# Ontology-based Image Retrieval

Wei Zheng, Yi Ouyang, James Ford, Fillia S. Makedon
Department of Computer Science
Dartmouth College
Hanover, NH 03755
USA
Wei.Zheng, Yi.Ouyang@Dartmouth.EDU, jford@cs.dartmouth.edu, Fillia.S.Makedon@Dartmouth.EDU

*Abstract:* - The current most desirable image retrieval feature is retrieving images based on their semantic content. Currently there are two major image retrieval paradigms that attempt to provide this: text-based metadata image retrieval and content-based image retrieval. In practical applications, both have limitations. In this paper, we discuss an ontology-based image retrieval approach that aims to standardize image description and the understanding of semantic content. Ontology-based image retrieval has the potential to fully describe the semantic content of an image, allowing the similarity between images and retrieval query to be computed accurately.

*Key-Words:* - Image Retrieval, Ontology-based, Similarity, Semantic, Combined Concept Entity, attributes

## 1 Introduction

Images are a major information source in the real world, and represent features of objects like their color, shape and other attributes. In this paper we consider the problem when an end-user is faced with a repository of images whose content is complicated and partly unknown to the user. Such situations recur frequently when using public image databases on the web. Facing such a massive image repository, users want to find some specific images satisfying their requirements. Considerable research has been done to propose effective mechanisms for image retrieval based on image content. Two major paradigms are: text-based metadata image retrieval and content-based image retrieval (CBIR).

### 1.1 Text-based Metadata Image Retrieval

In the text metadata-based approach, image retrieval is based on textual descriptions about images. An image may be described by a natural language description or by a set of keywords. Corresponding to this description, text-based metadata image retrieval is implemented by matching the text metadata with image retrieval queries. This mechanism is relatively simple to implement and easy to use. However in real application, it is normally that the documentation having the keyword is not relevant and relevant documentation may not include the same keyword. So the most obvious problem of text-based image retrieval is that it can sometimes provide too many redundant images or no image at all [2][11]. Finally, even if the accuracy of text-based image retrieval were optimal, text metadata is expensive to provide in terms of human effort, and alternative approaches are needed.

In fact, the accuracy of text metadata image retrieval is far from perfect. One reason is that the text metadata for an image is often text in natural language. It is very difficult to accurately extract exact keywords from such a representation, which result in inaccuracy in later retrievals. Another common reason for poor retrieval accuracy is that a user may know little about the domain, and thus can't specify the most appropriate keywords for image retrieval.

### 1.2 Content-Based Image Retrieval (CBIR)

In CBIR [3] images are retrieved without using externally provided metadata describing their content. The contents used to represent images include features such as color, texture, shape, and spatial location are extracted and used. So content-based image retrieval is in fact focused on the visual features. The examples of CBIR systems include: PicSOM system [8], QBIC[15], Virage[16] etc.

The implementation of content-based image retrieval depends mainly on advanced image processing and pattern recognition techniques, and so effective content-based image retrieval is difficult to realize. Even assuming perfect feature extraction, the accuracy of retrieval result is not optimal. For example, two images can be very similar in color, size, and shape, despite containing very different objects. Further, no amount of image processing will enable queries for semantic categories like "politician" to succeed.

## 2 Semantic Image Annotation

Semantic technologies like ontologies and the XML markup language provide tools for a promising new approach to image retrieval based on implementing semantic understanding of image content. Ontology-based image retrieval has two components: semantic image annotation and semantic image retrieval. Semantic image annotation focuses mainly on the description of image content, and tries to describe image content as fully as possible. Based on the resulting semantic content description, semantic image retrieval allows searching and retrieval based on image content. Compared with text-based metadata image retrieval and content-based image retrieval, ontology-based image retrieval is more focused on capturing semantic content, which has the potential to satisfy user requirements better.

### 2.1 Ontology

An ontology is a specification of an abstract, simplified view of the world [5]. An ontology defines a set of representational terms called concepts. An ontology can be constructed in two ways: domain-dependent or generic. Generic ontologies are definitions of concepts in general; such as WordNet [9], which defines the meaning and interrelationships of English words. A domain-dependent ontology generally provides concepts in a specific domain, which focuses on the knowledge in the limited area, while generic ontologies provide concepts more comprehensively.

