科技論文一

ELSEVIER

Available online at www.sciencedirect.com

ScienceDirect

Computer Standards & Interfaces 29 (2007) 196–204

COMPUTER STANDARDS & INTERFACES

www.elsevier.com/locate/csi

Standards for image annotation using Semantic Web

Latifur Khan

Department of Computer Science, University of Texas at Dallas, Richardson, TX 75083-0688, United States

Received 6 September 2005; accepted 20 March 2006
Available online 2 June 2006

Abstract

The impetus behind Semantic Web research remains the vision of supplementing availability with utility; that is, the World Wide Web provides availability of digital media, but the Semantic Web will allow presently available digital media to be used in unseen ways. An example of such an application is multimedia retrieval. At present, there are vast amounts of digital media available on the web. Once this media gets associated with machine-understandable metadata, the web can serve as a potentially unlimited supplier for multimedia web services, which could populate themselves by searching for keywords and subsequently retrieving images or articles, which is precisely the type of system that is proposed in this paper. Such a system requires solid interoperability, a central ontology, semantic agent search capabilities, and standards. Specifically, this paper explores this cross-section of image annotation and Semantic Web services, models the web service components that constitute such a system, discusses the sequential, cooperative execution of these Semantic Web services, and introduces intelligent storage of image semantics as part of a semantic link space.

Keywords: Semantic Web; Annotation; Image gallery; Ontology

1. Introduction

The impetus behind Semantic Web research remains the vision of supplementing availability with utility; that is, the World Wide Web provides availability of digital media, but the Semantic Web will allow presently available digital media to be used to serve new purposes, an example of which is image retrieval.

The Semantic Web is an extension of today's Web technology; it boasts the ability to make Web resources accessible by their semantic contents rather than merely by keywords and their syntactic forms. Due to its well-established mechanisms for expressing machine-interpretable information, information and Web services previously available for human consumption can be created in a well-defined, structured format from which machines can comprehend, process and interoperate in an open, distributed computing environment.

This proves to be quite advantageous with respect to data collection; intelligent software entities, or agents, can effectively search the web for items of interest, which they can determine with new semantic knowledge. For instance, sports images or articles can be retrieved from around the web and processed by the respective web services to enhance a website in terms of the sheer multimedia content available. In such a system, the Semantic Web serves as a large, automated image collection that may be used to populate an annotated image "gallery" [37]. This image "gallery" would be represented as a semantic link space that organizes like images together based on known image semantics; for example, all basketball images would be grouped as an image network, and so on.

Combining image retrieval with the Semantic Web, however, is not merely beneficial due to the availability of raw data or the potential for automated image annotation, but there is also the added benefit of using a web ontology, or a set of concepts and their interrelations. By using such ontologies not only to search for multimedia but also to classify it, the system ensures consistency in terminology, leading to more accurate and precise query results.

1.1. The approach

At present, there are vast amounts of digital media available on the web. Once this media gets associated with machine-understandable metadata, the web can serve as a potentially

E-mail address: lkhan@utdallas.edu.

0920-5489/$ - see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.csi.2006.03.006

P1

unlimited supplier for multimedia web services, which could populate themselves by searching via keywords and subsequently retrieving images or articles. This presents a novel approach to automatic image annotation or classification. In this case, not only is the annotation done automatically once both the support vector machine [1] and the Bayesian network are trained [2], but the source is replenished automatically, as well.

The image annotation task has been decomposed into classification of low-level, or atomic, concepts and classification of high-level concepts in a domain-specific ontology. In the general sense, concepts are atomic if they are terms that can describe specific objects or image segments. Examples would be ball, stick, net, and other well-defined objects. High-level semantic concepts, on the other hand, are used to describe an environment with a set of existing atomic concepts associated to it. For example, an image that contains a ball, a net, shoes, and humans can be described as a basketball game. The framework takes advantage of this natural gap in semantics, classifying atomic concepts using support vector machines and high-level concepts using Bayesian belief networks.

Upon classifying the image, the system would reflect the image semantics, its features, content, and semantic category, as part of a semantic link space. Fig. 1 illustrates the layered architecture. The bottom-most layer represents the original image, and the layer directly above it will represent the image semantics using an ontology. The semantic space can then, as mentioned, prepare image networks based on the available image semantics, and the features that correspond to the respective images will constitute the feature space. As part of the operation interface, a user or a web agent can query the system, which would search and retrieve image information from the underlying layer [3,30,4].

1.2. Experimental context

Results have shown that separating atomic classification from high-level classification improves the Bayesian classification by reducing the complexity of the directed acyclic graph associated with the Bayesian network. This way, image features are abstracted away and the Bayesian structure includes only semantic concepts. For atomic classification, we use support vector machine (SVM).

1.3. Contributions

This paper explores this cross-section of image annotation and Semantic Web services, models the web service components that constitute such a system, discusses the sequential, cooperative execution of these Semantic Web services, and presents the technical challenges. The main contributions deal with the integration of these new service-building technologies, the use of two classification methods to separate high-level and low-level semantic concepts, the hyperlink-based search and collection of fresh raw images using intelligent web agents, and of course the representation of key image information in terms of a semantic link space rather than a local image repository. Another key contribution is the use of a web ontology as a multipurpose tool that seamlessly integrates different service components; the ontology can serve as the Bayesian structure to classify images, a translator to understand user queries, and an instructor for agents that gather multimedia. Section 2 presents a proposed architecture for image annotation using Semantic Web. Section 3 presents technical challenges and implementation issues to build a framework for image annotation. Section 4 presents an experimental setup of our systems and results. Section 5 discusses the related work. Section 6 summarizes the paper and outlines some future research.

