Computers and Electrical Engineering 000 (2016) 1–14

Contents lists available at ScienceDirect

Computers and Electrical Engineering

journal homepage: www.elsevier.com/locate/compeleceng

An intelligent annotation-based image retrieval system based on RDF descriptions

Hua Chen[a,b,*], Antoine Trouve[b], Kazuaki J Murakami[b], Akira Fukuda[a]

[a] Graduate School of Information Science and Electrical Engineering, Kyushu University, 744, Motooka, Nishi-ku, Fukuoka, 819-0395 Fukuoka, Japan
[b] Institute of Systems, Information Technologies and Nanotechnologies (ISIT), 2-1-22-7F, Momochihama, Sawara-ku, Fukuoka, 814-0001 Fukuoka, Japan

ARTICLE INFO

Article history:
Received 3 June 2016
Revised 24 September 2016
Accepted 26 September 2016
Available online xxx

Keywords:
Image retrieval
Image annotation
Semantic image retrieval
RDF

ABSTRACT

In this paper, we aim at improving text-based image search using Semantic Web technologies. We introduce our notions of concept and instance in order to better express the semantics of images, and present an intelligent annotation-based image retrieval system. We test our approach on the Flickr8k dataset. From the provided captions, we generate annotations at three levels (sentence, concept and instance). These annotations are stored as RDF triples and can be queried to find images. The experimental results show that using concepts and instances to annotate images flexibly can improve the intelligence of the image retrieval system: (1) with annotations at concept level, it enables to create semantic links between concepts and then addresses many challenges, such as the problems of synonyms and homonyms; (2) with annotations at instance level, it can count things (e.g., "two people", "three animals") or identify a same concept.

1. Introduction

With the development of information technologies, a huge amount of digital images is being generated very rapidly. Consequently, how to quickly and accurately find relevant images has become a very hot research topic. Image retrieval solutions are generally classified into two types: content-based image retrieval (CBIR) solutions and annotation-based image retrieval (ABIR) solutions. Basically, CBIR solutions use visual features (such as color, texture, shape and object location) to retrieve images based on content properties. This technology has been widely used in many applications such as fingerprint identification, digital libraries, medicine and historical research, among others [1,2]. All these applications are similar: from an input image, they boil down to finding the similar images from a collection of images. However, because the visual features cannot accurately represent the semantics of images, the CBIR solutions suffer from the semantic gap problem [3]. On the other hand, ABIR solutions use textual descriptions as image metadata and find images with text retrieval techniques. The purpose of image annotation is to narrow the semantic gap between image visual features and semantics [4]. The crucial challenge is how to find relevant images accurately and intelligently.

It is well known that Flickr[1] is a famous image sharing website, which gathers millions of photos. It provides many kinds of image retrieval ways, including one with ABIR solution, which allows users to annotate and retrieve images by tags (a

* Corresponding author.
E-mail addresses: hua.chen@soc.ait.kyushu-u.ac.jp, chenhua5752@hotmail.com (H. Chen).

[1] See http://www.flickr.com.

http://dx.doi.org/10.1016/j.compeleceng.2016.09.031
0045-7906/

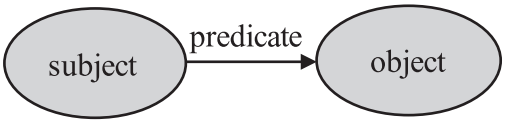

Fig. 1. An RDF triple graph(subject, predicate, object).

```
prefix foaf: <http://xmlns.com/foaf/0.1/>.
select ?age ?name where
{Image URI Creator ?person.
?person foaf:name ?name.
?person foaf:age ?age}
```

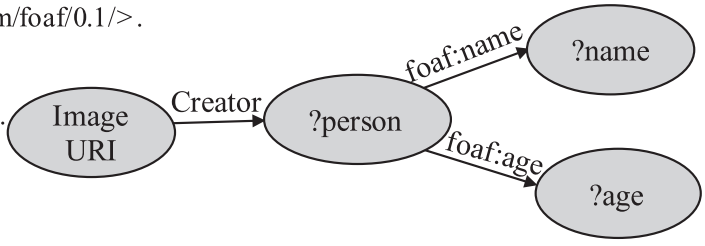

Fig. 2. An example of SPARQL query.

form of metadata). With this feature, users can find images easily and conveniently. However, in some cases, the retrieval results may be inaccurate. For example, when searching with “2 animals” or with “Apple” (a company name), the retrieval results are inaccurate; when searching with “TV” and with “television”, the retrieval results are different. The reasons may be as follows: (1) **synonyms**: “TV” and “television” have the same meaning, but machines cannot understand; (2) **homonyms**: it is unknown whether “Apple” is a kind of fruit or a company name; and (3) **count things**: it is unable to count things (e.g., “2 animals”). The problems of synonyms and homonyms have been mentioned many times by other researchers and have been solved in several ways. But the third problem is rarely discussed.

In this paper, we propose a new intelligent ABIR system. Our experiments are based on the Flickr8K [5] dataset, which consists of images extracted from the Flickr website, together with natural language sentences. We process those sentences and generate annotations for images at three levels of description: (1) sentence level: use natural language sentences, such as “two lizards fighting”; (2) concept level: use abstract or general concepts, such as “animal”, “cat” and “white cat”; and (3) instance level: use concrete and specific words, such as “1 cat”, “2 animals” and “a ‘my best buddy’ shirt”. We express annotations using Resource Description Framework (RDF) [6]. RDF is a foundation for processing metadata, it is designed to describe the resources and the relationships among them. Our main contributions are as follows: (1) we define our notions of concept and instance to describe images; (2) we propose an image annotation model to annotate images at three levels of description; (3) we count things (e.g., “2 animals”) intelligently in images with instances; and (4) we find things accurately by using unique identifiers of instances.

The rest of this paper is organized as follows. Section 2 briefly introduces the dataset and the basic technologies used in our work. Section 3 reviews some related works. In Section 4, we describe our image annotation model, the design of our system and evaluation methodology. Section 5 presents the results and discussion. Section 6 gives the conclusion and future work.

2. Dataset and related technologies

2.1. RDF

RDF is a standard data model to describe Web resources within the Semantic Web. By using RDF, we can define and use metadata vocabularies to make statements about resources; furthermore, we can also create links between different resources. A resource can be anything that is identifiable by a Uniform Resource Identifier (URI), the statements describe the properties of resources, and the links indicate the relationships between resources. RDF uses a graph data model, an RDF graph can be illustrated as a directed labeled graph (as shown in Fig. 1) as well as a triple of the form (subject, predicate, object) (it means that <subject> has a property <predicate>, whose value is <object>). RDF can be serialized in several different formats, one of which is Terse RDF Triple Language (Turtle) [7]. Turtle can display triples in a concise format, thus we use this format in our work.

2.2. SPARQL queries and triple store

SPARQL Protocol and RDF Query Language (SPARQL) [8] is a query language for RDF. It can be used to express queries across diverse data sources, whether the data is stored natively as RDF or viewed as RDF via middleware. Executing SPARQL queries often involves graph pattern matching. For example in Fig. 2, given an image URI, the query returns the age and the name of the image’s creator. SPARQL allows the definition of prefixes to make queries concise, in Fig. 2, the prefix “foaf” stands for “http://xmlns.com/foaf/0.1/ ”.

Please cite this article as: H. Chen et al., An intelligent annotation-based image retrieval system based on RDF descriptions, Computers and Electrical Engineering (2016), http://dx.doi.org/10.1016/j.compeleceng.2016.09.031