Computers in Human Behavior 26 (2010) 205–209

Contents lists available at ScienceDirect

Computers in Human Behavior

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

WESONet: Applying semantic web technologies and collaborative tagging to multimedia web information systems

Jose Emilio Labra Gayo [a,*], Patricia Ordóñez de Pablos [b], Juan Manuel Cueva Lovelle [a]

[a] Department of Computer Science, University of Oviedo, C/Calvo Sotelo, S/N 33007 Oviedo, Spain
[b] Department of Business Administration, Avda. El Cristo, S/N 33071 Oviedo, Spain

ARTICLE INFO

Article history:
Available online 7 November 2009

Keywords:
Multimedia web information systems
Semantic web technologies
Collaborative tagging

ABSTRACT

The publication of different media types, like images, audio and video in the World Wide Web is getting more importance each day. However, searching and locating content in multimedia sites is challenging. In this paper, we propose a platform for the development of multimedia web information systems. Our approach is based on the combination between semantic web technologies and collaborative tagging. Producers can add meta-data to multimedia content associating it with different domain-specific ontologies. At the same time, users can tag the content in a collaborative way. The proposed system uses a search engine that combines both kinds of meta-data to locate the desired content. It will also provide browsing capabilities through the ontology concepts and the developed tags.

1. Introduction

The increase of information of different media types available in the world wide web is overwhelming. The production of digital media content is no longer a difficult issue and it is expected that more and more people will publish multimedia content on the web. Although it is easy to generate and publish this kind of data, it is very difficult to search and index it.

Automatically extracted descriptions like the dominant color, texture, etc. tend to be low-level and there is a need for semantic descriptions that are closer to the knowledge domain of the users. This semantic gap has already been acknowledged (Mich, Brunelli, & Modena, 1999). We consider that current, state-of-the-art multimedia analysis technologies are not yet able to produce the higher-level descriptions that are needed to facilitate multimedia retrieval at the user level. To solve that problem, it is necessary to manually add the descriptions. The people who could add these descriptions can be either the creators or the users of the system.

In production time, the creator usually knows some information about the video. For example if a person is recording a conference talk he may know who the speaker, the title and the chairman are. That data can be stored in a database which contains the information in a non-uniform way. One problem of using non-standard databases is that once the video is published on the web, the meta-data is lost. A better solution would be to embed that meta-data in the same video that is being produced. In this way, the meta-data is maintained with the video and it does not depend on a particular system or platform. MPEG-7 was proposed as a vocabulary that could embed multimedia descriptions (Salembier & Smith, 2002). In this paper we will use that standard as a means to publish videos with their corresponding meta-data.

Images and videos can have different meanings depending on the context and the users who view them (Aurnhammer, Hanappe, & Steels, 2006). For that reason, it is not feasible to let the creators alone to describe the multimedia asset. It was long assumed that manual annotations would not be provided by end-users because it is a tedious and time consuming process. However, recent developments like Flickr[1], or YouTube[2] have shown that the users can add descriptions to resources in the form of a freely chosen set of keywords ("tags") or comments.

In this paper, we present an approach to develop multimedia web information systems which combine automatically obtained low-level descriptions, with higher-level descriptions obtained from the creators which link to ontology concepts and tags added collaboratively by the end-users.

The structure of the paper is as follows. We begin with a review of related work in Section 2. In Section 3 we describe the main semantic web technologies and their relationship with social networks. In Section 4 we propose an approach to add meta-data to multimedia systems. Section 5 describes a combination between ontologies and collaborative tagging. We also describe a hybrid search engine over these systems. Finally, Section 6 describes the

* Corresponding author. Tel.: +34 985106206.
E-mail addresses: labra@uniovi.es, patriciaordonezdepablos@yahoo.com (J.E. Labra Gayo), patriop@uniovi.es (P.O. de Pablos), cueva@uniovi.es (J.M. Cueva Lovelle).

[1] http://www.flickr.com.
[2] http://www.youtube.com.

0747-5632/$ - see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.chb.2009.10.004

architecture of the system that we are developing, which implements the proposed approach.

2. Related work

There have been numerous attempts to add semantic descriptions to multimedia files (Stamou & Kollias, 2005, Moenne-Loccoz, Janvier, Marchand-Maillet, & Bruno, 2005). One of the first attempts to use MPEG-7 to store meta-data was made by Hunter (2001). It initially developed an RDFS ontology enriched with DAML-OIL constructs, which was subsequently translated to OWL. In the aceMedia project (Bloehdorn et al., 2005), a Visual Descriptor Ontology was developed for describing low-level audio–visual features which were tightly coupled with the MPEG-7 Visual Part.