<table>
<tr><td colspan="2">Operation Interface</td></tr>
<tr><td>Feature Space</td><td>Semantic Link Space</td></tr>
<tr><td colspan="2">Semantic Web Representation (XML, RDF, Ontology, etc)</td></tr>
<tr><td colspan="2">Resource (Image) Entity Space</td></tr>
</table>

Fig. 1. Semantic space architecture.

2. Proposed architecture for the prototype system

The prototype system will consist of an interface through which users can query the system regarding specific topics (e.g., sports domain). This request would be processed by a service that would retrieve the relevant images and articles from their respective repositories and present them to the user. The system contains the central ontology (Fig. 2). This ontology will contain detailed information regarding a set of domains (e.g., the sports domain), specifically the different types of sports, the equipment involved, and so on. Moreover, the service ontology will be continually changing as agents are able to discover more sports and so on. For example, agents may discover a new sport, tennis for example, and add it into the ontology. Alternatively, agents may find additional gear that is associated to an existing sport, such as a baseball helmet. However, the system as discussed in this paper assumes a single ontology without any ontology merging, which is beyond the scope of this paper.

As depicted in Fig. 2, web agents collect unclassified images using a hyperlink-based approach in order to build the semantic link space containing image semantics and classifications; agents will discard images that cannot be supported by the system. Supported images are segmented into various objects, and the objects' features are subsequently extracted; features include hue, intensity, saturation, and shape. Then images will be annotated with the help of ontologies (see Section 3.2). Finally, the image semantics can then be stored as part of the semantic link space (see Section 3.3).

3. Technical challenges and implementation issues

To implement such a system involves an integration of several up and coming technologies. This section highlights some of the technologies involved and the challenges presented by each of them. First, we present segmentation technique to segment an image into a number of objects/visual tokens. Second, we will provide annotation for these objects (see Fig. 3). For this, ontology will be used which is a part of Semantic Web. This ontology will serve as a source for image objects and facilitate for the

P2

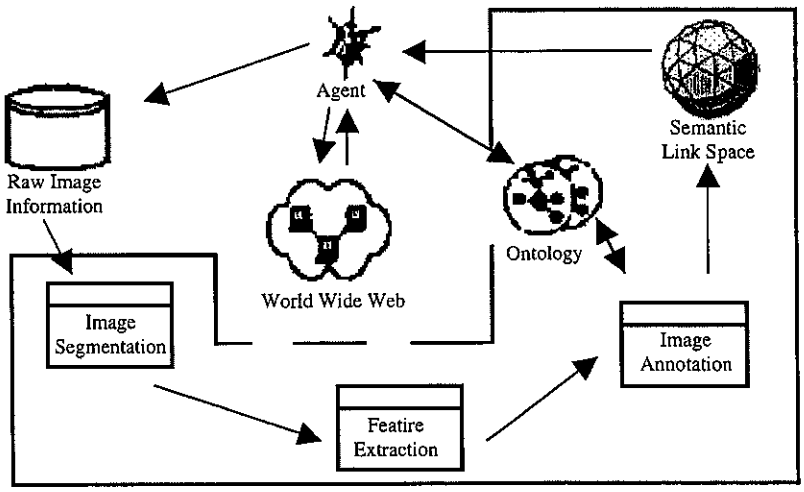


Fig. 2. Proposed architecture.

semantic contribution of semantic link space. Finally, we present semantic link space.

3.1. Segmentation

Therefore, we need first to identify object boundaries. For this, an object detection algorithm is invoked [5–10]. In this paper we only adopt the technique of Khan and Wang [11]; however, other techniques (e.g., normalized cut) can be used. This algorithm works in three stages. First, we detect all edge pixels in images and divide pixels into two sets, edge pixel and region pixel sets. Second, we grow a region from the region pixel set surrounded by edges taken from the edge pixel set. Finally, we may merge adjacent regions using an adjacency graph to avoid over segmentation of regions and to detect boundary of objects accurately. To illustrate the effectiveness of our algorithm in automatic image classification we implement a system aimed at the classification of images in the sports domain. By identifying objects in images, we observe how well our approach works when objects in images have varying degrees of complex organization along with shading and highlights.

3.2. Annotation

Annotation for an image is a process that facilitates descriptions of images. Annotation has been decomposed into identifying low-level and high-level semantic concepts, respectively. The former will be determined using support vector machines, and Bayesian's networks will determine the latter. After segmentation, features will be extracted from each image objects. Then feature vectors are sent into the support vector machine, which will identify low-level concepts that exist in

