SEMANTIC ANNOTATION AND RETRIEVAL OF IMAGE COLLECTIONS

Taha Osman[1], Dhavalkumar Thakker[1], Gerald Schaefer[2], Maxime Leroy[3], Alain Fournier[3]

[1] School of Computing & Informatics, Clifton Lane, Nottingham Trent University, Nottingham, NG11 8NS, UK
{Taha Osman, Dhavalkumar. Thakker,}@ntu.ac.uk

[2] School of Engineering and Applied Science, Aston University, Aston Triangle, Birmingham B4 7ET, UK
G.Schaefer@aston.ac.uk

[3] Naval Academy of France, Ecole Navale, BP 600, 29240 Brest, France
Contact: Professor Christophe Claramunt - claramunt@ecole-navale.fr

KEYWORDS

Digital Libraries, Semantic Web, Ontology Engineering, Image Retrieval, Image Annotation

ABSTRACT

In order to maximise the benefit of the colossal repository of digital images available both publicly and in private collections, intelligent matchmaking tools are required. Unfortunately, most image search engines rely on free-text search that often returns inaccurate sets of results based on the recurrence of the search keywords in the text accompanying the images. In this paper we present a semantically-enabled image annotation and retrieval engine that relies on methodically structured ontologies for image annotation, thus allowing for more intelligent reasoning about the image content and subsequently obtaining a more accurate set of results and a richer set of alternatives matchmaking the original query.

1. INTRODUCTION

Affordable access to digital technology and advances in Internet communications have contributed to the unprecedented growth of digital media repositories (audio, images, and video) over the past few years. Retrieving relevant media from these seemingly ever-increasing repositories is an impossible task for the user without the aid of search tools. Whether we are considering public media repositories such as Google™ images and YouTube™ [1] or commercial photo-libraries such as Empics™ [2], some kind of search engine is required to matchmake the user-query and the available media. This research effort focusses on image/photo retrieval techniques.

Most public image retrieval engines rely on analysing the text accompanying the image to matchmake it with the user query. Various optimisations were developed including the use of weighting systems where higher regard can be given to the proximity of the keyword to the image location, or where more frequently images are prioritised, or advanced text analysis techniques that use term weighting method, which relies on the proximity between the anchor to an image and each word in an HTML file [3].

Similar relevance-analysis and query expansion techniques [4] are used in annotation-enriched image collections, where usually a labour-intensive annotation process is utilised to describe the images with or without the aid of some domain-specific schema [5].

Despite the optimisation efforts, these search techniques remain hampered by the fact that they rely on free-text search that, while cost-effective to perform, can return irrelevant results as it primarily relies on the recurrence of exact words in the text accompanying the image. The inaccuracy of the results increases with the complexity of the query. For instance, using the Yahoo™ search engine to look for images of the football player Zico returns some good pictures of the player, mixed with photos of cute dogs (as apparently Zico is also a popular name for pet dogs), but if we add the action of scoring to the search text, this seems to completely confuse the Yahoo search engine and only one picture of Zico is returned, in which he is standing still!

Any significant contribution to the accuracy of matchmaking results can be achieved only if the search engine can "comprehend" the meaning of the data that describes the stored images, for instance, if the search engine can understand that scoring is an act associated with sport activities performed by humans. Semantic annotation techniques have gained wide popularity in associating plain data with "structured" concepts that software programs can reason about [6]. The aim of this research is to contribute towards the utilisation of semantic web technologies to improve the processes of image annotation and retrieval.

The paper begins with an overview of the Semantic web technologies. In section 3 we review the case study that was the motivation for this work. Sections 4, 5, and 6 detail the implementation roadmap of our semantic-based retrieval system, i.e. ontology engineering, annotation, and the retrieval engine. We discuss plans for further work and conclude in chapters 7 and 8 respectively.

