Are we Ready to Embrace the Semantic Web?

Shankar Vembu and Stephan Baumann

German Research Center for Artificial Intelligence
Erwin-Schroedinger-Strasse 57, 67663 Kaiserslautern, Germany
{vembu, baumann}@dfki.uni-kl.de

1 Multimedia and semantic web

The aim of the semantic web[1] is to describe resources using metadata elements that can be processed or interpreted by machines. MPEG-7 [1] is the result of a standardisation effort to annotate multimedia documents. It offers a rich suite of metadata descriptors for describing these documents at various levels of abstraction from low level features to high level semantics. Owing to the proliferation of multimedia content in the internet, there is widespread interest in the semantic web community in multimedia metadata standards in general, and MPEG-7 in particular. Despite the fact that the semantic web initiatives could benefit a lot from MPEG-7 for the annotation of multimedia documents, recent studies have underlined the limitations of MPEG-7 in describing the semantics of highly structured domains like sports or medicine. One way to solve this problem is to adopt an integrated approach to the design of multimedia ontologies.

The first step in bridging the gap between multimedia and semantic web is to map the MPEG-7 specification into a semantic web compliant metadata language like RDF Schema[2]. Details of such a mapping can be found in the seminal work of Hunter [2]. In our work, we use MPEG-7 to model only the structural and the low level aspects of multimedia documents. The reason behind not using the semantic descriptors of MEPG-7 is due to their lack of expressiveness for modeling highly structured domains like sports or medicine. We are interested in MPEG-7 concepts that provide storage features of multimedia documents like format, encoding and location. We also focus on structural aspects like spatial, temporal, spatio-temporal concepts, and also on certain low level features like colour, shape, texture, timbre and melody for images, video and audio files. High level semantics are exclusively captured using domain-specific vocabularies. For example, one could use an ontology that provides descriptors for soccer domain. Integrating these two ontologies is basically a design issue. The reader is referred to our recent work [3] for details.

2 Challenges and future work

Whilst a lot of research has been done in the design of multimedia ontologies, we observe that there is enough scope for further work in the automatic anno-

[1] http://www.w3.org/2001/sw/
[2] http://www.w3.org/TR/rdf-schema/

Dagstuhl Seminar Proceedings 06171
Content-Based Retrieval
http://drops.dagstuhl.de/opus/volltexte/2006/655

tation of multimedia content. It is possible to derive low level descriptors using state-of-the-art techniques in multimedia content analysis, but the same does not hold true when it comes to analysing multimedia content at a high level of abstraction that deals with semantics. The works of Naphade et al. [4], Barnard et al. [5] and Vogel and Schiele [6] analyse multimedia documents at semantic levels. These techniques would be helpful for the automatic population or annotation of multimedia ontologies described in the earlier section. An interesting yet unexplored line of research is ontology learning from multimedia documents. The underlying motivation is the developments that are taking place in learning ontologies from text documents [7] using natural language processing techniques. At some point of time in the future, we see the need to extend these techniques for ontology learning from multimedia documents that would in essential get rid of the knowledge-acquisition bottleneck. In the recent past, the so-called Web 2.0 or social software is gaining popularity. Flickr[3] and YouTube[4] are prominent examples. From a research point of view, these applications offer rich sources of information for analysing multimedia documents using machine learning techniques. Real-world data covering text, video, audio, tags and social relations are often accessible via open APIs or standardised XML-formats.

If we succeed in finding ways for semantic tagging and ontology learning, a new range of applications can be foreseen. Having annotated media snippets at hand enables users to create their own re-mixes of existing material. The recent success of such creative work on an amateur level – appearing at video sharing platforms – shows a tremendous interest in such activities. Cross-media and cross-domain applications could offer easy access to Creative Commons[5] licensed content via semantic search and subsequent production. Furthermore, we will see the national media broadcasters adopting new models of content distribution that give rise to multimedia sources in the large (e.g., BBC open news archive). A second wave of media-based applications will be in the area of recommendation systems. Last.fm[6] and Pandora[7] are well-known sites providing music recommendations based on either media or listening behaviour. In the future, trend spotting and recommendation engines for cross-domain taste based on cross-media data sources are feasible. Research results based on text analysis and social media have been recently published at the MIT Media Lab [8].

References

1. Chang, S.F., Sikora, T., Puri, A.: Overview of the MPEG-7 standard. IEEE Transactions on Circuits and Systems for Video Technology **11**(6) (2001) 688–695
2. Hunter, J.: Adding multimedia to the semantic web - Building an MPEG-7 ontology. In: Proceedings of the International Semantic Web Working Symposium, Stanford University, California, USA (2001)

[3] http://www.flickr.com
[4] http://youtube.com
[5] http://creativecommons.org/
[6] http://www.last.fm
[7] http://www.pandora.com

3. Vembu, S., Kiesel, M., Sintek, M., Baumann, S.: Towards bridging the semantic gap in multimedia annotation and retrieval. In: Proceedings of the 1st International Workshop on Semantic Web Annotations for Multimedia at the 15th International World Wide Web Conference, Edinburgh, Scotland (2006)
4. Naphade, M., Kristjansson, T., Frey, B., Huang, T.S.: Probabilistic multimedia objects (multijects): A novel approach to indexing and retrieval in multimedia systems. In: Proceedings of the 5th IEEE International Conference on Image Processing, Chicago, IL, USA (1998)
5. Barnard, K., Duygulu, P., Forsyth, D., de Freitas, N., Blei, D., Jordan, M.: Matching words and pictures. Journal of Machine Learning Research **3** (2003) 1107–1135
6. Vogel, J.: Semantic Scene Modeling and Retrieval. PhD thesis, Swiss Federal Institute of Technology, Zurich (2005)
7. P. Buitelaar, P. Cimiano, M.G., Sintek, M.: Tutorial on ontology learning from text. In: Proceedings of the 16th European Conference on Machine Learning, Porto, Portugal (2005)
8. Liu, H., Maes, P., Davenport, G.: Unraveling the taste fabric of social networks. Interntational Journal on Semantic Web and Information Systems **2**(1) (2006) 42–71