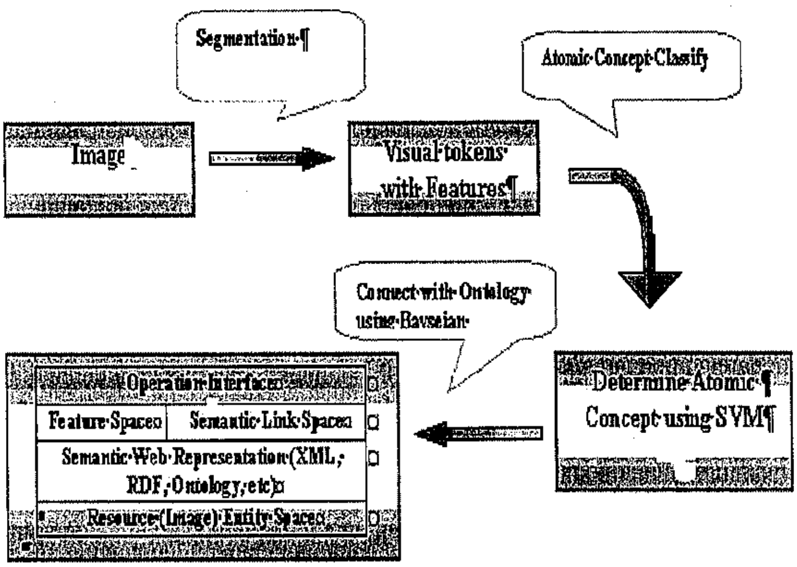


Fig. 3. Flow diagram of our approaches.

P3

each image. The set of low-level concepts are sent as known evidence into a trained Bayesian network, which will classify the images according to high-level semantic concepts.

The support vector method, or **SVM**, is a technique which is designed for efficient multidimensional function approximation; the aim is to determine a classifier or regression machine which minimizes the training set error. The basic procedure is to fix the empirical risk associated with an architecture and then to use a method to minimize the generalization error. The primary advantage of support vector machines as adaptive models for binary classification and regression is that they provide a classifier with a low expected probability of generalization errors. This approach can be trivially extended to multi-class classification by getting a binary response with respect to each atomic classification.

Bayesian networks are used to model the causal relationships that exist in the context of image semantics. This will allow a system to associate query keywords with other semantic concepts to some degree of belief. Hence, image queries will be more intelligently handled and will yield better results, a direct result of understanding the dependencies and relationships between different semantic concepts. The causal relationships can be patterned using the provided web ontology, which already represents a set of concepts and their interrelations. The variables that will make up the network will be a combination of high-level semantic classifiers as well as atomic level classifiers sent from the Support Vector Machine. For the purpose of this project, the Bayesian structure is precisely the web ontology (see Section 3.3.1).

3.3. Semantic link space

Upon linking the image with concepts of ontology, the system would reflect the image semantics, its features, content, and semantic category, as part of a semantic space. Fig. 1/bottom right of Fig. 3 illustrates the layered architecture of semantic link space. The bottom-most layer represents the original images, and the layer directly above it will represent the domain dependent ontology expressed in **OWL-S** including Semantic Web layers (see next paragraph). The next layer has two parts: the semantic link space represents linking between concepts and images, and linking across images; the feature space represents low-level image features. Now, the question is, why need this additional layer? There are several advantages of this layer. First, the web service can advertise its image content to other services by revealing images' semantic classification and features without requiring transfer images. Second, the agent can exploit this semantic link space by finding images of interest in bulk rather than blind individual search. Finally, this additional semantic layer minimizes image processing steps (e.g., downloading images, segmentation, and classification) within the system, since agents can discover similar images and combine similar image networks through this semantic link space. Furthermore, images will not be duplicated/downloaded, the systems using semantic link space will honor all original permission and rights of images. As part of the operation interface, a user or a web agent can query the system, which would search and retrieve image information from the underlying layer [4].

To enrich the Web by machine-processable information, the Semantic Web will be built up in different levels in an incremental fashion using the following layers (a) Unicode/**URI**, (b) **XML**/ Name Spaces/**XML** Schema, (c) **RDF**/**RDF** schema, (d) Ontology vocabulary. A common syntax is provided using the first two layers (i.e., Unicode, **XML**). **URIs** provide a standard way to refer to entities while **XML** fixes a notation for describing labeled trees. The next layer, **RDF** is a foundation for processing metadata; it provides interoperability between applications that exchange machine-understandable information on the Web. **RDF** emphasizes facilities which enable the automated processing of Web resources. Since **RDF** is designed to describe the resources and the relationship among them without assumption, the definition mechanism should be domain neutral, and can be applied to any domain. The next layers in the Semantic Web architecture are the ontology vocabulary (see Section 3.3.1 for details).

Therefore, information regarding classified images will be organized using this semantic link space. By associating like images together, image networks are created; similarity will be judged based on image semantics, including hue, saturation, intensity, and shape. By using this idea in conjunction with the hyperlink-based approach, user queries would be satisfied.

*3.3.1. **OWL-S***

To make use of a Web service, a software agent needs a computer-interpretable description of the service, and the means by which it is accessed. Semantic Web markup languages must not only establish a framework within which these descriptions are made and shared but also enable one web service to automatically locate and utilize services offered by other web services. **OWL-S** provides the solution, providing facilities for describing service capabilities, properties, pre-/postconditions, and /output specifications.

OWL-S is an ontology for describing Web services, enabling users and software agents to automatically discover, invoke, compose, and monitor Web resources offering services under specified constraints. **OWL-S** markup of Web services will facilitate the automation of Web service tasks, including automated Web service discovery, execution, composition and interoperation. First, we will present ontology and its characteristics. Next, we will present how **OWL-S** can be used to present the ontology.

An ontology is a specification of an abstract, simplified view of the world that we wish to represent for some purpose [12,13]. Therefore, an ontology defines a set of representational terms that we call *concepts*. Interrelationships among these concepts describe a target world. An ontology can be constructed in two ways, domain dependent and generic. **CYC** WordNet [14], or Sensus [15] are examples of generic ontologies. For our purposes, we choose a domain dependent ontology. This is because a domain dependent ontology provides concepts in a fine grain, while generic ontologies provide concepts in coarser grain.

