1D-7

Maximizing Context with Desktop Search Integration for Semi-Automatic Metadata Generation of Digital Photographs

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

In our previous work, we presented SemiANNOTATE - a novel semi-automatic personal digital photographs annotation system which takes the advantages of public and personal information to leverage contextual metadata for consumers' photographs [9]. This paper describes some modifications of our previous research to maximize consumers' ambient contextual information with practical implementation. In the current work, we integrate Google Desktop Search (GDS) into our system with some adaptive components. Experiments of our new system were conducted with 5 subjects and 172 photos. The results are presented.

2 System Development

Figure 1 shows the steps in generating automatic metadata. We group them into 3 parts namely, *Source Selection*, *Information Extraction* and *Automatic Metadata Suggestion*.

2.1 Sources Selection

In this step, we try to get relevant sources from all readily-available sources by matching time and location information of each photo against them. We assume that we can get location information from the photo.

2.1.1 Sources

One of the main targets of our proposed model is to gather free or cheap readily-available sources of information from users and their surrounding environment. This is because we would like to keep minimal effort from user yet get relevant sources and make it easy to implement. In this regard, we identify our sources as follow and we divide them into two categories namely *public* and *personal information*.

1. **Public information** refers to information from public repositories such as community news, worldwide news, online encyclopedia, tourist sites or other main public information websites. In our case, we use MDN-Mainichi English news [8], Asahi English News [1] and Wikipedia [10].
2. **Personal information** refers to sources of information from user's private information such as their schedules, notes, emails, chats, web browsing histories and all other documents in their computer.

2.1.2 Acquisition & Selection Mechanism

We integrate GDS [4] to our system via its JAVA API [5]. GDS is configured to index all the files from user's hard disk. For public information, the above mentioned public repositories are crawled and stored in user's computer so that Google Desktop can index all of them together with personal information. HTTrack [6] is used for this purpose. We perform both exact query matching as well as loose query matching. We select only the first top 100 relevant sources to keep high relevant quality of sources as well as to reduce computing time in the *Information Extraction* step.

2.2 Information Extraction

In this part, we try to extract some named entities (NE) such as names of people, organization and some other important keywords automatically from relevant sources obtained from the previous step. To do this, we have built two modules : (A) *Named Entity Extraction module* and, (B) *Keywords Extraction module*.

2.2.1 Named Entity Extraction

We apply an information extraction (IE) engine to extract NE from the sources. A Nearly-New Information Extraction system (ANNIE) [3] has been used as our IE engine. ANNIE is composed of the following sub-modules : Unicode Tokenizer, Gazetteer Lookup, Sentence Splitter, Semantic Tagger and Name Matcher. We have modified the gazetteer list to include more resources to suite our case. We have also built our own sub-module *Named Entity Sorter* to rank the obtained NE in order to suggest better results. 4 categories of NE have been extracted namely, *people's names*, *organization names*, *date* and *location names*.

2.2.2 Keyword Extraction

To retrieve potential keywords, we integrate Lucene [7] into our module. The latter indexes the relevant sources. Then it calculates the ranking of each term in the sources by their frequencies of occurrence. In our case, we select the top 30 keywords.

2.3 Automatic Keyword Suggestion

All the metadata candidates (of *Who*, *Organization*, *Where*, *When* and *Free Keyword* field) are presented to the user. Top suggested keywords of each field are shown to users. They may consult more keywords by clicking on the *magnifying icon* of each field. To be able to answer all the questions related to the photos and to improve semantic integrity, we include three other fields: *Event*, *How*, and *Free Text*. However they are optional. *Figure 2* depicts our annotation engine interface with keywords suggestion feature.

In order to assure the quality of the keywords for each photo, users have to verify them and make some modifications if necessary. Once user validates the metadata, it will be sent to our XML Native Database - eXist [2]. A subset of MPEG-7 MDS has been selected as our metadata format.

2.4 Experiments and Results

The objective of our experiments is to evaluate the time difference when using our annotation system with built-in keywords suggestion feature and without this feature. In addition, we also would like to evaluate the accuracy of our proposed NE and keywords by calculating their hit rates.

5 subjects were recruited for the experiments. Each subject contributed at least 30 photos for the experiments. We came up with the total number of 172 photos with a 6-month interval of time. Each subject was asked to install GDS and