Another approach taken in Tsinaraki, Polydoros, Kazasis, and Christodoulakis (2005) was to translate the Multimedia Description Schemes of MPEG-7 into OWL. The obtained ontology serves as a core ontology for the integration of domain specific knowledge into the MPEG-7 semantic data. This approach was also developed in García and Celma (2006) using an automatic conversion tool from XML Schemas to OWL. Hollink and Worring (2005) developed a visual ontology linking MPEG-7 low-level concepts with WordNet concepts.

On the other side, web sites like Amazon (Linden, Smith, & York, 2003) developed collaborative filtering algorithms which take into account the similarity between users in order to offer item recommendations. Other web sites started to allow their users to add tags to different types of resources. This approach has recently acquired a lot of popularity as a categorization system (Cattuto, 2006). Marlow, Naaman, and Boyd (2006) describe a taxonomy of tagging systems and apply their model to the photo-sharing site Flickr. Aurnhammer et al. (2006) propose a combination between collaborative tagging and image retrieval and classification. It introduces the navigation map concept that links between users, tags and data elements. It shows that similarity search based on image features yields additional links on this map. The system also works with Flickr and is able to separate different meanings of the same tag using visual features.

Another line of research is the application of social networks tools to improve information retrieval taking into account provenance of descriptions and trust. Halaschek-Wiener et al. (2005) describe an approach to include semantic web descriptions in images and to link them to ontology concepts. Their system allows the users to navigate through different conceptual axis. Recently, there have been several proposals that use FOAF[3] or extensions of FOAF. Celma (2006) presents a system that recommends music based on FOAF profiles. In the audio–visual field, Golbeck and Hendler (2006) proposed an algorithm to create movie recommendations based on trust.

Finally, in the search process, our approach is based on the hybrid approach proposed by Rocha, Schwabe, and de Aragão (2004) that we already applied to develop a search engine over public administration legal documents (Berrueta, Labra, & Polo, 2006). The main contribution of this paper is the adaptation of that approach to multimedia collections where we have added a collaborative tagging system.

3. Semantic web technologies and social networks

Social networks are representations of the relationships between groups and individuals in a community. Analysis of social ties and social networks is an established field and has found numerous applications in several fields, like sociology, psychology and computer science (Wasserman & Faust, 1994).

Internet has enabled the creation of on-line virtual communities. The technological infrastructure of those communities facilitates the development of systems which can analyze the structure of the networks. Apart from analyzing that structure, it is possible to create tools which allow the users to access and manage the knowledge of those communities.

The semantic web has emerged as a vision of the future of the web where it is promoted to publish data on the web that can automatically be manipulated. Currently, the information that is published in the web is usually thought for human consumption. Browsers are mainly interpreters which recognize the HTML markup and visualize its content. But the most difficult part, content interpretation, selection and management, must be done by human beings which are able to understand natural language, or in the case of multimedia, the contents that appear in the images and videos.

In order to develop the semantic web vision, it was necessary to create a technology to describe the data that appears inside web pages, breaking the small unit of a web page and having more fine granularity. One of the proposed vocabularies has been RDF, which enables the description of resources and the relationships between them. In this way, it is possible to make descriptions about data in a greater granularity, enabling an automatic processing of them. The main point of RDF is that the relationships between resources are globally identified by URIs. In this way, it is possible to establish standard relationships. For example, in order to declare that a person has a name, it is possible to use the property `http://xmlns.com/foaf/0.1/name` which is a standard property defined in the FOAF specification. Using those standard properties, it is possible, for example, to collect all people names in a given resource.

For example, the following RDF fragment can be used to describe that a given resource contains an actor whose name is `Steven Pemberton`, which has the role of `Speaker` and that the title of the resource is `The power of declarative thinking`.

```
<rdf:RDF ...
   xmlns:foaf="http://xmlns.com/foaf/0.1/"
   xmlns:weso="http://purl.oclc.org/NET/weso/">
<rdf:Description rdf:about="">
   <weso:contains rdf:parseType="Resource">
      <foaf:Person>
         <foaf:name>Steven Pemberton</foaf:name>
         <foaf:mbox_sha1sum
            xml:lang="">811408...</foaf:mbox_sha1sum>
      </foaf:Person>
      <weso:role
   rdf:resource="http://purl.oclc.org/NET/weso/speaker"/>
      </weso:contains>
      <weso:title>The power of declarative thinking</weso:
      title>
   </rdf:Description>
</rdf:RDF>
```

In the FOAF vocabulary, the property `mbox_sha1sum` is used to encapsulate the email of a person. That property can be used to have a unique identifier of people without giving access to its email to prevent spam.