Fig. 4 shows an example ontology for sports news, and Fig. 5 shows in Ontology Web Language [16,17]. Such an ontology is usually obtained from generic sports terminology and domain experts. Here, we represent our ontology as a directed acyclic graph (**DAG**). Each node in the **DAG** represents a concept.

P4

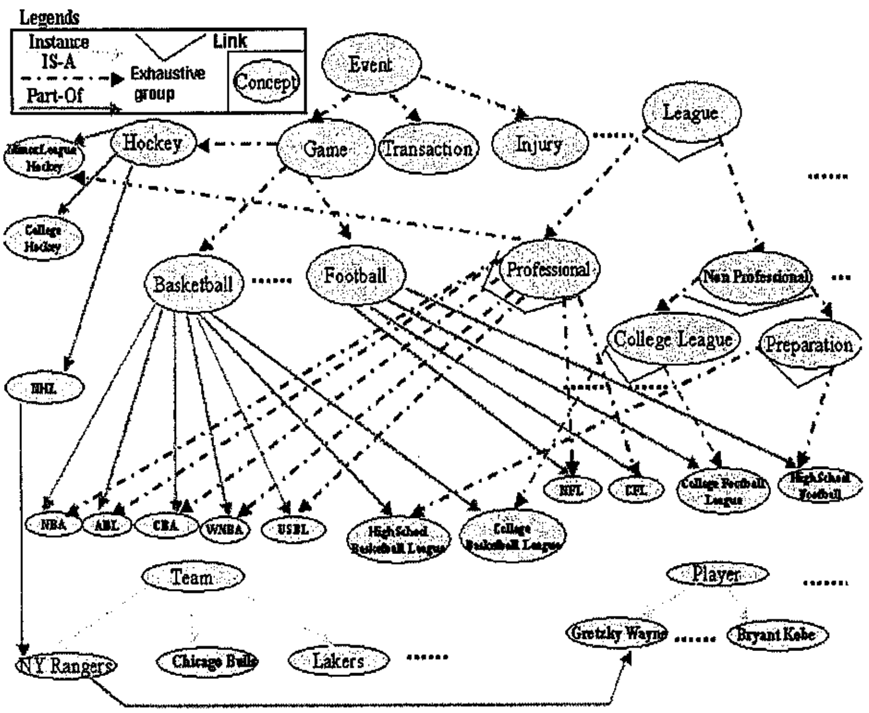


Fig. 4. A small portion of an ontology for a sports domain.

3.3.1.1. Interrelationships. Concepts are interconnected by means of interrelationships. If there is an interrelationship R between concepts C_i and C_j, then there is also a interrelationship R'(R inverse) between concepts C_j and C_i. In Fig. 4, interrelationships are represented by labeled arcs/links. Three kinds of interrelationships are used to create our sports ontology: specialization (Is-a), instantiation (Instance-of), and component membership (Part-of). These correspond to key abstraction primitives in typical object-based and semantic data models [18].

3.3.1.1.1. Please provide bookmark: Is-a. This interrelationship is used to represent specialization (concept inclusion). A concept represented by C_j is said to be a specialization of the concept represented by C_i if C_j is kind of C_i. For example, "**NFL**" is a kind of "Professional" league. In other words, "Professional" league is the generalization of "NFL." In Fig. 1, the Is-a interrelationship between C_i and C_j goes from generic concept C_i to specific concept, Cj represented by a broken line. Is-a interrelationship can be further categorized into two types: *exhaustive* and *non-exhaustive*. An exhaustive group consists of a number of Is-a interrelationships between a generalized concept and a set of specialized concepts, and places the generalized concept into a categorical relation with a set of specialized concepts in such a way so that the union of these specialized concepts is equal to the generalized concept. For example, "Professional" relates to a set of concepts, "**NBA**", "**ABL**", "**CBA**", ..., by exhaustive group (denoted by caps in Fig. 4).

3.3.1.1.2. Please provide bookmark: Instance-of. The Instance-Of relationship denotes concept instantiation. If a concept Cj is an example of concept Ci, the interrelationship between them corresponds to an Instance-Of denoted by a dotted line. For example, player "Wayne Gretzky" is an instance of a concept "Player." In general, all players and teams are instances of the concepts, "Player" and "Team" respectively.

3.3.1.1.3. Please provide bookmark: Part-of. A concept is represented by C_j is Part-Of a concept represented by C_i if C_i has a C_j (as a part) or C_j is a part of C_i. For example, the concept "**NFL**" is Part-Of the concept "Football" and player, "Wayne Gretzky" is Part-Of the concept "**NY** Rangers".

3.3.1.2. Disjunct concept. When a number of concepts are associated with a parent concept through Is-a interrelationships it is important to note when these concepts are disjoint, and are referred to as concepts of a disjoint type. For example, the concepts **NBA**, **CBA**, or **NFL** are associated with the parent concept Professional through association with Is-a, they become disjoint concepts.

Concepts are not disjoint, on the other hand, when they are associated with a parent concept through Instance-Of or Part-Of.