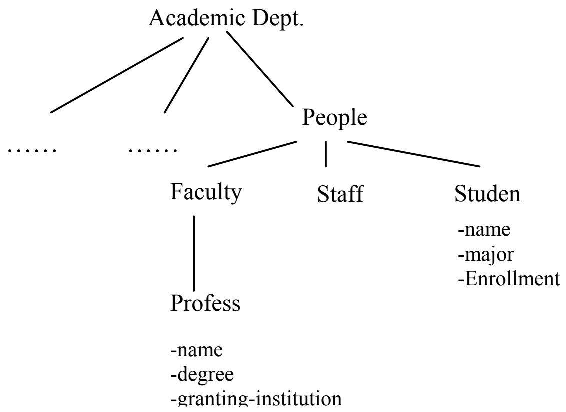


Figure 1: Ontology Example

The implementation of an ontology is generally taxonomy of concepts and corresponding relations [4]. In an ontology, concepts are the fundamental units for specification, and provide a foundation for information description. In general, each concept has three basic components: terms, attributes and relations. Terms are the names used to refer to a specific concept, and can include a set of synonyms that specify the same concepts. Attributes are features of a concept that describe the concept in more detail. Finally relations are used to represent relationships among different concepts and to provide a general structure to the ontology. Figure 1 is an example of a simple ontology about the organization of an academic department.

In this ontology example, every node is a concept defined about the organization. For each concept, there should be a set of attributes used to specify the corresponding concept. For instance, for the concept "Professor", the attributes of name, degree, and granting-institution are shown, and help explain the corresponding concept. In a complete ontology definition, there would of course be attributes for all the other concepts, including "Staff", "Faculty", *etc.* and then would likely be many more concepts and attributes.

The relations between different concepts are also simplified. In a real application, several types of concept relations are used.

Three kinds of inter-relationships are generally represented in an ontology: "IS-A", "Instance-Of", and "Part-Of". These relations correspond to key abstraction primitives in object-based and semantic data models [5]. "Instance-Of" relation shows membership between concepts, while "Part-Of" shows composition relationships. In this paper, we mainly focus on the "IS-A" relation, which show concept inclusion, because the similarity comparison in our ontology-based image retrieval is mainly based on "IS-A" relations. When concept has an "IS-A" relation to another concept, this means that the second concept is more general than the first concept. If concept A has relation "IS-A" to concept B, we call concept A a subconcept and call concept B a superconcept. For instance, in the ontology example above, concept "Professor" has "IS-A" relation to concept "Faculty", which is a more general concept compared with "Professor". Thus "Faculty" is superconcept to the concept "Professor". One characteristic of "IS-A" relation is that all the attributes of a superconcept can be inherited by its subconcepts. Subconcepts normally have more attributes than superconcepts and as a result, correspondingly subconcepts are more specific.

The “IS-A” relation is very important in similarity comparison between retrieval a query and images. If we try to find an image including a superconcept, the image with subconcepts should also be considered similar in image retrieval. For example, if the retrieval query is to find an image including a faculty member, an image with a professor should satisfy the retrieval request.

### 2.2 Semantic Annotation

The objective of semantic annotation is to describe the semantic content in images and retrieval queries. Currently, semantic annotation of an image or retrieval query still needs the intervention of a human being. Semantic annotation requires some understanding of the semantic meaning in images and retrieval query, and standardization of representation of images. Based on the semantic annotation of images and retrieval queries, we can compare semantic similarity between images and a retrieval query.

Figure 2: Example Image: A discussion between a professor and a student

At present, semantic annotation is implemented by some markup language such as XML based on a shared ontology definition. The markup language provides a mechanism for describing information in a structural way. The shared ontology definition provides a standard repository of concepts, which are used to describe all the entities that may be related to the image content. Figure 2 is an image of a discussion between a professor and a student. If we base our semantic annotation on ontology like the example in Figure 1, using XML we can get the semantic annotation shown in Figure 3.