Another key technology in the development of the semantic web is the possibility to define concepts and relationships between them. To that end RDF Schema and afterwards OWL, were developed. OWL has been developed with the goal to provide a common language for ontology definition and has acquired some popularity in the knowledge representation field. It is mainly based on the

[3] http://www.foaf-project.org.

description logic formalism, a subset of first order predicate logic. The development of ontologies in a decentralized way, will facilitate the integration of data from different sources and different kinds of applications and there are already a number of ontologies for different purposes and domains.

Following the example, an ontology could declare that if a resource contains something which has the role of `Speaker`, then that resource is of type `Talk` and that a `Talk` is a subclass of a `Conference`. So an inference system could infer that the above resource is describing a `Conference`. This information can be described in description logics terminology as:

$$\exists contains(\exists role\ \{Speaker\}) \subseteq Talk$$
$$Talk \subseteq Conference$$

4. Adding semantics to multimedia data

The semantic web approach breaks the concept of a web page as a unit of information enabling the creation of resource descriptions with finer granularity. For example, instead of the homepage of a person, it would be possible to refer to the phone number of that person.

In the case of textual information, that break-up is more affordable, although not easy, given that it is possible to access to paragraphs, words, etc. facilitating syntactic searches from keywords. In this way, traditional search engines do a reasonable good job to access text based information.

However, in the case of multimedia information, the minimum unit is much more difficult to divide. The availability of tools that analyze the contents of an image or video to automatically obtain its content is scarce. It is necessary to provide other means to add semantics to these kind of resources.

MPEG-7 has been designed to provide detailed formatting information and fine-grained descriptions of the structural, and low-level audio–visual features of multimedia content. It was divided into four components: the Description Definition Language (the basic building blocks of MPEG-7), Audio (the descriptive elements for audio), Visual (those for video) and Multimedia Description Schemes (the descriptors for capturing the semantic aspects of contents, e.g. places, objects, events, etc.).

In our approach, we propose the use of the MPEG-7 standard and to add three levels of meta-data to a multimedia resource (see Fig. 1):

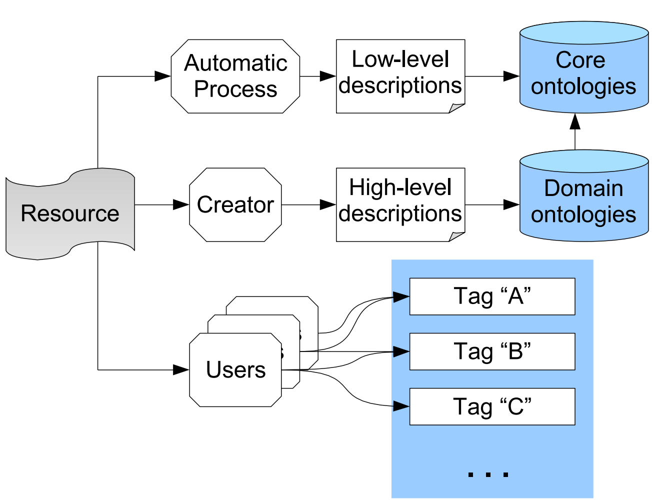

Fig. 1. Describing multimedia resources.

- Low-level descriptions, which can be automatically generated by the authoring tool. Following García and Celma (2006), our system converts those descriptions into RDF descriptions that can be combined with the core ontology that is obtained from the MPEG-7 specification.
- High-level descriptions, which are added by the creators of the video. The tool allows the creator to select parts of a multimedia file and to link them to instances of concepts in several domain-specific ontologies. Although we are developing some ontologies in our domain (conferences and courses) we are trying to reuse several existing ontologies like DOLCE[4], Dublin Core[5], FOAF, etc. In our ontologies, each concept has an associated synset with terms that can refer to it.
- Finally, our system allows the end-users to collaboratively add tags to the resources. Although the tags are free keywords, our tool uses the approach described in Xu, Fu, Mao, and Su (2006) to suggest tags to the users, so the set of tags tends to converge. This feature is important, given that the search tool tries to find resources using the tags associated with them and the similarity between those tags and the synsets defined in the ontologies.