2. OVERVIEW OF THE SEMANTIC WEB

Introduction

The fundamental premise of the semantic web is to extend the Web's current human-oriented interface to a format that is comprehensible to software programmes. For instance, in a future scenario of the Semantic Web, intelligent agents should be able to set up an appointment between a patient and the doctor, looking at both timetables, and then finding the best way to the clinic without the patient having to interfere in the

process. Hence, the user would only have to specify the appointment requirements and the semantic agent will complete the task on its own [7].

Ontology support

The concept of ontologies is fundamental to the Semantic Web. According to the Collins dictionary, ontology is "the branch of philosophy that deals with the nature of existence". Considering a domain (e.g. science, sports, etc.), "its ontology forms the heart of any system of knowledge representation for that domain. Without ontologies, or the conceptualisations that underlie knowledge, there cannot be a *common* vocabulary for representing and sharing knowledge."

From a computing science point of view, an ontology represents an area of knowledge that is used by people, databases, and applications that need to share domain information. Ontologies include computer-usable definitions of basic concepts in the domain and the relationships among them. [8].

The Ontology Working Language (OWL) [9] has become the de-facto standard for expressing ontologies. It adds extensive vocabulary to describe properties and classes and express relations between them (such as disjointness), cardinality (for example, "exactly one"), equality, richer typing of properties, and characteristics of properties (such as symmetry). OWL is designed for use by applications that need to process the content of information rather than just present information to humans.

Semantics and image retrieval

Applied to image retrieval, the semantic annotation of images creates a conceptual understanding of the domains that the image represents, enabling software agents, i.e. search engines, to make more intelligent decisions about the relevance of the image to a particular user query. For example, when searching the Google™ Image Search engine for some pictures of English football star David Beckham angry, it seems relevant to type the keywords 'David Beckham angry'. The search engine returns 14 results, within which only six represent a picture of David Beckham while only in two of them does he really look angry. The other retrieved images comprise many irrelevancies and have little to do with the initial query: one shows a moose, another one a picture of David Beckham's wife, or the front cover of American singer Eminem's book entitled Angry Blonde.

The use of the Semantic Web in image retrieval is likely to improve the computer's understanding of the image objects and their interactions. The goal is to make the machine understand that David Beckham is a person, and that he is also an English footballer playing Real Madrid FC. He also used to play for Manchester United and for the England National Team. Because Beckham is understood as a person, he is thus likely to express emotions. For this query the emotion is rather strong and negative.

The ontology relating David Beckham to human emotions should be able to retrieve all the pictures where David Beckham appears to be expressing a rather negative strong emotion, such as angry, furious, frustrated or even disappointed. To attain such expanded results, the data needs a better structure, so as to make sense for a machine that feelings are attached to people and can be either positive or negative. Here, the Semantic Web is likely to bring such a structure that integrates concepts and inter-entity relations from different domains, such as Person, Sport, and Media in relation to the query above.

3. CASE STUDY FOR SEMANTIC-BASED IMAGE RETRIEVAL

An opportunity to experiment with our research findings in semantic-based search technology was gratefully provided by Empics™. Empics is a Nottingham-based company which is part of the Press Association Photo Group Company [2]. As well as owning a huge image database in excess of 4 million annotated images which date back to the early 1900's, the company processes a colossal amount of images each day from varying events ranging from sport to politics and entertainment. The company also receives annotated images from a number of partners that rely on a different photo indexing schema.

More significantly, initial investigation has proven that the accuracy of the results sets matching the user queries do not measure up to the rich repository of photos in the company's library.

The goal of the case study is two-fold. Initially, we intend to investigate the use of semantic technology to build a classification and indexing system that critically unifies the annotation infrastructure for all the sources of incoming stream of photos. Subsequently, we'll conduct a feasibility study aiming to improve the end-user experience of their images search engine. At the moment Empics' search engine relies on Free-Text search to return a set of images matching the user requests. Therefore the returned results often go off-tangent if the search keywords do not exactly recur in the photo annotations. A significant improvement can result from semantically enabling the photo search engine. Semantic-based image search will ultimately enable the search engine software to understand the "concept" or "meaning" of the user request and hence return more accurate results (images) and a richer set of alternatives.

We decided to focus on sports events as this is currently the main domain of the company's business.