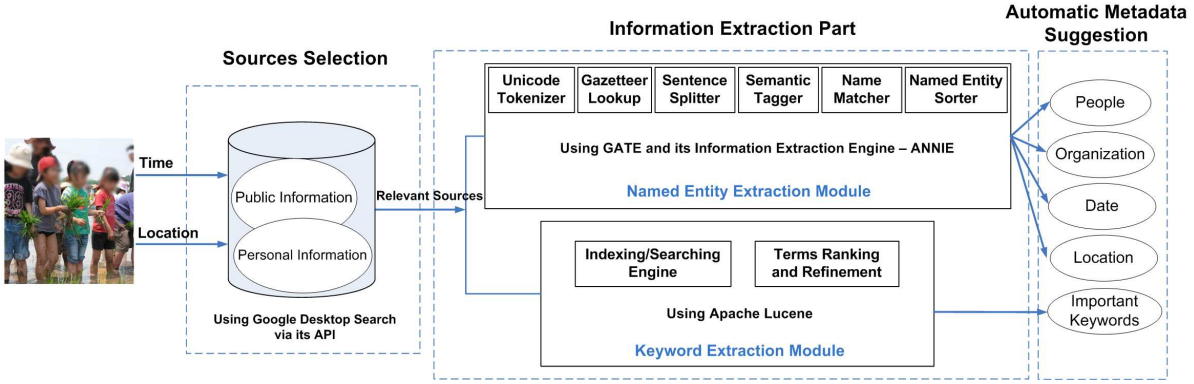

Figure 1: Processing steps in generating automatic metadata from user's context

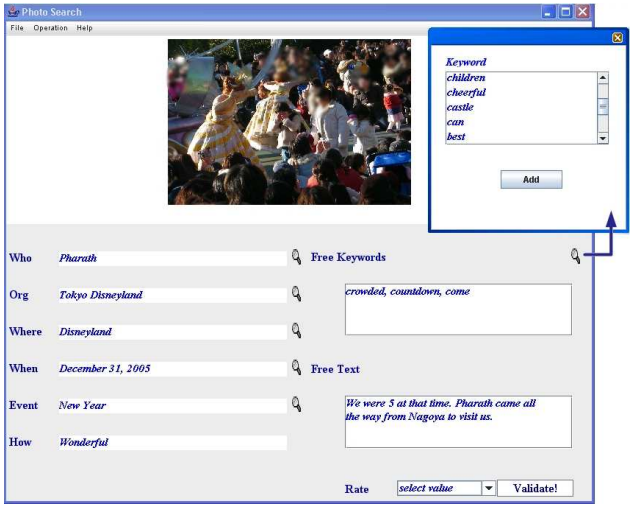

Figure 2: Annotation with Keywords Suggestion

activate it when they use their computer. Subjects also installed our prototype system in their own machine.

In case of annotation without metadata suggestion feature, subjects were asked to annotate their photos using a blank interface. We asked subjects once again to annotate their photos using our system with metadata suggestion feature enabled. In each experiment, time required to annotate each photo was recorded. For the purpose of evaluation, for each photo, we also generate 30 free keywords, 5 person names and 5 organization names to files before the second experiment. We left a period of 2-3 days between the first and second experiment to avoid influence of subject's memory about their input metadata from the first experiment. After the second experiment, we asked subjects to judge the metadata of each photo file that we saved.

We arrive with the results shown in *Figure 3*:

1. **Time performance**: In (A), we can reduce up to 39% of annotation time when enabling the keywords suggestion feature.
2. **Accuracy**: (B) and (C) show that the acceptable hit rate of proposed person's names and organization names both hold up to nearly 30% for the first name suggested, and drop gradually to around 16% and 8% respectively if we suggest the fifth name. However, by suggesting the top 5 names of each category, these results explain that 83% of photos will have at least one acceptably correct person name while 41% of photos will have at least 1 acceptably correct organization name. (D) shows the acceptable hit rate up of keywords up to 60% if we suggest only 7 top keywords and about 55% if 11 keywords are suggested. In other words, if we suggest 7 keywords, then 4 keywords are

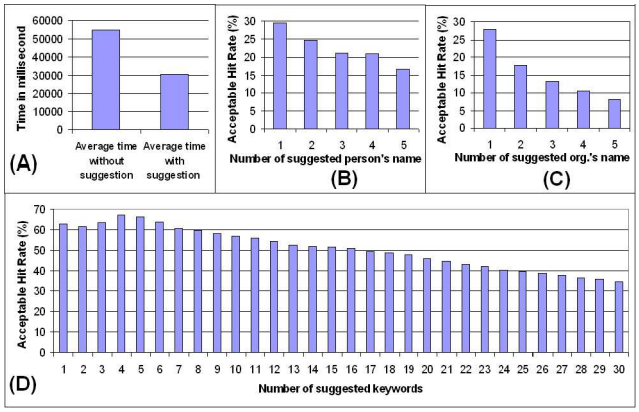

Figure 3: (A) Time difference between *annotation without keyword suggestion* and *annotation with keyword suggestion*, (B) Acceptable Hit Rate of suggested People's name, (C) Acceptable Hit Rate of suggested Organization's name, (D) Acceptable Hit Rate of suggested Keywords

acceptable.

3 Conclusions

We propose a novel and practical paradigm in integrating and generating contextual metadata for photos from readily available public and personal sources. Our experiments give us very encouraging results. We are now doing extensive experiment to assure the effectiveness of our system. In our future work, we would like to focus on improving our IE part to get better metadata as well as integrating the current available CBIR technologies. Thus we will get richer metadata that describes not only the abstract meaning of the photographs but also its content features.

References

[1] Asahi English News. http://www.asahi.com/english/.
[2] eXist XML Database. http://exist.sourceforge.net.
[3] GATE and ANNIE. http://gate.ac.uk.
[4] Google Desktop Search. http://desktop.google.com/.
[5] GDS Java API. http://sourceforge.net/projects/gdapi/.
[6] Httrack. http://www.httrack.com/.
[7] Apache Lucene. http://lucene.apache.org.
[8] Mainichi English News. http://mdn.mainichi-msn.co.jp/.
[9] S. Sarin, T. Nagahashi, T. Miyosawa, and W. Kameyama. Semi-ANNOTATE: A Semi-automatic Approach to Personal Photo Album Annotation based on Public and Personal Information. In *Proceedings of the 5th Forum of Information Technology (FIT2006)*, 2006.
[10] Wikipedia. http://en.wikipedia.org/.