5. Combining ontologies and collaborative tagging

The development of collaborative tagging sites has attracted a huge number of users. This kind of social software allows the users to tag a given resource with some text. Some examples of sites with collaborative tagging are Flickr, del.icio.us, Technorati, CiteULike, Buzznet, etc. They allow to share photos, URLs, blogs, article references and music titles. A tag is a free text string chosen by the user. Usually, these tools offer some recommendations about similar tags given to the same resource so the tags tend to be similar.

Cattuto (2006) states that collaborative tagging is different to ontology based approaches, where a group of experts develops an ontology and the resources are linked to that ontology. In this case, the approach is bottom-up, because the users are the ones who develop their own tags which group the information. From a global point of view, the set of tags, although it is developed without explicit coordination, evolves along the time and carries certain terminological patterns which are shared by the user community. In this way, one can observe the appearance of a categorization system – commonly referred as folksonomies – that can be used to navigate through the large collection of possibly heterogeneus resources (see Fig. 2).

As an example, the following table shows a possible set of tags associated with the example resource.

Tag	Users
"XHTML 2.0"	U2, U3
"XFORMS"	U1, U3
"Pemberton"	U1
"Tutorial"	U3

Our search algorithm combines the information stored in the collaborative tag database with the ontology based meta-data associated with each resource. To that end, we will extend the architecture of the system that we have described in Berrueta et al. (2006). In that system, given a query α we created a semantically enriched query β using the domain ontologies. The new syntactic query was used to search through a large set of legal documents. In the case of multimedia, we will search over the tags and the meta-data associated with them.

[4] `http://dolce.semanticweb.org/`.
[5] `http://dublincore.org/`.

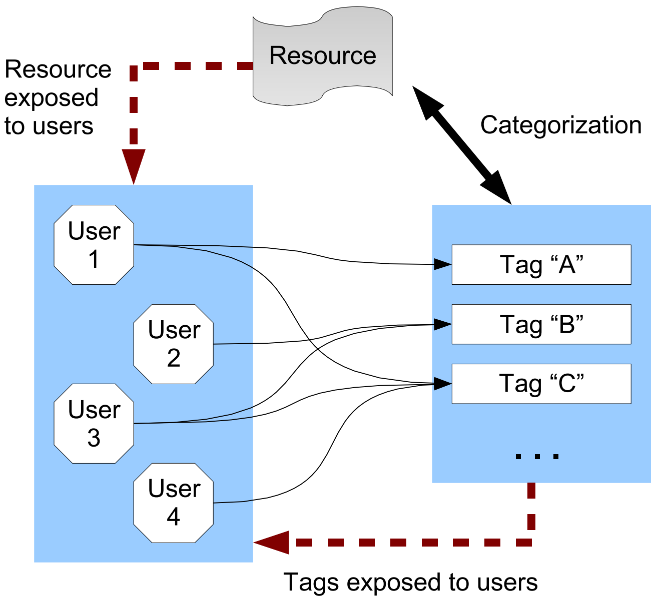

Fig. 2. Collaborative tagging.

The main steps of the algorithm are:

- Let α be the initial query formed by a set of words $\{w_1, \ldots, w_n\}$. The first step is to obtain a normalized set of terms $\tau = \{t_1, \ldots, t_n\}$ by removing plurals, death words, verb forms, etc.
- The next step is to obtain the ontology concepts Ω whose synset match with the terms in τ.
- Next, we apply spreading activation (Crestani, 1997) to Ω to obtain a new set Ω' of ontology concepts related with the initial ones. Each concept c_i contains a weight w_i which indicates the strength of its relationship with the concepts in Ω.
- We create a new syntactic query β formed by the synsets of the concepts in Ω' with their associated weights. We also include the terms t_j for which there was no related concept in order to take into account the information in the collaborative tags that has not been included in the ontologies.

The new semantically enriched query β is executed over the collaborative tags and the synsets that appear in the meta-data of each resource. For efficiency reasons, in this phase of the search we are not taking into account the relationship between the user who makes the query and the users who tagged the resources. The above algorithm could be extended to accomodate a similarity measure between the profiles of the different users giving more weight to tags declared by more similar users.

As an example, lets suppose that the user typed the query α = "*Conference about XForms*" After the first, the normalized set could be τ = {"conference", "xform"}. In the second step, the system would obtain the set of concepts in the ontology that match with the terms in τ. In this case Ω = {#Conference}. In the next step, after applying spreading activation, the system would obtain the set of related concepts, for example Ω' = {(#Conference,1), (#Talk,0.7), (#Symposium,0.8)} Finally, the system would create a new syntactic query from the synsets associated to those concepts. In this case, the query could be β = ["Conference"$\wedge$1 "Talk"$\wedge$0.7 "Symposium"$\wedge$0.8 "xform"$\wedge$1].