4. ONTOLOGY DEVELOPMENT

Domain Analysis

Our domain analysis started from an advanced point as we had access to the photo agency's current classification system. However, we felt that we need to compile an integrative structure that combines the image attributes and the domain specifications. Hence, two main ontology classes were developed:

- Classes to localise the event in the picture (stadium, area, etc.),

- Classes to characterise the picture itself (frame, size, creation date, etc.).

As this project uses data related to sports, many other classes are also required to describe it, such as Player, Team, Match, and Tournament. We also added action and emotion classes to enable expressing relationships between objects in the image. A subset of the ontology tree is shown in Figure 1.
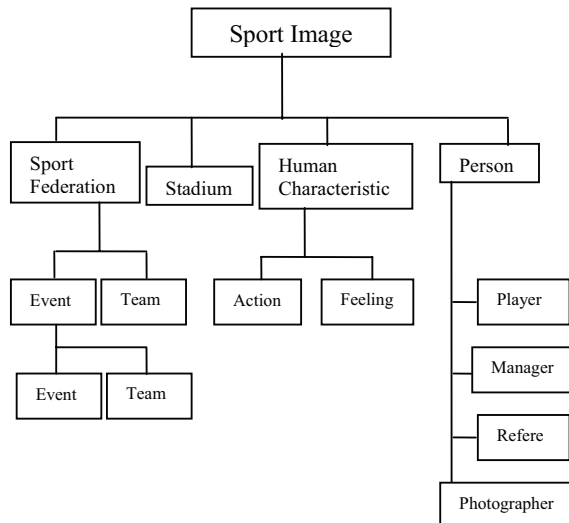

Figure 1: Subset of the ontology tree

Consistency Checking

The semantics of OWL uses properties to build relationships between classes. Following the rational design techniques for data structures, it is difficult to avoid inconsistencies in Ontology development [10]. For instance, when designing relationships for players performing in a particular event, the following is a commonly taken approach:
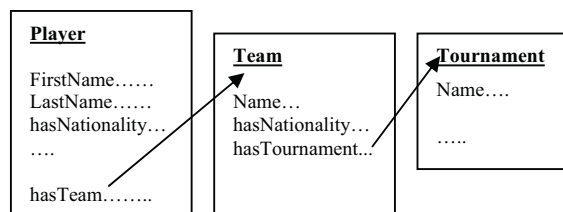

Figure 2: Example of class relationships in an ontology

In the example above, a player plays for a team which takes part in a tournament. However logical this description appears at first sight, further analysis reveals a serious inconsistency problem. When a player plays for two different teams at the same time (e.g. his club and his national team) or changes clubs every year, it is almost impossible to determine which team he plays for. When reusing such player as an individual, this one is likely to play for the wrong team. The same problem occurs for a team which can take part in different tournaments, like FA Premier League and UEFA Champions' League for a British football team.

A better approach to the classification is to start with the tournament, which solves the problem above by storing the information only within the player class.
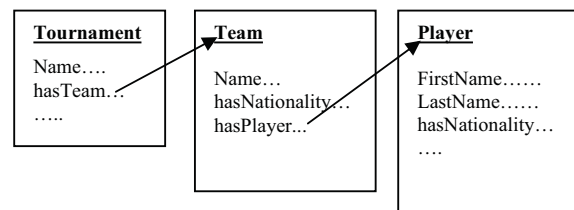

Figure 3: Re-organization of the player classification

All semantic models use two types of properties to build relationships between individuals (classes), Datatype properties, and Object properties. When assigning properties to a class, all its sub-classes inherit their parent class properties.

Deciding on the appropriate type of property to use is not a trivial task. Every use of these properties has to be done thoughtfully, whereas object properties link individuals of different classes together, datatype properties can just point to immediate values (e.g. text strings), which are meaningless to a reasoning software, except for performing a string-based search. For instance, allocating datatype properties to the person class in order to give each new instance a first name, a last name is a correct use of datatype properties, because they cannot be reused by another individual. On the other hand, object properties are required to assign someone a nationality. Indeed, a country is more than a mere string. A country can have properties such as a currency, a capital city, many towns and villages, a language, a national flag and anthem, etc. A country needs thus to be an instance. Furthermore, such useful classes might be defined in already existing ontologies which enables their reusability.

