Multimedia and the Semantic Web

David Hulme
Electronics and Computer Science
University of Southampton
dh7g10@ecs.soton.ac.uk

ABSTRACT

With the advent of Web 2.0 and user-generated content, the amount of multimedia on the Web is growing. This vast amount of data, coupled with its rich nature, make it challenging to bring to the Semantic Web. This paper researches the process of making multimedia data accessible to the Semantic Web and applications that could become possible.

1. INTRODUCTION

The amount of multimedia data on the web is growing fast. Popular web sites such as YouTube[1] and Flickr[2], encourage users to post content, often with no restrictions on the amount they can upload. Currently, multimedia retrieval from these services relies on text-based keyword searches from annotations such as *Title* and *Author*. These attributes are manually added by users; they typically have no formal meaning and cannot describe the media content in depth.

This paper examines the way in which multimedia data can be annotated and how this data can be structured for use by the Semantic Web. It will also examine applications that could be possible with a multimedia rich Semantic Web.

2. DESCRIPTORS

As multimedia databases of web sites such as YouTube grow, the ability to retrieve relevant information from them becomes increasingly important. Alemu et al. discuss two general approaches for image retrieval: text-based and content-based [2]. Descriptors can be added to content manually, or automatically, through content analysis. Automatic methods can annotate large numbers of documents quickly, but can be difficult to develop and can lead to incorrect annotations being applied to media. Manual methods produce accurate annotations, however they do not scale well with the corpus size. This section will examine how the data for text and content based approaches can be sourced and how high level semantic descriptors can be produced from low level data.

2.1 Text-based

Text-based descriptors can be sourced from multimedia metadata. There are a number of different standards in use across different media types.

The Exif (Exchangeable image file format) standard specifies the format to store images captured by digital cameras, including fields such as location, ISO speed and exposure time [4]. Video container formats, such as MP4, can contain metadata formatted according to their specification, or often support embedding XMP (Extensible Metadata Platform) data, an ISO standard that specifies a data model for storing metadata. XMP data can be stored using a number of namespaces, such as Dublin Core[3] [1].

Text-based metadata can also be stored in a separate location to the media, for example the *Title* and *Description* of user generated content uploaded to a Web 2.0 service are not embedded in the uploaded multimedia file, rather they are stored in a database. This aids performing queries, however it requires links to be maintained between data and metadata.

Zakaria et al. combined text-based processing with knowledge bases to create an information retrieval solution for images from the Flickr service [16]. The textual description associated with each image was processed using natural language analysis (NLA) tools. The domain of interest was Malaysian tourism, leading to concept descriptors such as *Island* and *Malaysia*.

2.2 Content-based

Alemu et al. note that often the bottleneck of the efficiency of retrieval is the semantic gap between high level requests by users and low level features stored in multimedia databases [2]. Over the past 25 years, much work has been done to develop techniques to design algorithms that extract low level features to produce semantic high level descriptors [7]. The first stage of research, up until the mid 1990s, focussed on developing algorithms to extract low level features, creating systems that could respond to queries such as *'find all images containing a dark green area near the top*

[1] YouTube: `https://www.youtube.com/`

[2] Flickr: `https://www.flickr.com/`

[3] Dublin Core: `http://dublincore.org/`

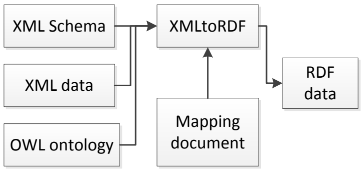

Figure 1: XML to RDF conversion process, adapted from [14]

of the image' [7]. Such queries are unlikely to be thought of by users. The second phase has brought a shift to higher semantic levels, allowing for more natural queries from users such as *'find all shots from a football match containing someone scoring a goal'* [6], coupled with standardisation efforts such as MPEG-7 [7].

MPEG-7 is an ISO standard for describing features of multimedia content. The standard is comprised of four main elements. Descriptors that define the syntax and semantics of each feature, Description Schemes that specify the relationships between components, a Description Definition Language (DDL) to allow for the creation of new Description Schemes and finally tools to support the binary representation of the data [9].

Systems have been developed that analyse multimedia data and generate MPEG-7 metadata, such as the work of Lin et al. in [10]. They were able to identify and store concepts such as *People*, *Landscape* and *Monologue*, with their solution performing better than 18 other comparable systems [10].

2.3 Structured

For descriptors to be used in semantic reasoning, they must be converted from semi-structured formats, such as XML, to structured formats, typically RDF (Resource Description Framework).