In Fig. 5 we show a part of ontology in ontology representation language, **OWL**. Concepts like, "Professional, "**NBA**", and "Team" are treated as Class in **OWL**; on the other hand, "**NY** Rangers", and "Gretzky Wayne" are treated as instance of some classes. Is-a interrelationship is expressed as subClassOf relationship which is defined in **OWL** (e.g.,<owl:Class rdf: ID="Injury"><rdfs:subClassOf><owl:Classrdf:ID="Event"/> </rdfs:subClassOf></owl:Class>). Here Part-Of relationship is defined as an ObjectProperty. For example, if concept, "**NY** Rangers" is a part of concept "**NHL**" then this Part-Of interrelationship will be stated in the following way: <Team rdf: ID="NY_Rangers"><PartOf rdf:resource="#NHL"/></Team>. In this example, "**NY** Rangers" is declared as an Instance-Of

P5

```
<?xml version="1.0"?>
<rdf:RDF
  xmlns:rss="http://purl.org/rss/1.0/"
  xmlns="http://a.com/ontology#"
  xmlns:jms="http://jena.hpl.hp.com/2003/08/jms#"
  xmlns:protege="http://protege.stanford.edu/plugins/owl/protege#"
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
  xmlns:owl="http://www.w3.org/2002/07/owl#"
  xmlns:vcard="http://www.w3.org/2001/vcard-rdf/3.0#"
  xmlns:daml="http://www.daml.org/2001/03/daml+oil#"
  xmlns:dc="http://purl.org/dc/elements/1.1/"
 xml:base="http://a.com/ontology">
 <owl:Ontology rdf:about="">
  <owl:imports rdf:resource="http://protege.stanford.edu/plugins/owl/protege"/>
 </owl:Ontology>
 <owl:Class rdf:ID="Professional">
  <owl:disjointWith rdf:resource="#Non-professional"/>
  <rdfs:subClassOf rdf:resource="#League"/>
 </owl:Class>
 <owl:Class rdf:ID="Injury">
  <rdfs:subClassOf>
   <owl:Class rdf:ID="Event"/>
  </rdfs:subClassOf>
 </owl:Class>
 <owl:Class rdf:ID="NHL">
      <PartOf rdf:resource="#Hockey"/>
 </owl:Class>
 <owl:Class rdf:ID="Team"/>
 <owl:Class rdf:ID="Player"/>
 <owl:Class rdf:ID="Basketball">
  <rdfs:subClassOf>
   <owl:Class rdf:about="#Game"/>
  </rdfs:subClassOf>
 </owl:Class>
 <owl:Class rdf:ID="MinorLeagueHockey">
      <PartOf rdf:resource="#Hockey"/>
 </owl:Class>
 <owl:Class rdf:ID="Game">
  <rdfs:subClassOf rdf:resource="#Event"/>
 </owl:Class>
 <owl:ObjectProperty rdf:ID="PartOf">
  <rdfs:range rdf:resource="http://www.w3.org/2002/07/owl#Class"/>
 </owl:ObjectProperty>
 <owl:Class rdf:ID="CollegeHockey">
      <PartOf rdf:resource="#Hockey"/>
 </owl:Class>
 <owl:Class rdf:ID="Transaction">
  <rdfs:subClassOf rdf:resource="#Event"/>
 </owl:Class>
 <owl:Class rdf:ID="Hockey">
  <rdfs:subClassOf rdf:resource="#Game"/>
 </owl:Class>
 <Player rdf:ID="Wayne_Gretzky">
  <PartOf rdf:resource="#NY_Rangers"/>
 </Player>
 <Player rdf:ID="Cobe_Bryant"/>
 <Team rdf:ID="NY_Rangers">
  <PartOf rdf:resource="#NHL"/>
 </Team>
 .........
</rdf:RDF>
```

Fig. 5. A small portion of an ontology for a sports domain in **OWL**.

"Team" (using Team tag/class). Disjoint/Disjunct properties of concepts are expressed as disjointWith property in OWL (e.g., <owl:Class rdf:ID="Professional"><owl:disjointWith rdf:resource="#Non-professional"/><rdfs:subClassOf rdf:resource="#League"/></owl:Class>).

By studying Fig. 4, it is clear that an ontology can be used as a central semantic library. In this case, a web service can present in a hierarchical manner all valid queries that may be made in the system. For instance, a user may be presented with choices in a top down manner where he or she may specify his query in greater and detail. To begin with, a user may want to retrieve images of outdoor sports, and then specify basketball as the sport of choice. By traversing the ontology, the web service will be able to refine the query further by allowing the user to choose specific items that are related to basketball, as per the ontology. Other web services, such as the **SVM** service, will also make use of the ontology to determine atomic concepts which must be recognized.

*3.3.1.3. **WSDL**.* Web Services Description Language (**WSDL**) is a new specification to describe networked **XML**based services [19]. It provides a simple way for service providers to describe the basic format of requests to their systems regardless of the underlying protocol, in our case **SOAP**. Under the **WSDL** standard, network services are viewed as a set of endpoints operating on messages containing either document-oriented or procedure-oriented information. The operations and messages are described abstractly, and then bound to a concrete network protocol and message format to define an endpoint. Related concrete endpoints are combined into abstract endpoints, or services. **WSDL** is extensible to allow description of endpoints and their messages regardless of what message formats or network protocols are used to communicate [4].

WSDL documents describe operations, messages, datatypes, and communication protocols specific to a web service. To carry out the communication between web services, **SOAP** will be used. **SOAP** provides the framework by which application-specific information may be conveyed in an extensible manner. Also, **SOAP** provides a full description of the required actions taken by a **SOAP** node on receiving a **SOAP** message. The **SOAP** stack will convert **SOAP** requests into native requests that the web service can make use of. Similarly, the web services' responses must be designed as **SOAP** responses.