Coverage

Although consistent, the structural solution in Figure 3 is incomplete as every year a team will have a range of players leaving it either to retire or to play for another team. The same problem occurs with tournaments as from one year to another, teams taking part in the tournament change. This problem can be easily solved by adding a start and end date for the tournament, rather than by engineering more complex object property solutions. Hence, as far as the semantic reasoner is concerned, the "FIFA World Cup 2004" is a different instance from "FIFA World Cup 2008". The same reasoning can be applied to the class team, as player can change team every season, or even sometimes twice in a season. These indications, although basic for a human reasoning, need to be explicitly defined in the ontology.

Furthermore, in order to increase the automation of available data, inverse properties are used. Therefore, there is no privileged way to reason; all the properties are added in a dynamic manner. The example in Figure 4 presents this stage of the ontology.
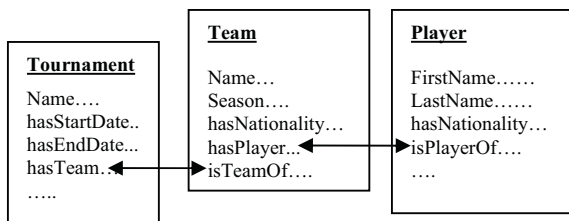
Figure 4: Resolving Coverage problems in ontology

Object Properties such as hasPlayer can have inverse property isPlayerOf. Such properties reveal the power of the Semantic Web where everything may be reused and easily spread. Such inverse properties automatically state that "if Team A hasPlayer Player B, then Player B isPlayerOf Team A". A high degree of understanding is given to the computer along with automation.

Normalisation: reducing the redundancy

Adopting the latest evolution of our ontology in Figure 4 above will result in creating new tournaments every year, which is rather inefficient. No matter which players are playing for the team, or which tournament a team takes part into, a team is a non-temporal class. That is, Arsenal Football Club, Manchester United or the Glasgow Rangers need to remain abstract entities. Even though the managers change, the players come and go, the Club entity that gathers players, managers and fans is nothing but temporal. A higher degree of abstraction is required here.

The underlying idea consists in the fact that a player is a member of a team, which is a member of a particular tournament. A new class is thus required: a membership class. This is the indispensable link between teams and players, as well as between teams and tournaments as illustrated in Figure 5.
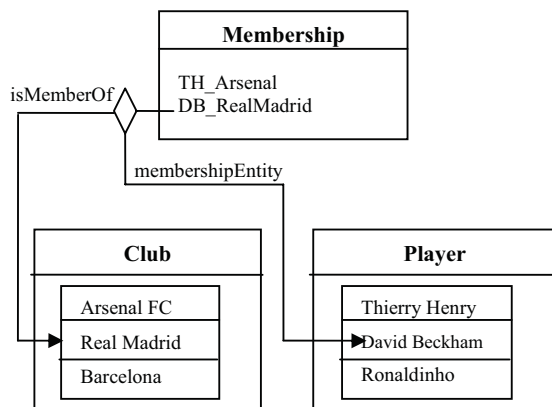

Figure 5: Membership class in the final ontology

The instances from the Membership class link together instances from two different classes. In the figure above, the statement "David Beckham plays for Real Madrid" is translated into instance DB_Real_Madrid. Owing to the properties of the Membership class, all its individuals possess the following properties:

- memberEntity, which links to a person (Player, Manager, Supporter, Photographer, etc.)

- isMemberOf, which refers to the organisation (Club, Press Association, Company, etc.)
- fromPeriod, which depicts the day when the person was hired by the organisation
- toPeriod, which states the day when the person leaves the organisation.

