Semantics for Multimedia on the Web

B.J. Wielinga A. Th. Schreiber J. Wielemaker

University of Amsterdam, Social Science Informatics
Rotersstraat 15, NL 1018 WB Amsterdam, The Netherlands
E-mail: {wielinga,schreiber,jan}@swi.psy.uva.nl

Position paper for the Semantic Web Workshop SWWS
Stanford 30-31 July 2001

The vision of the Semantic Web entails that large amounts of multimedia data should be annotated with semantic meta data. Current technology for content-based image interpretation is too limited for automated annotation of visual material. Techniques used by image search engines are also very poor and are unlikely to be improved in the near future. So, human annotations are required to make large annotated image corpora available on the Semantic Web. Currently, image archives use human annotators to select a set of keywords to describe an image. There are several problems with this approach: consistency of vocabulary is hard to enforce, keywords cannot describe relational properties of the images and the keyword approach does not scale up. Searching an image archive such as GETTYONE[1] which contains some 350.000 images, with the keyword `ape` delivers 575 hits. The Lycos multimedia engine finds over 5000 images for the query `GORILLA`. More specialized queries tend to give unstable results due to inconsistent indexing methods. Searching the full Semantic Web with tenths of millions of images with a simple keyword approach is not realistic anymore. A more sensitive method for creating semantic annotations for multimedia material is needed.

We are developing a paradigm for indexing and retrieval of images and other multimedia materials for the Semantic Web. The paradigm is based on the following key notions:

- Annotations are built from structured sets of descriptive elements, which are extensions of the Dublin Core [1] and VRA 3.0 [2] metadata element sets.
- Relational information is represented using complex instantiated structures, resembling sentence structures [3].
- Annotations can be hierarchically structured. The subject slot of an image description can contain references to other descriptions of objects in the image.
- Existing thesauri such as WordNet and AAT [4] are used as sources for closed vocabularies.
- The thesauri are augmented with additional knowledge to create proper ontologies. These ontologies are used to infer additional information from a partial description.
- Annotation tools and retrieval tools (search engines and browsers) are automatically configured on the basis of the ontology.

[1] UEL: http://www.gettyone.com.

- Representation of the ontologies and the indexed material, and implementation of the tools is compliant with current W3C standards, notably RDF, RDFS, and the usability of emerging standards such as RULE-ML are being investigated.

In an earlier project on stolen art objects [5] we have used this approach (without the RDF representations) successfully for the structured description of stolen art and antique objects. More recently we are applying the paradigm in a number of domains (photographs of animals, landscapes, buildings, paintings) [6]. In these experiments a number of tools were used and developed. We use Protg 2000 [7] for construction of the ontologies and for generating RDFS data. The Protg WordNet plugin was used to select certain parts of the WordNet ontology that were relevant to a particular domein. An RDFS version and browser for the Art and Architecture Thesaurus was developed and integrated with the description ontologies. An image annotation toolkit that automatically generates a user interface from the RDFS specification of an ontology was used in a number of the experiments. A preliminary conclusion from the experiments is that the highly structured annotation of images allows a much more fine-grained retrieval of images than the standard keyword approach. The use of RDFS technology turned out to be staifactory, with the exception of the representation of inference rules [8]. In structured annotations it is often possible to aid the annotation process by limiting the set of possible values based on constraints. In many situations it is also feasible to provide sensible defaults based on a partial annotation to speedup the annotation process. This requires knowledge about constraints between properties as well as default knowledge, neither of which can be expressed in RDFS.

References

[1] Dublin Core Metadata Initiative. *Dublin Core Metadata Element Set Version 1.1: Reference Description*, July 1999. Url: http://dublincore.org/documents/1999/07/02/dces/.

[2] Visual Resources Association Standards Committee. VRA core categories, version 3.0. Technical report, Visual Resources Association, July 2000. URL: http://www.gsd.harvard.edu/ staffaw3/vra/vracore3.htm.

[3] A. M. Tam and C. H. C. Leung. Structured natural-language description for semantic content retrieval. *Journal of the American Society for Information Science*, to appear.

[4] T. Peterson. *Introduction to the Art and Architecture Thesaurus*. Oxford University Press, 1994. See also: http://shiva.pub.getty.edu.

[5] B. J. Wielinga, J. A. C. Sandberg, and A. Th. Schreiber. Methods and techniques for knowledge management: What has knowledge engineering to offer? *Expert Systems With Applications*, 13(1):73–84, 1997.

[6] A. Th. Schreiber, B. Dubbeldam, J. Wielemaker, and B. J. Wielinga. Ontology-based photo annotation. *IEEE Intelligent Systems*, 2001. To appear. URL: http://www.swi.psy.uva.nl/usr/Schreiber/Schreiber01a.pdf.

[7] N. Fridman Noy, R. W. Fergerson, and M. A. Musen. The knowledge model of Protégé-2000: combining interoperability and flexibility. In *Knowledge Engineering and Knowledge Management: 12th International Conference EKAW2000, Juan-les-Pins*, volume 1937 of *Lecture Notes in Artificial Intelligence*, pages 17–32, Berlin/Heidelberg, 2000. Springer-Verlag. Also as: Technical Report Stanford University, School of Medicine, SMI-2000-0830.

[8] B. J. Wielinga, A. Th. Schreiber, J. Wielemaker, and J. A. C. Sandberg. From thesaurus to ontology. Report ICES-MIA project, University of Amsterdam, Social Science Informatics, 2001. Submitted for publication. URL: http://www.swi.psy.uva.nl/usr/Schreiber/Wielinga01a.pdf.