The semantic annotation just describes who the people in the image are, which is confined by the definition of the corresponding ontology. The more comprehensive the ontology, the more accurate the semantic annotation will be. If a behavior concept such as “Talk” were defined in the ontology above, we could use it to specify what the professor and student are doing. Since “Student” and “Professor” are both concepts defined in the ontology, if we assume that attributes “name”, “enrollment” and “major” are available for the concept “student”, the semantic annotation shown in Figure 3 results.

```
<?xml version= “1.0”?>
<?xml-stylesheet href=“img.xsl” type =“text/xsl”?>
<image>
  <Professor>
     <name> K.Cook</name>
     <degree>Ph.D.</degree>
     <granting-institution>Univ. of Michigan</granting-institution>
   </Professor>
   <Student>
     <name>Jane</name>
     <major>computer science</major>
     <enrollment>2002</enrollment>
   </Student>
</image>
```

Figure 3: Semantic Annotation

Image retrieval queries can be formalized in the same way. Based on the shared ontology, we can design the query interface to use only concepts defined in the ontology. Together with the concepts, users should also be allowed to propose some constraints on the attributes for corresponding concepts. The system can then construct an XML file to specify the retrieval query. Using the formalizations for both the images and a retrieval query, we can compare the semantic similarity between each image and the user’s query.

## 3 Ontology-based Image Retrieval

Through semantic annotation, both images and retrieval queries can be formalized as XML files. In semantic annotation, the semantic meanings of images and queries are described based on a combination of concepts defined in an ontology. In image retrieval, the goal is to determine the similarity between images and a retrieval query. To achieve this objective in ontology-based image retrieval, we implement similarity comparison in two steps: extraction of combined concept entities and similarity comparison between images and a retrieval query.

### 3.1 Extraction of Combined Concept Entity

Since each semantic annotation is a description based on the concepts in an ontology, understanding of the corresponding concepts is the first step needed to understand the whole semantic annotation. To do this, it is necessary first to determine what concepts are used to constitute a semantic annotation.

Because XML files are structural documents, it is easy to extract the concepts used as tags in an XML file. By parsing the XML files, one can get a set of concepts in the corresponding images or queries. Thus, in the example of Figure 3, we can extract the concepts “Student” and “Professor”. Together with these two concepts, we can also extract several attributes used to specify information about instantiations of these two concepts such as “name”, “degree”, “major”, “Enrollment”, *etc*. In general we use: $CS=(c_1, c_2, \ldots, c_n)$ refer to the set of concepts used in the semantic annotations. $c_i (i= 1 \ldots n)$ is one specific concept in this set.

In the concept set, there may be semantic affiliations among concepts. Some of the concepts actually describe attributes of another concept, for example in Figure 3, “name”, “degree”, and “granting-institution” specify attributes of the concept “Professor”, that might also be represented in the full ontology as concepts. These concepts can be considered to form a concept entity that represents a semantic unit. Therefore, concepts in the concept set are combined into combined concept entities when one serves as an attribute of another. After combining the attribute concepts in the concept set $CS=(c_1, c_2, \ldots, c_n)$, a set of combined concept entities results: $CES=(ce_1, ce_2, \ldots, ce_m)(m<=n)$, here each $ce_i (i= 1 \ldots m)$ is a combined concept entity.

In each combined concept entity, there is one core concept that uses other concepts to describe its attributes. The core concept is the theme of the combined concept entity, and decides what the combined concept entity represents. We represent a combined concept entity as:

$$ce = (cc, (ac_1, ac_2, \ldots ac_k)) \quad (1).$$

Here cc is the core concept, and $ac_i (i=1 \ldots k)$ refer to the concepts that are attributes of the core concept. For example in Figure 3, we get the following set of combined concept entities:

*((Professor, (name, degree, granting-institution)), (Student, (name, major, enrollment)))*

After the extraction of combined concept entities, the semantic annotation in an XML file is converted to a set of combined concept entities. The combined concept entity is the fundamental semantic unit in the described model of semantic annotation, and is also the basis for the similarity comparison between images and retrieval queries described in the following section.

