Facilitating Image Exploratory Search with Relations

Qi Li, Daqing He
School of Information Sciences
University of Pittsburgh
{qil14,dah44}@pitt.edu

ABSTRACT
Traditional text-based image retrieval does not fully support queries on semantic relationships between two entities. To better help the exploratory search on image collections, this paper presents a system for automatically extracting the relations between entities by analyzing the sentence dependency on the descriptions of the images. Our results demonstrate that using the extracted relations is not only beneficial for understanding the data set but also an effective way to facilitate users' exploratory searches.

Keywords
Exploratory search, relations, text-based image retrieval.

INTRODUCTION
Image retrieval on generic photographic collections (i.e., containing everyday real-world photographs) is usually based on the metadata of the images, which include short captions such as the titles, locations, dates or additional notes, as well as the descriptions of the photographs stored in a database. Explicit conceptual dimensions (Hearst, 2006) or entities (Lin et al, 2010) are often used to describe the images for image findings. In this paper, we propose a novel image retrieval system that extracts relations from the descriptions of images to represent the contents of images. In order to exact the relations from the image's descriptions, we apply the dependency parse trees to analyze the dependency between the components among the sentences. With well represented objects as well as their relations within an image, a relation retrieval system is designed to facilitate object and relation exploratory searches.

We believe that the content of an image consists of objects, events, people, times, and other information. The understanding of an image starts with the recognition of the existences of these pieces of information. Moreover, it does not stop at concept recognition. Concepts are componentized, aggregated, and related to other concepts to represent the whole meanings or contents of the images. Relationships among components or other distinct concepts are critical to the definition of concepts and the object's context in the image. For example, basic relationships such as "part of" or "is-a" frequently appear in the thesauri (e.g., synonyms, broader/narrower terms). However, it is far from enough to represent the relation between the concepts. To understand an image, people must first identify what components are? What is the nature of the image? What is its form and extent? What are the entities in the images? Once the basic objects in images are understood, more complex questions can be raised (Möller and Neumann, 2008). For example, with two people in an image, users would like to search for the position between these two people, e.g. in front of or near. Therefore, with the relationships in the images, how can such relationships contribute to the image retrieval?

We believe that the relation retrieval can contribute to or improve the entity-based exploratory search in the following two cases: searching for common entities and searching for an entity in different scenarios. When retrieving a specific entity with the proper name, it is very easy and effective to issue queries with the name of entities, such as "Lawrence, David Leo" or "Pittsburgh, Pennsylvania". But when this collection is about the average people as well as their daily lives, it might be hard to generate queries with proper terms. For example, the real description of an image is "Portrait of man wearing fedora, eyeglasses, polo shirt, light colored plaid suit, striped belt, and watch chain, standing with left hand on light colored automobile with open hood inscribed". The subject in this image is "man", which usually is treated as a stop-word in the conventional information retrieval system and thus is useless in the retrieval. Although it is a stop-word in document retrieval, it is a meaningful searching object in the images, which can be part of the image, such as describing a man's wardrobe (such as, fedora, eyeglasses, shirt etc.)

Relations can also help to distinguish an entity in different contexts. The entity of "flowers" can exist in the images in them, such as a person wearing a flower, as the main object of flowers or the images with flowers as a decoration or brooch. Even the Thesaurus for Graphic Materials only covers the indexing of subject terms and genre and physical characteristic terms (Library of Congress. Prints and Photographs Division, 1995). In this scenario, they annotate the flower as a subject without indicating the relation with other subjects. It is not enough to represent the image with just the subject. The relationships can help to distinguish the role of the flower in different scenarios. For example,

ASIST 2011, October 9–13, 2011, New Orleans, LA, USA.

the flower in "Group portrait of six women, including one wearing light colored belted dress with large button or brooch, eyeglasses, corsage, light colored flower in hair …" can be represented as decoration (Hair, Flower). The flower in "Woman wearing dress or robe with large flower print posing behind block in Harris studio" can be represented as decoration (Dress, Flower). The flower in "Man's body in open casket with flowers" can be represented as position (casket, flowers). Relationships can represent the content better.

In this paper, we propose an image retrieval system, which searches on the extracted relations from the descriptions of images. We expect these relations can better represent the contents of images. Many approaches are applied to relation extraction. Pattern-based extraction approaches were set up to increase specificity of the extraction, but they achieve significantly lower recall (Hearst M. , 1992). This method predefines some patterns to extract the "is-a" relation from the sentences. Later, the pattern detection approach is expanded to non-hierarch structures. Some relation extraction algorithms are supervised learning based approaches. That is, with the large training data set, the algorithm learned a model from a large annotated training corpus and used the model for further relation extractions (Huang, 2005). Semi-supervised learning, such as bootstrapping methods, is also applied in the relation extraction (Li, He, & Mao, 2009). This approach learns from some good quality seeds to extract the patterns and then relies on these patterns to further extract the relations from the Web with web mining techniques. Because of the lack of training data set, however, in this study, we propose relying on the sentence dependency analysis to extract the relations from the image descriptions.