Van Deursen et al. describe a generic approach for converting XML to RDF in [14]. Their process takes XML data, an XML Schema, an OWL ontology and a mapping document as inputs, then outputs RDF data, see Figure 1. The mapping document is the key component, providing the link between the XML Schema and OWL ontology. Their tool was successfully used to convert XML data formatted according to DIG35, an XML-based metadata description standard for digital still images [14].

3. ONTOLOGIES

Once descriptors have been extracted, ontologies can be formed to enable semantic reasoning. Shirahama et al. produced a video ontology, see Figure 2, from descriptors such as *Action* and *Location* [11]. They were able to correctly categorise daytime and night-time with 90% and 63% accuracy respectively [11].

The work of Zakaria et al. in [16], describe how descriptors were given semantic meaning through the use of ontologies. Concepts extracted from textual descriptors were passed to knowledge bases made up of ontologies, such as the Malaysian Tourism Ontology, to map descriptors to ontological concepts [16]. They created a prototype to retrieve data from an RDF store. Testing showed that their retrieval system recalled more images than full-text or tag searches and gave improved precision when compared to tag-based searches [16]. Their future goal was to combine these two components with content analysis to give higher confidence when describing the images [16].

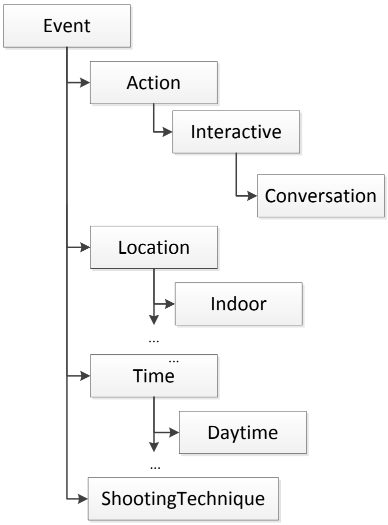

Figure 2: Video ontology [11]

With the arrival of MPEG-7, a standardised multimedia video ontology exists. XML Schema has been chosen as the DDL for MPEG-7 [9]. However, as Hunter noted, "XML Schema provides little support for expressing semantic knowledge" [8]. She attempted to develop an ontology for MPEG-7 in RDF Schema, however noted that "although RDF Schema is capable of expressing the semantics of MPEG-7 Description Schemes and Descriptors, it does have certain serious limitations" that could be addressed through the use of more expressive ontology languages [8]. A year later, Yoo et al. successfully showed how the MPEG-7 specification can be converted into RDF Schema [15].

Tsinaraki et al. furthered this work, presenting a methodology for interoperability support between MPEG-7 and the more expressive ontology language, OWL [12]. However, as García and Celma note, this work did not taken the step of completely moving MPEG-7 to the Semantic Web [5]. They achieved this goal by converting the XML Schema for MPEG-7 into OWL using the ReDeFer project tool: XSD2OWL[4].

4. APPLICATIONS

As more multimedia data has semantic descriptors associated with it, new types of applications become possible. This section explores some of the new applications that have been developed with a multimedia-enhanced Semantic Web.

[4] ReDeFer project: http://rhizomik.net/html/redefer/

4.1 Semantic Search

For semantic multimedia search to be successful, Bonino et al. note critical features such applications must support. Firstly, systems should be able to deal with all multimedia documents uniformly, whether textual, audio, video or a mixture and secondly, systems should support searching across different information sources [3]. They developed a system to meet these requirements, by representing resources through RDF 'Resource Descriptors' [3]. Two datasets were combined and could be queried to return results based on concepts determined from queries.

4.2 Metadata Propagation

Tummarello et al. defined 'Semantic Audio Hyperlinks' using MPEG-7 descriptors, allowing them to express relations between audio tracks such as *'sounds like'* or *'same instrument playing'* [13]. Using these hyperlinks, they were able to expand *Genre* metadata from some multimedia resources to others [13]. Starting from 20% of the corpus containing genre annotations, up to 75% of the corpus was correctly annotated [13]. Such a tool could be useful to annotate the vast amount of multimedia data on the web that may have metadata missing, although further work is needed to increase the accuracy.

5. CONCLUSIONS AND FUTURE DEVELOPMENTS

As a response to the huge amount of multimedia content on the Web, progress in being made in bringing all this data to the Semantic Web. This work has been aided by the MPEG-7 standard for describing multimedia content. The development of OWL ontologies for MPEG-7 and the ability to represent MPEG-7 data in RDF, has enabled new applications such as semantic multimedia search engines.