## 3.2 Similarity Comparison between Image and Retrieval Query

From XML files based on a shared ontology, semantic annotations of images and retrieval queries can be derived as described above, and from them a set of concept entities can be extracted. Based on corresponding sets of combined concept entities, the next step is to compare the semantic similarity between images and retrieval queries.

Using combined concept entities, the similarity comparison problem can be formalized as follows: given two sets of combined concept entities, $CES_{image}=(ce_{image,1}, ce_{imgae,2}, \ldots, ce_{image,u})$ describing images and $CES_{query}=(ce_{query,1}, ce_{query,2}, \ldots, ce_{query,v})$ describing retrieval queries, similarity probability $P_{similarity}(CES_{image}, CES_{query})$ can be calculated. This paper focuses only on how to calculate the similarity probability, and not about how to use this probability to retrieve images. We believe that this depends on the specific application and domain, so our similarity probability is aimed at providing a basis for further image retrieval decisions.

For each combined concept entity in a retrieval query, a satisfaction probability $P_{satisfaction}(ce_{query,i})(i=1 \ldots v)$ is introduced. This probability specifies what proportion of each combined concept entities in the retrieval query is satisfied in each image’s set of combined concept entities. In addition to the satisfaction probability, each combined concept entity in a retrieval query also has a weight $w_i (i=1 \ldots v)$ related to it, which allows the user to express the retrieval priority for a concept. Using the satisfaction probability and retrieval weight, the similarity between an image and a retrieval query can be represented as:

$$P_{similarity}\left(CES_{image}, CES_{query}\right) = \sum_{i=1}^{v} w_i * P_{satisfaction}\left(ce_{query,i}\right)$$

Here it is required that $w_1 + w_2 + \ldots + w_v = 1$, which can easily be achieved by normalizing any set of concept entities.

In order to calculate the required satisfaction probability of each combined concept entity, we introduce the similarity probability between combined concept entities $P_{sim}(ce_a, ce_b)$ in the following section. Based on this, we define $P_{satisfaction}(ce_{query,j})(j=1 \ldots v)$ as:

$$P_{satisfaction}(ce_{query,j}) = \max(P_{sim}(ce_{image,i}, ce_{query,j}))(i \le u)$$
.

### 3.3 Similarity Comparison between Two Combined Concept Entities

To compute the analysis above, the final problem needing resolution is the calculation of the similarity between two combined concept entities. The problem can be formalized as: given two combined concept entities, $ce_a=(cc_a,(ac_{a,1},ac_{a,2},...,ac_{a,p}))$ and $ce_b=(cc_b,(ac_{b,1},ac_{b,2},...,ac_{b,q}))$, calculate $P_{sim}(ce_a,ce_b)$, the similarity probability between them.

If there is a sequence of concepts $c_1, c_2,...,c_s$ such that each $c_i$ is the subconcept of $c_{i+1}(i=1...s)$, we say that $c_1$ is a descendent concept of $c_s$ and $c_s$ is the ancestor concept of $c_1$. Based on this definition, we discuss calculation of $P_{sim}(ce_a,ce_b)$ in two cases, based on the relation of their attribute concepts (recall that attribute concepts are defined as components of a combined concept entity in Equation (1)).

For any attribute concept $ac_{b,i}(i=1...q)$, if there is an attribute concept $ac_{a,j}(j<=p)$, such that $ac_{b,i}$ and $ac_{a,j}$ are the same concept with different values, $P_{sim}(ce_a,ce_b)=0$. Otherwise, we calculate $P_{sim}(ce_a,ce_b)$ as:

$$P_{sim}(ce_a, ce_b) = Ratio * \frac{1}{q} * \sum_{i=1}^{q} P(ac_{b,i})$$

$$Ratio = \begin{cases} 1 & (cc_a \text{ is the same or descendent concept of } cc_b) \\ \dfrac{Num_b}{Num_a} & (cc_a \text{ is an ancestor concept of } cc_b) \end{cases}$$

