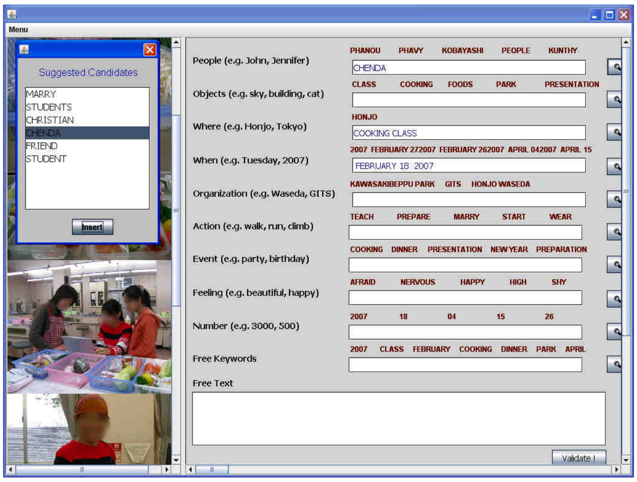


Fig. 3: Run-Time Annotation Architecture

We divide the annotation task into two kinds: *single annotation* in which photo is annotated one by one and *batch annotation* in which photos from the same event are annotated all at once. Figure 4 shows our batch annotation interface with keywords suggestion feature.



MARRY
STUDENTS
CHRISTIAN
CHENDA
FRIEND
STUDENT

People (e.g. John, Jennifer)
PHANOU PHAVY KOBAYASHI PEOPLE KUNTHY
CHENDA

Objects (e.g. sky, building, cat)
CLASS COOKING FOODS PARK PRESENTATION

Where (e.g. Honjo, Tokyo)
HONJO
COOKING CLASS

When (e.g. Tuesday, 2007)
2007 FEBRUARY 272007 FEBRUARY 262007 APRIL 042007 APRIL 15
FEBRUARY 18 2007

Organization (e.g. Waseda, GITS)
KAWASAKIBEPPU PARK GITS HONJO WASEDA

Action (e.g. walk, run, climb)
TEACH PREPARE MARRY START WEAR

Event (e.g. party, birthday)
COOKING DINNER PRESENTATION NEW YEAR PREPARATION

Feeling (e.g. beautiful, happy)
AFRAID NERVOUS HAPPY HIGH SHY

Number (e.g. 3000, 500)
2007 18 04 15 26

Free Keywords
2007 CLASS FEBRUARY COOKING DINNER PARK APRIL

Free Text

Fig. 4: Batch Annotation with Keyword Suggestion

3 Evaluations and Results

Experiments of our system were conducted. 10 subjects were recruited for the experiments. Each subject contributes about 50 photos from at least 5 different events. The events are from a period of 3 months.

During the experiments, we generate 30 proposed keyword candidates of each category and ask the subjects to evaluate each of them whether it is related to the photo and can be acceptable as a keyword for the photo. We arrive at the results of the acceptable hit rate of each annotation category, namely single annotation and batch annotation which are illustrated in Figure 2 part (A) and (B), respectively. Results show that in both categories of annotation, we could obtain the average accuracy of around 22%-32% when introducing the first 5 keywords. This means we could assure at least 1 acceptable keyword of each category if we suggest 5 keywords. We also observe from our experimental data that the results are obtained with the average number of sources used around 40 and each one with the average size of 70KBytes. This relationship is quite interesting because we can somehow predict the number and size of documents needed to assure this degree of hit rate. However, extended experiments and improvements are still needed to increase the current accuracy and generalize the results for real application implementation.

4 Related Work

Pei-Jeng Kuo et al. proposed MPEG-7 based ontology for personal digital photographs [3]. Some recent works of Naaman et al. [5] enable digital photos to be annotated with metadata of spatial context using exact location information. In the same context, Tuffield et al. suggest a way to get metadata from other sources in addition to the photo content[6].

5 Conclusions

In this paper, a novel approach in generating contextual metadata for photos is presented. We are able to produce 11 categories of keywords with an encouraging acceptable hit rate. We believe that with these 11 categories of metadata, we could cover most of the related information about the photos to semantically interpret them. In short, with this proposed approach we achieve 3 main goals: (1) semi-automize the annotation task, (2) reduce semantic gap and (3) provide a practical implementation framework without any training task.

References

[1] GATE/ANNIE. http://gate.ac.uk.

[2] Google Deskop Search. http://desktop.google.com/.

[3] P.-J. Kuo, T. Aoki, and H. Yasuda. Building personal digital photograph libraries: An approach with ontology-based mpeg-7 dozen dimensional digital content architecture. In *CGI '04: Proceedings of the Computer Graphics International*, pages 482–489, Washington, DC, USA, 2004.

[4] Apache Lucene. http://lucene.apache.org.

[5] M. Naaman, A. Paepcke, and H. Garcia-Molina. From where to what: Metadata sharing for digital photographs with geographic coordinates. In *10th CoopIS*, 2003.

[6] M. Tuffield, S. Harris, D. P. Dupplaw, A. Chakravarthy, C. Brewster, N. Gibbins, K. O'Hara, F. Ciravegna, D. Sleeman, Y. Wilks, and N. R. Shadbolt. Image annotation with photocopain. In *Proceedings of the First International Workshop on Semantic Web Annotations for Multimedia (SWAMM)*, May, 2006.