EXPLORATORY SEARCH ON ENTITY RELATIONSHIPS

Our system is consisted with two parts—relation extraction, and relation indexing and retrieval. The overall architecture of the system is as shown in the Figure 1.

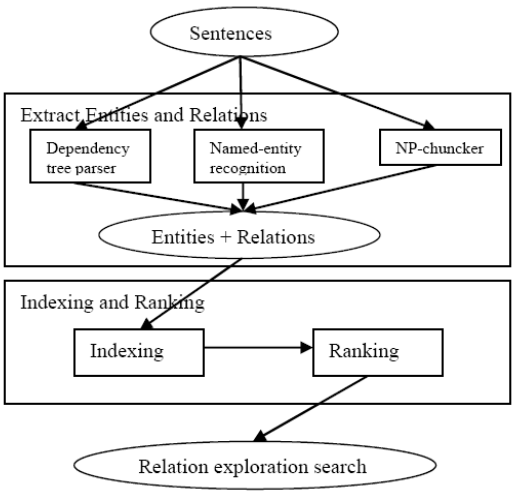

Figure 1. The overall architecture of relation extraction for museum archives

The first part is relation extraction. The corpus is preprocessed with named entity recognizer, dependency parser, and noun-phrase chunker. The heuristic rules are applied on the results to extract the relations. The second part is the relation indexing and retrieval system. Hierarchical language model are used for the indexing and retrieval.

Sentence dependency analysis for entity relationships

Texts are preprocessed, firstly, with named entity recognition tagger. For example, the description is as follows:

"Group portrait of three men, including Joe Palmer on left, and man in center wearing plaid jacket, and two women, seated at table with food, in interior with large windows."

The entity recognition tools can identify the named entities of "Joe Palmer" as person. Secondly, in order to include the common entities, we apply the noun-phrase chuncker into the pre-processing to detect the noun phrases in the sentence as the candidate entities. Therefore, in the above example, we have more noun phrases, such as "group portrait", "three men", "plaid jacket", etc. Thirdly, dependency parse are applied for the sentence component dependency analysis, and the sample result is as follows, Figure 2.

Some heuristic rules are applied to the dependency parse results for the relation extractions. There are 26 rules applied in the system. It can be summarized as the following three categories.

1. Rules extracting the noun phrases and entities are generated. Noun phrases and entities from the preprocessing steps are all treated as entities in the relation extractions. Therefore, the algorithm needs to identify these building blocks. The further conjunction terms connecting two entities will be extracted to be a noun phrase as an "entity", such as "man and woman". Although "man" and "woman" are extracted as two entities, they are connected with "and" to set up an "entity" of "man and woman".

2. Rules extract the structures of the same verb connecting two entities as a relation. For example, nsubj(verb, nn_1) and prep_*(verb, nn_2) is extracted as verb_prep_*(nn_1,nn_2). In the case of nsubj (gathered-4,Men and women-1) and prep_for(gathered-4,Thomas Dewey and Earl Warren Republican presidential campaign-8), it will be extracted as gathered for-4(Men and women-1, Thomas Dewey and Earl Warren Republican presidential campaign-8)

3. Rules extract the structures with preposition word to represent the "ownership" or "position" between two entities. The example of position relation is like prep_of (portrait-2, men-5), prep_including (men-5, Palmer-9), prep_on(Palmer-9, left-11), prep_in

(women-23, interior-32), WITH (Women-1, dictaphones and typewriters-3), prep_outside (Open-12,the Pittsburgh Board-22), and so on.

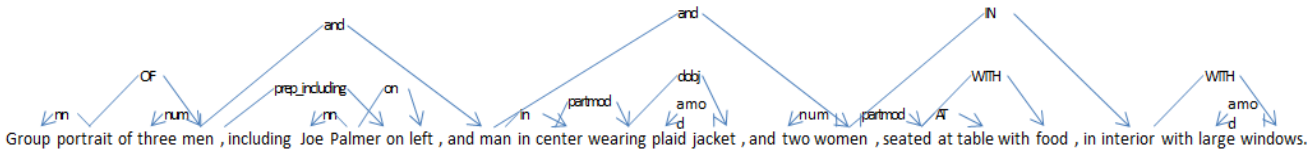

Figure 3. The visualization of dependency parse tree

With the aid of the dependency parser and the relation extraction rules, the sample output of the above example is like figure 4.