Both languages are required for the full specification of a grounding, because the two languages do not cover the same conceptual space. The two languages do overlap in the area of providing for the specification of what **WSDL** calls "abstract types", which in turn are used to characterize the inputs and outputs of services. **WSDL** is unable to express the semantics of an **OWL** class. Similarly, **OWL-S** has no means, to express the binding information that **WSDL** captures. As shown in Fig. 6, **WSDL** documents describe operations, messages, datatypes, and communication protocols specific to a web service.

Fig. 6 above shows a **WSDL** document for a web service, Images_Service. The <operation> tag encapsulates the message exchange protocol; the input and output messages are simply abstract definitions of the data being communicated as part of this web service. The operation is defined within the <portType> tags, after which the concrete protocol is defined within the <binding> tags.

4. Results

We have developed a prototype system for image annotation in sports domain. Images are collected from the web and segmented [11]. A total of 3000 image objects are generated. For further tuning of image segmentation, human intervention was applied.

P6

```xml
<?xml version="1.0" encoding="UTF-8"?>
<definitions name="ImagesService"
  targetNamespace="http://www.dblab.utdallas.edu/wsdl/images.wsdl"
  xmlns="http://schemas.xmlsoap.org/wsdl/"
  xmlns:soap="http://schemas.xmlsoap.org/wsdl/soap/"
  xmlns:tns="http://www.dblab.utdallas.edu/wsdl/images.wsdl"
  xmlns:xsd="http://www.w3.org/2001/XMLSchema">

  <message name="ImagesRequest">
    <part name="keywordList" type="xsd:string"/> <!-- | is the delimeter -->
  </message>
  <message name="ImagesResponse">
    <part name="URLs" type="xsd:string"/> <!-- | is the delimeter -->
  </message>

  <portType name="Images_PortType">
    <operation name="getImageURLs">
      <input message="tns:ImagesRequest"/>
      <output message="tns:ImagesResponse"/>
    </operation>
  </portType>

  <binding name="Images_Binding" type="tns:Images_PortType">
    <soap:binding style="rpc"
      transport="http://schemas.xmlsoap.org/soap/http"/>
    <operation name="getImageURLs">
      <soap:operation soapAction="sendImageURLs"/>
      <input>
        <soap:body encodingStyle="http://schemas.xmlsoap.org/soap/encoding/"
        namespace="urn:examples:ImagesService" use="encoded"/>
      </input>
      <output>
        <soap:body encodingStyle="http://schemas.xmlsoap.org/soap/encoding/"
          namespace="urn:examples:ImagesService" use="encoded"/>
      </output>
    </operation>
  </binding>

  <service name="Images_Service">
```

Fig. 6. **WSDL** document for ImageService.

We use the **LIBSVM** [20] for **SVSM** implementation and use the $\nu-$**SVM** with **RBF** kernel. In our experiments we set ν very low (ν=0.5). Since, we address the problem of multiclass classification (i.e., various atomic concepts), we implement "one-vs-one" scheme. For the construction of sports domain dependent ontologies, first, we list all possible objects necessary to cover a given sports domain. This possible object list should include different sports, such as basketball, football, baseball, hockey etc, and different leagues within a given sport, as in basketball, **NBA**, **ABL**, **CBA** and so forth. Furthermore, different sports can be qualified by characteristics such as injuries, player transactions, strikes, etc.

We assume that the Bayesian structure is precisely this ontology. Furthermore, conditional probability tables will be constructed using gradient decent learning to find hidden variables [21].

Results from image annotation have proved that a properly trained support vector machine and Bayesian network can work alongside one another to produce satisfactory results. The **SVM** was trained with a mix of basketball, baseball, bat, soccer, hoop, and grass images that total 3000 training objects. The training set captured key characteristics of each image segment: hue, saturation, intensity, and shape. Once the support vector machine was trained, the training set was also used as test data in order to judge the training accuracy, which averaged at 98.5%.

The recall values indicate how well the system fared in recognizing all the segments that depict the same object. For instance, in the case of grass, the system is able to recognize 74 of the 80 grass objects, resulting in a 93% recall. However, there are 88 grass objects recognized, so the precision value is used to indicate how many of the retrieved positively classified images truly depict the object in question. In this case, 74 of the 88 positively classified grass objects are actually grass objects, leading to an 84% precision. Some of the problems with atomic classification are intuitive. In the case of a basketball hoop, the support vector machine is attempting to recognize an object that lacks a definite shape or color. Soccer balls are also tough to recognize because they are composed of two distinct colors: black and white. Complete results are presented in Tables 1 and 2.

Higher-level classification suffers in all instances where atomic classification falls short; however, one idea that deserves mention is the difference between misclassification and unclassification. For example, even if a basketball is not recognized as a basketball, there is still an inherent benefit of not recognizing the basketball as another type of object, a soccer ball for instance. In the case of the 150 basketball pictures, 136 were retrieved due to limited misclassification. On the other hand, of the 80 soccer images, 35 were misclassified at the atomic level as containing baseballs. In this case, the Bayesian network will be unable to accurately classify the image, which will subsequently be discarded.

The results, particularly the precision values, show that there are too many multiple classifications. For example, an image that contains a soccer ball, a bat, and a baseball will be retrieved both as a baseball image as well as a soccer image. Another important note is that, due to a simple Bayesian structure, images were often classified correctly if one of two objects were recognized. The system will be extended to recognize details pertaining to the environment so images can be classified on the basis of indoors or outdoors and team or individual.