As these applications continue to improve, a new development phase will begin where multimedia information retrieval systems will be integrated into practical solutions [7]. The dependability of multimedia retrieval systems will need to improve to enable integration with technologies such as personal video recorders and critical systems, such as analysis of video surveillance data [7].

In conclusion, the future of the multimedia and the Semantic Web is bright and the 'Holy Grail' of multimedia information retrieval may finally be in reach.

6. REFERENCES

[1] Adobe Systems Incorporated. XMP specification, 2012.

[2] Y. Alemu, J.-B. Koh, M. Ikram, and D.-K. Kim. Image retrieval in multimedia databases: A survey. In *Intelligent Information Hiding and Multimedia Signal Processing, 2009. IIH-MSP '09. Fifth International Conference on*, pages 681–689, Sept 2009.

[3] D. Bonino, F. Corno, and P. Pellegrino. Versatile rdf representation for multimedia semantic search. In *Tools with Artificial Intelligence, 2007. ICTAI 2007. 19th IEEE International Conference on*, volume 2, pages 32–38, Oct 2007.

[4] Camera and Imaging Products Association. Exchangeable image file format for digital still cameras: Version 2.3, 2012.

[5] R. García and O. Celma. Semantic integration and retrieval of multimedia metadata. In S. Handschuh, T. Declerck, and M. Koivunen, editors, *Proceedings of the ISWC 2005 Workshop on Knowledge Markup and Semantic Annotation (Semannot'2005)*, volume 185, pages 69–80. CEUR Workshop Proceedings, 2005.

[6] Y. Gong, L. T. Sin, C. H. Chuan, H. Zhang, and M. Sakauchi. Automatic parsing of tv soccer programs. In *Multimedia Computing and Systems, 1995., Proceedings of the International Conference on*, pages 167–174, May 1995.

[7] A. Hanjalic, R. Lienhart, W.-Y. Ma, and J. R. Smith. The holy grail of multimedia information retrieval: So close or yet so far away? *Proceedings of the IEEE*, 96(4):541–547, April 2008.

[8] J. Hunter. *Adding Multimedia to the SemanticWeb: Building and Applying an MPEG-7 Ontology*, pages 75–106. John Wiley & Sons, Ltd, 2005.

[9] ISO. MPEG-7 overview, 2004.

[10] C.-Y. Lin, B. L. Tseng, M. Naphade, A. Natsev, and J. R. Smith. Mpeg-7 video automatic labeling system. In *Proceedings of the Eleventh ACM International Conference on Multimedia*, MULTIMEDIA '03, pages 98–99, New York, NY, USA, 2003. ACM.

[11] K. Shirahama, K. Otaka, and K. Uehara. Content-based video retrieval using video ontology. In *Multimedia Workshops, 2007. ISMW '07. Ninth IEEE International Symposium on*, pages 283–289, Dec 2007.

[12] C. Tsinaraki, P. Polydoros, and S. Christodoulakis. Interoperability support for ontology-based video retrieval applications. In P. Enser, Y. Kompatsiaris, N. O'Connor, A. Smeaton, and A. Smeulders, editors, *Image and Video Retrieval*, volume 3115 of *Lecture Notes in Computer Science*, pages 582–591. Springer Berlin Heidelberg, 2004.

[13] G. Tummarello, C. Morbidoni, P. Puliti, and F. Piazza. Semantic audio hyperlinking: a multimedia-semantic web scenario. In *Automated Production of Cross Media Content for Multi-Channel Distribution, 2005. AXMEDIS 2005. First International Conference on*, pages 4 pp.–, Nov 2005.

[14] D. Van Deursen, C. Poppe, G. Martens, E. Mannens, and R. Walle. Xml to rdf conversion: A generic approach. In *Automated solutions for Cross Media Content and Multi-channel Distribution, 2008. AXMEDIS '08. International Conference on*, pages 138–144, Nov 2008.

[15] J. Yoo, S. Myaeng, S. Kim, and H. Lee. Automatic conversion of mpeg-7 specification and data into rdf(s) for semantic interoperability in information retrieval. In *Computational Intelligence for Modelling, Control and Automation, 2005 and International Conference on Intelligent Agents, Web Technologies and Internet Commerce, International Conference on*, volume 1, pages 45–50, Nov 2005.

[16] L. Q. Zakaria, W. Hall, and P. Lewis. Modelling image semantic descriptions from web 2.0 documents using a hybrid approach. In *Proceedings of the 11th International Conference on Information Integration and Web-based Applications &Amp; Services*, iiWAS '09, pages 306–312, New York, NY, USA, 2009. ACM.