Expected Relations:

R1: Group portrait of three men, and man, and two women
R2: three men including Joe Palmer
R3: Joe Palmer on left
R4: man in center
R5: man wearing plaid jacket
R6: two women seated at table

Figure 4. The sample output relations

Indexing and Ranking algorithm

The goal of the retrieval task is to return a ranked list of relevant images which are represented as multiple relationships as the form of triples (E_1, R, E_2) according to the user's information need which are also represented as a query triple (e_1, r, e_2). In the pre-processes, we do not distinct the entity E_1 and E_2, so we only denote the entities as E. Therefore, we formalize the relation retrieval task as the following probability: $p(e_1, r, e_2 \mid E, R, D)$. In order to estimate the probability, we rely on a generative model. First, if we assume the elements in the query are independent, then we have the following formula:

$$p(e_1, r, e_2 \mid E, R, D) = p(e_1 \mid E, R, D)\, p(r \mid E, R, D)\, p(e_2 \mid E, R, D)$$

If we further assume that the document, entities and relationships are independent, then we have the following:

$$p(e_1, r, e_2 \mid E, R, D) = p(e_1 \mid E, D)\, p(r \mid R, D)\, p(e_2 \mid E, D)$$

Two-level Jelinek-Mercer smoothing is applied to estimate the probability. The estimation for the first quantifier of relation retrieval is as shown in follows:

$$p(e_1 \mid E, D) = \lambda_1 \frac{tf(e_1, E, D)}{tf(E, D)} + \lambda_2 p(e_1 \mid D) + \lambda_3 p(e_1 \mid C)$$

The first part is to estimate $p(e_1 \mid E, D)$ using the documents containing the entities, and the second part and the third part are using the document and corpus for the smoothing respectively.

Similarly, we can calculate the other two quantifiers as follows:

$$p(e_2 \mid E, D) = \lambda_1 \frac{tf(e_2, E, D)}{tf(E, D)} + \lambda_2 p(e_2 \mid D) + \lambda_3 p(e_2 \mid C)$$

$$p(r \mid R, D) = \lambda_1 \frac{tf(r, R, D)}{tf(R, D)} + \lambda_2 p(r \mid D) + \lambda_3 p(r \mid C)$$

The structured retrieval approach is used to index the corpus which has been pre-processed with the annotations of entities and relations.

EXPERIMENTS

This work uses the Tennie Harris Archive (THA) [1] for the demo and the evaluation. THA is an image collection of Charles "Teenie" Harris (1908–1998) who photographed the events and daily life of African Americans for the Pittsburgh Courier from the 1930s to 1970s. The collection has nearly 80,000 photographic negatives, few of which are titled and dated. Stanford named entity recognition tool and the dependency parser with the rules are applied to identify the relations in the collection. Indri retrieval system based on statistical language model is used for the indexing and retrievals in the field.

The goal of the study is to assess the usefulness and the value of relation-based faceted search. Our hypotheses are that the relation would be useful for users to satisfy their requirements. The search interface and the sample results are as shown in Figure 5. All the pictures are ranked according to the relevant scores. The left part is the description of the picture, and the right part is the corresponding picture.

In the evaluation, we test the two scenario mentions mentioned at the beginning of the paper. The first one is searching for the common entities with special task. We design the information need as "finding the photos with woman and flowers". With the query of one entity for "woman" and the other entity for "flower", we get the results including a woman with flower in her hair, a woman in casket with flowers, a women presenting flowers to someone, woman handing small vase with flowers, woman wearing a hat with flowers, woman standing in front of

[1] http://www.cmoa.org/teenie/intro.asp

walls painting of flowers, woman standing behind the table with flowers, and so on. The common entity's retrieval can be improved with the aid of the relation with another entity. The second scenario is to evaluate the retrieval of the special relationships with well definition. With the information need of "finding the woman having flowers in the hair", the query of one entity for "woman" and the other entity for "flower" as well as the relation for "wearing" can find the proper results.

CONCLUSION AND FUTURE WORK

This paper proposes a system to help people explore the images with the descriptions represented as the relationships across different entities. These relationships enable users to understand the scope and the extent of the corpus through active exploration of different entities. This is our system to investigate the exploratory search with the aid of relation retrieval. Although the system can search on relations, it is far from a well-structure system. For relation extraction, the current system is built on the dependency parsing results with heuristic rules. In the future work, we would like to have automatic methods for the extraction. It is more important to detect the same relationships but they are represented in different word surfaces. For example, both "woman with flower in her hair" and "woman wearing flowers in hair" can represent the same relation. Unfortunately, the current system cannot detect them. In fact, this example also raises another limitation of current system, that is, the relations are binary ones. It cannot represent triple relations such as "someone has something in some place". Solving this problem will be another work in the future.