5. Related works

Here, we will summarize key related efforts in image retrieval domain. First, we will discuss some image retrieval systems

Table 1
SVM classification results

Objects	Re-	Preci-
Soccer	39%	51%
Grass	93%	84%
Basket-	89%	100%
Baseball	83%	67%
Bat	100%	99%
Hoop	100%	37%

Table 2
Bayesian classification results

Concepts	Recall	Precision
Basketball	92.7%	76.8%
Soccer	100%	72.1%
Baseball	96%	71.1%

P7

based on low-level features and then we will present current state of the art of image annotations based on semantic level.

QBIC [22], Virage [23], RetrievalWare [24], Photobook [25], VisualSEEK [26], Netra [27], Surfimage [28] and **MARS** [29] are few image retrieval systems which are mostly focused on the retrieval based on low-level features rather than high-level concepts.

On the other hand, [3,30–36] propose semantic level image retrieval based on high-level concepts; however, linking between image low-level features and high-level concepts require human intervention which seriously cripples scalability of the approaches. In this paper we propose various mechanisms that automatically (at least semi-automatically) link low-level features with high-level concepts and facilitate higher accuracy retrieval.

6. Conclusion and future work

In this paper, we have presented a framework that can be used to provide annotation for images with the usage of Semantic Web. This Semantic Web will facilitate semantic link space along with image semantics, its feature, content and semantic category. To achieve semantic category of images, we have used ontologies. Lower level concepts of ontologies are linked images using support vector machines, and for higher-level concepts we have used Bayesian belief networks.

Future developments will focus on interoperability issues. The ultimate goal is to develop a system that uses truly independent web services with independent ontologies in a seamless fashion. This would give way to Intelligent Web Services, or **IWSs**. Ontology merging and mapping will be key research issues, and would utilize the unified modeling language **XML** Declarative Description, or **XDD**, and an **XML**-based declarative programming language **XML** Equivalent Transformation, or **XET**. Also, future enhancements will attempt to focus on two ideas. The first is the implementation of web agents that can effectively search the web for semantically relevant multimedia, whether these are images, articles, or video clips. The second is introducing a semi-automatic annotation system, which would use manual training in conjunction with feedback to intelligently annotate multimedia, thereby greatly improving the results.

References

[1] C. Cortes, V. Vapnik, Support vector networks, Machine Learning 20 (1995) 1–25.

[2] R.E. Neapolitan, Learning Bayesian Networks, Prentice-Hall, Upper Saddle River, NJ, 2004.

[3] O. Marques, N. Barman, Semi-automatic semantic annotation of images using machine learning techniques, The Semantic Web — ISWC 2003 Proceedings, 2003, pp. 550–565.

[4] Hai Zhuge. Semantic-Based Web Image Retrieval. http://www2003.org/cdrom/papers/poster/p172/p172-zhuge/p172-zhuge.htm.

[5] Jianbo Shi, Jitendra Malik, Normalized Cuts and Image Segmentation, IEEE Conf. Computer Vision and Pattern Recognition(CVPR), Puerto Rico, June 17–19, 1997, p. 731.

[6] M. Lybanon, S. Lea, S. Himes, Segmentation of diverse image types using opening and closing, Proc. IEEE Int. Conf. on Image Proc., 1994.

[7] Michael Hansen, William Haggins, Watershed-driven relaxation labeling for image segmentation, Proc. IEEE Int. Conf. on Image Proc., 1994.

[8] Ramin Samadani, Cecilia Han, Computer-assisted extraction of boundaries from images, Proc. SPIE Storage and Retrieval for Image and video databases, 1993.

[9] Yong Rui, Thomas S. Huang, Shih-Fu Chang, Image retrieval: current techniques, promising directions and open issues, Journal of Visual Communication and Image Representation 10 (March 1999) 39–62.

[10] B. Sumengen, B.S. Manjunath, C. Kenney, Image segmentation using multiregion stability and edge strength, Proc. IEEE International Conference on Image Processing(ICIP), Barcelona, Spain, Sep. 2003.

[11] L. Khan, L. Wang, Object boundary detection for hierarchical image classification, in: C. Djeraba, S.J. Simoff, O.R. Zaiane (Eds.), Mining Multimedia and Complex Data, Springer-Verlag Publishing, Heidelberg, Germany, ISBN: 3-540-20305-2, 2003, pp. 36–49, doi:10.1007/b12031, Chapter 6.

[12] N. Guarino, C. Masolo, G. Vetere, OntoSeek: content-based access to the Web, IEEE Intelligent Systems 14 (3) (1999) 70–80.

[13] T.R. Gruber, A translation approach to portable ontology specifications, Knowledge acquisition, An International Journal of Knowledge Acquisition for Knowledge-based Systems 5 (2) (1993) 199–220.

[14] D.B. Lenat, Cyc: a large-scale investment in knowledge infrastructure, Communications of the ACM 38 (11) (1995) 33–38.

[15] B. Swartout, R. Patil, K. Knight, T. Ross, Toward distributed use of large-scale ontologies, Proc. of the Tenth Workshop on Knowledge Acquisition for Knowledge-Based Systems, Banff, Canada, 1996.

[16] The OWL Services Coalition. OWL-S: semantic markup for web services. http://www.daml.org/ services/owl-s/1.0/owl-s.html.

[17] Protégé-2000. http://protege.stanford.edu/.