Thus, such a class added with these properties eventually solves all the problems encountered during the building of this ontology. The club remains a static entity, as well as player or manager. Moreover, this relational description allows retracing the career of a person, having just to describe pictures of him. While this example is built upon sports it is clear that is expandable to every domain of the society. Describing images of anyone at different stages or ages can thus help retracing his career. Through this example, the power of the Semantic Web is really tangible.

5. IMAGE ANNOTATION

The developed ontology represents the shared vocabulary to be used for describing any sports image. The annotation stage requires not only precise understanding of the ontology, but also an end-user focussed approach that methodically considers the dynamics of the subsequent retrieval process. Reasoning about the annotation requires the utilisation of an ontology editor. In this project, we intended to use Jena [11], an ontology editor that is available as a Java package API enabling the user to work with OWL files under Eclipse [12].

The method used to store data is very similar to the structure of a database, using several tables to store and organise the data. In this project, the central part of the annotation is the object of the picture, as it is the only link between the image and its content. Thus, particular care needs to be given to the image library as illustrated in Figure 6.

In the schema in Figure 6, the central part is the image library. Each image possesses an object, whose main features are stored within an object library, distinct from the image library. By using this means of storage, every object created can be reused if needed. Thus, its features are entered just once, reducing redundancy and enhancing the information provided to the annotation.
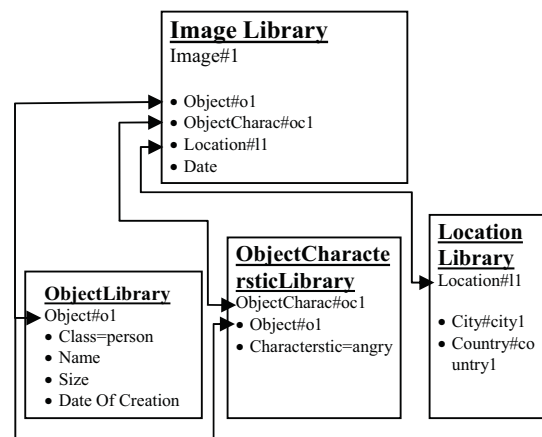

Figure 6: Architecture of the annotation

For example, when annotating a new image where David Beckham appears, the user only needs to specify the URI (Uniform Resource Identifier) referring to the description of the player. However, the description will implicitly make the computer understand that in the picture, there is a person, who is a player, who plays football for Real Madrid, and who is an English citizen. All this data is implicitly given to the annotation by using this architecture.

The main philosophy behind building relationships in the annotation is built around the sentence structure [13]: Actor – Action – Object, for instance, Beckham – Smiles – null, or Zidane – Receives – Red_Card, or Gerrard – Tackles – Henry.

6. IMAGE RETRIEVAL

The Image Retrieval Algorithm uses a tree comparison that traverses the ontology classes in order to find a path that represents the "nearest neighbour" to the query. The user is also allowed to assign preferences to different segments of the query, for instance priority might be given to the player or the tournament in the specific query. In order to allow a dialog between the query and the annotated files, a semantic description generator needs to be created. Figure 7 gives a high level view of the annotation mechanism.
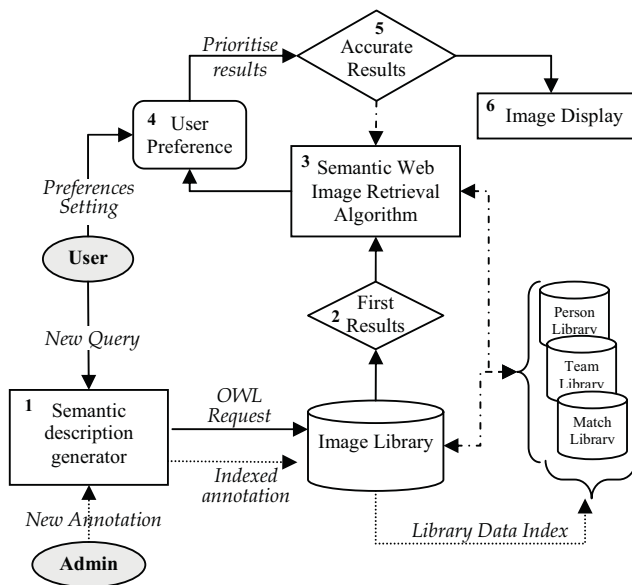