REFERENCES

Hearst, M. (1992). Automatic acquisition of hyponyms from large text corpora. *In the 14th International Conference on Computational Linguistics (COLING-92)*, (pp. 539-545). Nantes, France.

Hearst, M. A. (2006). Design recommendatons for hierarchial faceted search interfaces. the Faceted Search Workshop at *SIGIR*.

Huang, M. a. (2005). Discovering patterns to extract protein protein interactions from full texts. *Bioinformatics*, 3604-3612.

Li, Q., He, D., & Mao, M. (2009). A Study of Relation Annotation in Business Environment. *In the Third IEEE International Conference on Semantic Computing*.

Library of Congress. Prints and Photographs Division, E. B. (1995). Thesaurus for graphic materials. *Cataloging Distribution Service, Library of Congress*.

Lin, Y., Ahn, J.-w., Brusilovsky, P., He, D., & Real, W. (2010). ImageSieve: Exploratory Search of Museum Archives with Named Entity-Based Faceted Browsing. *ASIST2010*. Pittsburgh, PA, USA.

Möller, R.and Neumann, B. (2008) Ontology-based Reasoning Techniques for Multimedia Interpretation and Retrieval. In *Semantic Multimedia and Ontologies: Theory and Applications*. Springer.

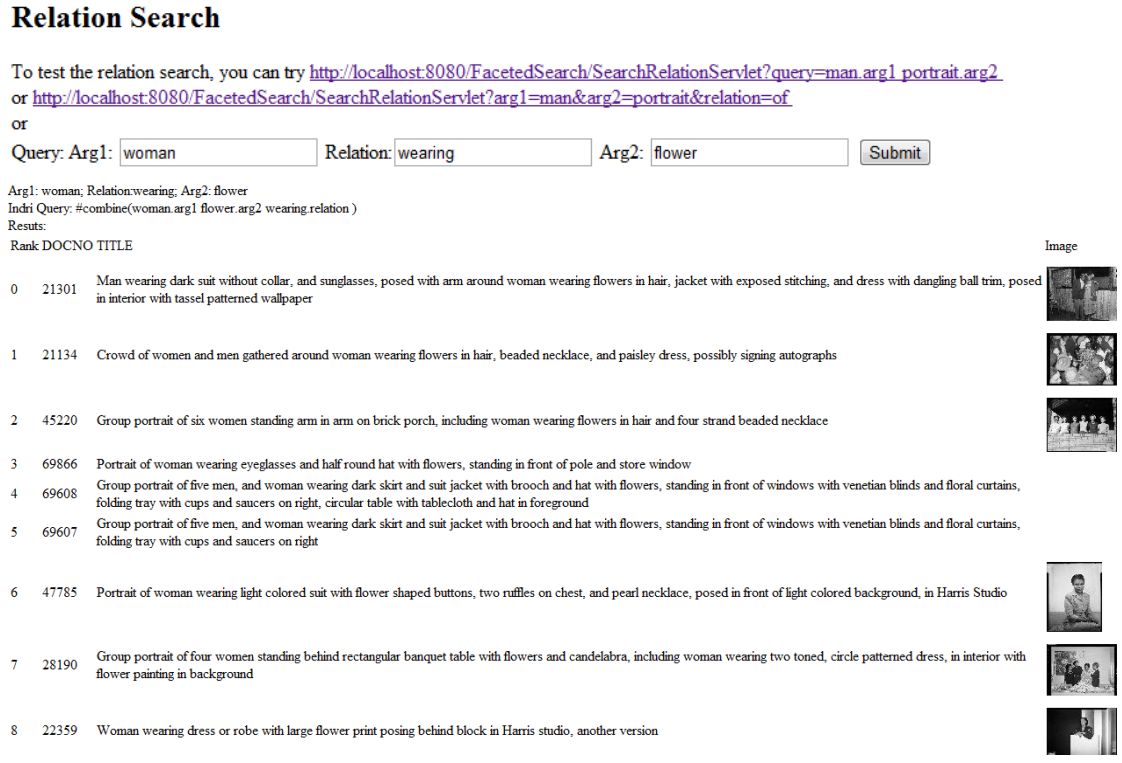

Figure 5. The interface of THA relation exploratory search