Notice that the new query β would find the resource, because the creator has added information that enables the system to infer that it is a conference and some users have tagged it with the word "XForms".

6. WESONet project

The main goal of our project was the development of a video collection for our University. Most of the videos are about conferences and courses, although there are several videos about other subject matters. The number of videos available has increased in the last years given that there is a videoconference service provided by the University which publishes almost around 3 or 4 conferences each day. In the current system, it was quite difficult to find information given that there was not much meta-data added to the videos. We are developing an ontology in the domain of universities, courses and conferences, which will enable the hybrid search approach described above.

The architecture of the system is presented in Fig. 3. The main modules are:

- **DGen** (Description Generator tool): It allows the video producers to generate descriptions of the content. Although some of the descriptions are automatically generated by the default values, this tool also allows to select parts of a frame (using SVG figures: ellipses, rectangles and polyshapes) and temporal intervals and to bind them to ontology concepts.
- **MBrowser** (Mediateca Browser): This is the public web site accessible to users. It allows the users to browse the different multimedia contents in the repository. At the same time, it is possible to navigate through the ontology concepts and the tags stored by the users. In each phase of this navigation it is possible to search contents that match with the corresponding information.
- **MSearch** (Multimedia Search): This module implements the hybrid search algorithm described in last section. The search combines the collaborative tags with the ontology concepts and user profiles enabling a richer set of results.

In order to implement the above modules, we keep a repository with the different videos, their meta-data and user profiles. User profiles are stored in FOAF descriptions augmented with some information specific for our application. This approach facilitates future integration with other semantic web enabled tools.

The core ontology is a low level multimedia ontology which has been obtained from the MPEG-7 specification following the approach described by García and Celma (2006). The domain specific ontologies are being developed using the OWL language. At this moment, we have developed a some ontologies for conferences,

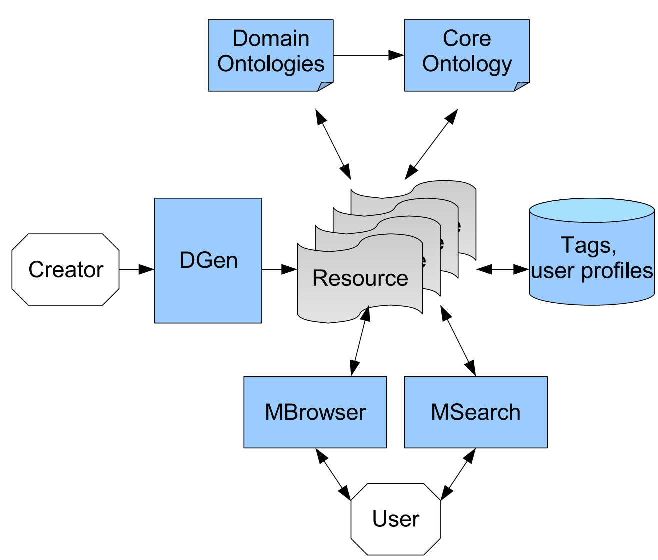

Fig. 3. Architecture of WESONet.

courses and Universities. We are planning to integrate our domain ontologies with other ontologies like DOLCE or WordNet.

7. Conclusions

The creation of image and video libraries is a difficult challenge. It is necessary to develop hybrid approaches which combine collective tagging with ontology based tagging for a better integration and user experience.

Although our system is being developed as a prototype in a specific domain – a collection of images and videos from our University – we consider that it could be applied to generic domains, like YouTube or similar web sites. We also believe that the use and production of digital media data will be increasingly popular and that there is a need of this kind of systems to facilitate the management of vast quantities of information.

With regards to future work, it is very important to obtain a testing framework for our search algorithms so we can assess the importance of the different parameters. The main challenge is how to take into account the end user, because the quality of the collaborative tagging can influence in the quality of the search procedures. Another line of future research is to study the scalability of our approach to larger collections and the use of more complex ontologies. Some work in this line has been developed by Pan, Qasem, and Heflin (2006).

Finally, our approach is our first attempt to take into account the similarity between the behaviour of users through their tagging process. Further research must be done to compare user profiles taking into account user reputation to avoid the effects of malintended users and to have more means to incentivate users in the collaborative tagging process.