Figure 7: Schematic diagram of the Semantic Web Image Retrieval software

The diagram reveals that user and administrator play two different roles. First the administrator annotates the images in order to give more consistency and structure to the annotation libraries. Thus, the user can only create new requests. The semantic description generator (1) transforms his query into an OWL query, which determines a first set of images (2) on which the Semantic Web Image Retrieval Algorithm can be

applied (3). During this stage, a dialog between all the libraries is required. This algorithm uses a weightings system (4) previously set by the user in order to fine-tune the final results in accordance to the user queries (5). The last stage consists in displaying the results (6).

The nearest neighbour matchmaking algorithm continues traversing back to the upper class of the ontology and matching instances until there are no super classes in the class hierarchy, i.e. the leaf node for the tree is reached, giving degree of match equal to 0. The degree of match (DoM) degree is calculated according to the following equation:

Where the MN is the total number of matching nodes in the selected traversal path, and GN the total

$$DoM = \frac{MN}{GN}$$

number of nodes in the selected traversal path. This is exemplified in Figure 8. Then the comparison values are weighted [14] using the user preferences according to the formula:

$$m = |l_r - l_a| \; ; \quad \forall p \in [0,1[\; , \quad v = p^m$$

v: value assigned to the comparison,
m: matching level of the individuals,
p: user preference setting,
l_r: level of the request,
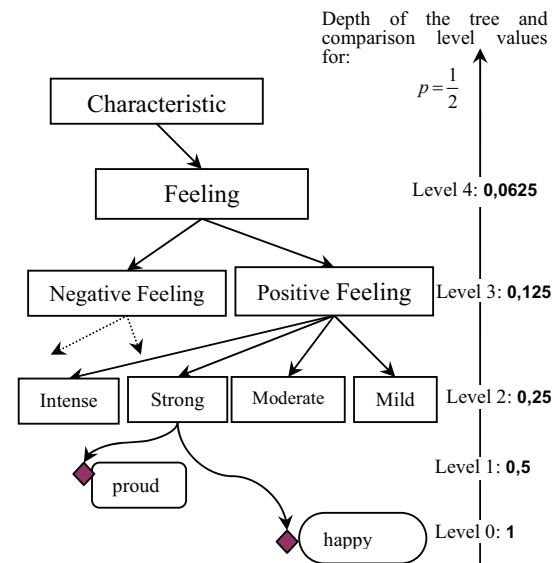l_a: level of the annotation.

Figure 8: Traversing the Ontology Tree

For example, in figure 8, comparing the hasCharacterstic object property values for the new problem request and the retrieved cases, which matches till level 0, will give degree of match(DoM) =(1/1)=1. While cases matching with problem case at the level 1 will result in the (DoM) = (1/2) =0.5 and so on. If there are no matches found while traversing the ontology tree until the parent node is reached, then DoM will be 0.

Thus, when two individuals match at level 0, it means that they are identical. The comparison method will return the value v = 1 for this comparison, as it represents an exact matching. If the two instances compared belong to the same class, the method will return 0.5

7. FUTURE WORK

Our immediate plans involve implementing a web-based, user-transparent annotation tool and developing a web-service based API for the retrieval engine. This will allow us to conduct an exhaustive evaluation study comparing the performance of our engine to the company's current indexing and retrieval tools.

We also intend to look at benefiting from advances in content-based techniques to aid in semi-automating the labour-intensive annotation process and boost the accuracy of the search engine. Content-based approaches are designed to extract useful image information directly from the raw images by deriving image features that describe the image content in terms of colour, texture, shape, etc. Image features can either be computed globally (i.e. for the whole image) or locally (i.e. for an image region which ideally represents an object in the scene). Content-based image retrieval based on an underlying semantic model (see e.g. [15]) currently represents an area of intensive research although at the moment still in its infancy. Enriching our semantics-based retrieval approach presented in this paper with content-based techniques, for example based on learning strategies that correlate certain image features with certain parts of the ontology will hence provide even more accurate retrieval results and can also be employed to avoid ambiguities as well as to spot cases of incorrect annotations.