$P(ac_b,i)(i=1...q)$ is a probability function indicating how much of the attribute concept $ac_{b,i}(i=1...q)$ is satisfied. If there is an attribute concept $ac_{a,j}(j<=p)$ such that $ac_{b,i}$ and $ac_{a,j}$ are the same concept with same value, $P(ac_{b,i})=1$; if there is no attribute concept $ac_{a,j}(j<=p)$ such that $ac_{b,i}$ and $ac_{a,j}$ are the same concept, we need to calculate $P(ac_{b,i})$ according the value domain of attribute concept $ac_{b,i}$. Because the value domain of different concepts should be different, calculation here is limited to specific attribute concept. $Num_a$ is the number of all the descendent concepts of $cc_a$; $Num_b$ is the number of all the descendent concepts of $cc_b$.

## 4 Conclusions

In ontology-based image retrieval, both images and retrieval queries are represented by XML files based on a shared ontology that encode semantic annotations. By extracting combined concept entity, fundamental semantic units can be extracted for both. Based on sets of combined concept entities, the mechanisms are presented to compare the semantic similarity between images and a retrieval query using concept relation "IS-A". Compared with other approaches, ontology-based image retrieval provides better standardization of information descriptions and potentially allows understanding of semantic content.

*References:*

[1] Blobworld image retrieval using regions. Technical report, University of California, Berkeley.

[2] *The SMART Retrieval System: Experiments in Automatic Document Processing.* Prentice-Hall, 1971.

[3] Automatic image content retrieval, *Are we getting anywhere?* De Montfort University, 1996.

[4] *Ontologies: Silver Bullet for Knowledge Management and Electronic Commerce.* Springer-Verlag, 2001.

[5] D. Brickley and R. V. Guha. Resource description framework (RDF) schema specification 1.0. *In W3C Candidate Recommendation* 2000-03-27.

[6] D. Fensel. The semantic web and its languages. *IEEE Intelligence Systems*, 2000.

[7] Oasim Iqbal and J.K. Aggarwal. Image retrieval via isotropic and anisotropic mapping. *In IAPR Workshop on Pattern Recognition in Information Systems.*

[8] M. Koskela, J. Laaksonen, S. Laakso, and E. Oja. The picsom retrieval system: Description and evaluations. *In In The Challenge of Image Retrieval*, Brighton, UK.

[9] B. McBride, A. Seaborne, and J. Carroll. Jena tutorial for release 1.4.0. *Technical report*, Hewlett-Packard Laboratories, Bristol, UK, 2002.

[10] Wolfgang Muller and Henning Muller. MRML: A communication protocol for content-based image retrieval. *Technical report*, University of Geneva, 2001.8

[11] Gerard Salton and Christopher Buckley. Term-weighting approaches in automatic text retrieva, *Information Processing and Management*, 1998.

[12] Arnold W.M. Smeulders. Content-based image retrieval at the end of the early years. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2000. 9

[13] E. Hyvönen, A. Styrman, and S. Saarela. *Ontology-based Image Retrieval.* Number 2002-03 in HIIT Publications, pages 15-27.

Helsinki Institute for Information echnology (HIIT), Helsinki, Finland, 2002. http://www.hiit.fi (31.3.2003).

[14] Von-Wun Soo, Chen-Yu Lee, Chung-Cheng Li, Shu Lei Chen, National Tsing Hua University**,** Ching-chih Chen , Simmons College**.** Automated Semantic Annotation and Retrieval Based on Sharable Ontology and Case-Based Learning Techniques. *Proceedings of the third ACM/IEEE-CS joint conference on Digital libraries*. Pages: 61 - 72, May 2003.

[15] W. Niblack, R. Baeber, and et al. The QBIC project: Querying images by content using color, texture and shape. *In Proc. SPIE Storage and Retrieval for Image and Video Databases*, Feb 1994.

[16] Jeffrey R. Bach, Charles Fuller, Amarnath Gupta, Arun Hampapur, Brandley Horowitz, Rich Humphrey, Ramesh Jain, and Chiao fe Shu. The Virage image search engine: An open framework for image management. *In Proc. SPIE Storage and Retrieval for Image and Video Databases.*