References

Aurnhammer, M., Hanappe, P., & Steels, L. (2006). Augmenting navigation for collaborative tagging with emergent semantics. In I. Cruz, S. Decker, D. Allemang, C. Preist, D. Schwabe, & P. Mika, et al. (Eds.), *5th international semantic web conference, ISWC 2006, Athens, GA, USA. LNCS* (Vol. 4273, pp. 58–71). Springer.

Berrueta, D., Labra, J. E., & Polo, L. (2006). Searching over public administration legal documents using ontologies. In E. Tyugu & T. Yamaguchi (Eds.), *Joint conference on knowledge based software engineering* (pp. 167–175). IOS Press.

Bloehdorn, S., Petridis, K., Saathoff, C., Simou, N., Tzouvaras, V., Avrithis, Y., et al. (2005). Semantic annotation of images and videos for multimedia analysis. In *2nd European semantic web conference (ESWC 2005)*. Greece: Heraklion.

Cattuto, C. (2006). Semiotic dynamics in online social communities. *The European Physical Journal C, 46*, 33–37.

Celma, O. (2006). Foafing the music: Bridging the semantic gap in music recommendation. In *5th international semantic web conference*.

Crestani, F. (1997). Application of spreading activation techniques in information retrieval. *Artificial Intelligence Review*(11), 453–482.

García, R., Celma, O. (2006). Semantic integration and retrieval of multimedia metadata. In *5th knowledge markup and semantic annotation workshop* (Vol. 185, pp. 69–80). CEUR workshop proceedings.

Golbeck, J., Hendler, J. (2006). Filmtrust: Movie recommendations using trust in web-based social networks. In *IEEE consumer communications and networking conference*.

Halaschek-Wiener, C., Schain, A., Golbeck, J., Grove, M., Parsia, B., Hendler, J. (2005). A flexible approach for managing digital images on the semantic web. In *5th international workshop on knowledge markup and semantic annotation*. ISWC 2005, Galway, Ireland.

Hollink, L., & Worring, M. (2005). Building a visual ontology for video retrieval. In *MULTIMEDIA '05: Proceedings of the 13th annual ACM international conference on multimedia* (pp. 479–482). New York, NY, USA: ACM Press.

Hunter, J. (2001). Adding multimedia to the semantic web – building an mpeg-7 ontology. In *International semantic web working symposium, SWWS*.

Linden, G., Smith, B., & York, J. (2003). Amazon.com recommendations: Item-to-item collaborative filtering. *IEEE Internet Computing, 4*(1), 76–80.

Marlow, C., Naaman, M., & Boyd, D. (2006). Ht06, tagging paper, taxonomy, flickr, academic article, toread. In *Hypertext 2006*. New York: ACM Press.

Mich, O., Brunelli, R., & Modena, C. M. (1999). A survey on video indexing. *Journal of Visual Communications and Image Representation*(10), 78–112.

Moenne-Loccoz, N., Janvier, B., Marchand-Maillet, S., Bruno, E. (2005). An integrating framework for the management of video collections. In *First workshop on machine learning and multimodal interaction* (Vol. MLMIÔ4). Lecture notes in computer science 3361. Switzerland: Martigny.

Pan, Z., Qasem, A., Heflin, J. (2006). An investigation into the feasibility of the semantic web. In *Proceeding of the twenty first national conference on artificial intelligence* (AAAI 2006). Boston, USA.

Rocha, C., Schwabe, D., de Aragão, M. P. (2004). A hybrid approach for searching in the semantic web. In *WWW* (pp. 374–383).

Salembier, P., Smith, J. (2002). Introduction to MPEG-7: Multimedia content description interface. John Wiley & Sons. Ch. Overview of MPEG-7 multimedia description schemes and schema tools.

Stamou, G., & Kollias, S. (2005). *Multimedia content and semantic web: Standards, methods and tools*. John Wiley and Sons.

Tsinaraki, C., Polydoros, P., Kazasis, F., & Christodoulakis, S. (2005). Ontology-based semantic indexing for mpeg-7 and tv-anytime audiovisual content. *Multimedia Tools Applications, 26*(3), 299–325.

Wasserman, S., & Faust, K. (1994). *Social network analysis: Methods and applications*. Cambridge University Press.

Xu, Z., Fu, Y., Mao, J., Su, D. (2006). Towards the semantic web: Collaborative tag suggestions. In *Proceedings of workshop on collaborative Web tagging*. Edinburgh, Scotland.