8. CONCLUSIONS

In this paper we presented a detailed road-map for re-engineering image search engines to take full advantage of advances in semantic web technologies.

The first stage of the development was producing ontologies (domain vocabularies) that provide the search engine with a consistent view of the domain entities and possible relations between them. Annotations based on our ontology are expected to have little redundancies as we took great care in normalising our ontology structure by applying pragmatic non-semantic techniques were required.

Our approach to engineering the annotation process was user-focused and took into account the dynamics of the retrieval process. It was based on a *sentence-based* template suitable for modelling user-queries that can involve complex relationships between the query keywords.

The retrieval algorithm is based on a variation of the nearest-neighbour search technique for traversing the ontology tree and can accommodate complex, relationship-driven user queries. The algorithm also

provides for using pre-defined weightings to qualify the search result in accordance to user preferences.

REFERENCES

[1] Ka-Ping Yee *et* al. 2003. Facated meta data for image search and browsing. *Proceedings of the SIGCHI conference on Human factors in computing systems.* 401 - 408

[2] PA Photos. 2007. http://www.paphotos.com/

[3] FUJII A., ISHIKAWA T. 2005. Toward the Automatic Compilation of Multimedia Encyclopaedias: Association Images with Term Descriptions on the Web. *In Proceedings of the 2005 International Conference on Web Intelligence – WIC05.* Compiègne, France. September 19-22, 2005. 536-542.

[4] J. Jeon *et al*. 2003. Automatic image annotation and retrieval using cross-media relevance models. *Proceedings of the 26th annual international ACM SIGIR conference on Research and development in information retrieval.* Toronto, Canada.

[5] Hare JS *et al*. 2006. Mind the gap: another look at the problem of the semantic gap in image retrieval. *Multimedia Content Analysis, Management, and Retrieval 2006, Vol. 6073, No. 1. (2006).*

[6] Huan Wang *et al*. 2006. Does ontology help in image retrieval?: a comparison between keyword, text ontology and multi-modality ontology approaches. *Proceedings of the 14th annual ACM international conference on Multimedia.* 109 - 112

[7] BERNERS-LEE, Tim. 2000. Weaving the Web: the original design of the World Wide Web by its inventor / Tim Berners-Lee with Mark Fischetti. Harper Collins. 2000. pp 157-160.

[8] W3C Corporation [online]. 2004: OWL Web Ontology Language Overview. http://www.w3.org/TR/owl-features

[9] Parsia, B., and Sirin, E. 2004. Pellet: An OWL DL Reasoner. *In Third International Semantic Web Conference (ISWC2004)*, Hiroshima, Japan.

[10] Rector A. 2003. Modularisation of domain ontologies implemented in description logics and related formalisms including OWL. *Proceedings of the 2nd international conference on Knowledge capture.* 121 - 128

[11] Jeremy J. Carroll *et al*. 2004. Jena: implementing the semantic web recommendations. *Proceedings of the 13th international World Wide Web conference.* New York, USA. 74-83.

[12] Gamma E, Beck K. 2003. Contributing to Eclipse: Principles, Patterns, and Plug-Ins. Published by Addison Wesley Professional. ISBN-10: 0-321-20575-8;

[13] L. Hollink *et al*. 2003. Semantic annotation of image collections. *In Workshop on Knowledge Markup and Semantic Annotation, KCAP'03.*

[14] Osman T, Thakker D, Al-Dabass D. 2006. Semantic-Driven Matchmaking of Web Services Using Case-Based Reasoning, IEEE International Conference on Web Services (ICWS'06), September 2006. pp. 29-36.

[15] J. Li and J.Z. Wang. 2003. "Automatic Linguistic Indexing of Pictures by a Statistical Modeling Approach", IEEE Trans. Pattern Analysis and Machine Intelligence, Vol. 25, No. 9, pp. 1075-1088.