[18] L. Khan, D. McLeod, E. Hovy, Retrieval effectiveness of ontology-based model for information selection, The VLDB Journal: The International Journal on Very Large Databases, vol. 13(1), ACM/Springer-Verlag Publishing, 2004, pp. 71–85.

[19] E. Christensen, F. Curbera, G. Meredith, S. Weerawarana. Web Services Description Language (WSDL) 1.1. http://www.w3.org/TR/wsdl.

[20] Chih-Chung Chang, Chih-Jen Lin, LIBSVM: a library for support vector machines, 2001. http://www.csie.ntu.edu.tw/~cjlin/libsvm.

[21] Jiawei Han, Micheline Kamber, Data Mining: Concepts and Techniques, The Morgan Kaufmann Series in Data Management Systems, Jim Gray, Series Editor, Morgan Kaufmann Publishers, August 2000. 550 pages. ISBN 1-55860-489-8.

[22] W. Niblack, R. Barber, et al., The QBIC Project: querying images by content using color, texture, and shape, Proc. SPIE Storage and Retrieval for Image and Video Databases, Feb 1994.

[23] A. Gupta, R. Jain, Visual information retrieval, Communications of the ACM 40 (5) (May 1997) 71–79.

[24] James Dowe, Content-based retrieval in multimedia imaging, Proc. SPIE Storage and Retrieval for Image and Video Databases, 1993.

[25] A. Pentland, R.W. Picard, S. Sclaroff, Photobook: content-based manipulation of image databases, International Journal of Computer Vision 18 (3) (1996) 233–254.

[26] John R. Smith, Shih-Fu Chang, VisualSEEk: a fully automated content-based image query system, ACM Multimedia, 1996, pp. 87–98.

[27] W.Y. Ma, B.S. Manjunath, Netra: a toolbox for navigating large image databases, Proc. IEEE Int. Conf. on Image Proc., 1997.

[28] C. Nastar, M. Mitschke, C. Meilhac, N. Boujemaa, Surfimage: a flexible content-based image retrieval system, Proceedings of the 6th ACM International Conference on Multimedia, September 1998, pp. 339–344.

[29] Shara Mehrotra, Y. Rui, M. Ortega, T. Hunag, Supporting content-based queries over images in MARS, Proc. of IEEE Int. Conf. on Multimedia Computing and Systems, 1997.

[30] Oge Marques, Nitish Barman, Semi-automatic semantic annotation of images using machine learning techniques, International Semantic Web Conference, 2003, pp. 550–565.

[31] Th. Schreiber, B. Dubbeldam, J. Wielemaker, B.J. Wielinga, Ontology-based photo annotation, IEEE Intelligent Systems (May/June 2001).

P8

[32] Shiyong Lu, Rong Huang, Artem Chebotko, Yu Deng, Farshad Fotouhi, ImageSpace: an environment for image ontology management, International Journal of Information Theories and Applications 11 (1) (June 2004) (Also appeared in Proc. of ICTP'2004, Bulgaria, Macedonia, Greece).

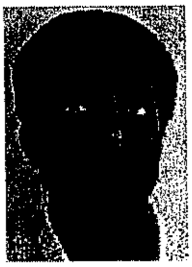
[33] Shiyong Lu, Rong Huang, Artem Chebotko, Yu Deng, Farshad Fotouhi, ImageSpace: an environment for image ontology management, Proc. of the Second Special Workshop on Multimedia Semantics of the 29th International Conference on Information and Communication Technologies and Programming, Bulgaria, Macedonia, Greece, June 2004.

[34] K. Goh, E.Y. Chang, W.-C. Lai, Concept-dependent multimodal active learning for image retrieval, ACM International Conference on Multimedia, New York, October 2004.

[35] Sally Goldman, Qi Zhang, Wei Yu, et al., Content-based image retrieval using multiple-instance learning, Proceedings of the Nineteenth International Conference on Machine Learning (ICML 2002), July 2002.

[36] G. Sheikholeslami, W. Chang, A. Zhang, SemQuery: semantic clustering and querying on heterogeneous features for visual data, IEEE Transactions on Knowledge and Data Engineering (TKDE) 14 (5) (September/October 2002) 988–1002.

[37] Eero Hyvönen, Samppa Saarela, Avril Styrman, Kim Viljanen, Ontology-Based Image Retrieval, http://www.cs.helsinki.fi/group/seco/presentations/www2003/p199-hyvonen.html.



Dr. Latifur R. Khan is currently an Assistant Professor in the Computer Science Department at the UTD, where he has taught and conducted research since September 2000. His research work is currently supported by grants from the Air Force Office of Scientific Research (AFOSR), National Science Foundation (NSF), the Nokia Research Center, Alcatel, Raytheon, and the SUN Academic Equipment Grant program. Dr. Khan is one of the principal investigators at the CyberSecurity and Emergency Preparedness Institute at UTD, where he is involved with finding solutions to deal with the rapidly growing Homeland Security problems in cybercrime, information assurance, and emergency preparedness. In addition, Dr. Khan is the director of the state-of-the-art DBL@UTD, UTD Data Mining/Database Laboratory, which is the primary center of research related to data mining and image/video annotation at University of Texas–Dallas. Dr. Khan's research areas cover data mining, multimedia information management, semantic web and database systems with the primary focus on first three research disciplines. He has served as a committee member in numerous prestigious conferences, symposiums and workshops including the ACM SIGKDD Conference on Knowledge Discovery and Data Mining. Dr. Khan has published over 80 papers in prestigious journals and conferences.

P9