Standards

John R. Smith
IBM

Multimedia Annotations on the Semantic Web

Giorgos Stamou
National Technical University of Athens

Jacco van Ossenbruggen
CWI, the Netherlands

Jeff Z. Pan
University of Aberdeen

Guus Schreiber
Free University of Amsterdam

In recent years, the amount of digital multimedia information distributed over the Web has increased tremendously because everyone can now follow the production line of digital multimedia content. Production houses shoot high-quality video in digital format; organizations that hold multimedia content (such as TV channels, film archives, museums, and libraries) digitize analog material and use digital formats for preservation, management, and distribution; and even novices use digital devices to produce image and video content in MPEG and JPEG formats that's ready for delivery. Moreover, because of the maturity of storage and data network technologies, digital formats now provide the cheapest, safest, and easiest way to store and deliver multimedia content, even in high resolutions. Standardization activities, which provided MPEG, JPEG, and other digital coding formats, have played a key role in these developments.

As a result of this proliferation, however, those dealing with multimedia content—from professional archivists to amateur photographers—are faced with daunting problems when it comes to storing, annotating, and retrieving media items from multimedia repositories, whether via the Web or otherwise. Although the standardization activities of the International Organization for Standardization (ISO) as well as other communities (see the "Related Standardization Activities" sidebar for details) have provided standards for describing content, they haven't been widely used for several reasons. First, it's difficult, time consuming, and thus expensive to manually annotate multimedia content. Second, many organizations feel that the complexity of many standards makes multimedia annotation unnecessarily difficult. Third, there's little incentive for organizations to provide, for example, MPEG-7 metadata because there are insufficient applications that would benefit from its use.

Most of these problems could be solved by merging and aligning existing good practices in multimedia industry with the current technological advances of the Semantic Web. Such integration would give metadata providers immediate payoff because they could directly benefit from publicly available Semantic Web software. In addition, it would enable the deployment of more intelligent applications that could reason over multimedia metadata in a way that's impossible today because current multimedia metadata standards are usually (XML) syntax oriented and thus lack formal semantics.

The Semantic Web's "open world" approach would simplify integrating multiple vocabularies from different communities. The ultimate goal is to provide small and simple but extensible vocabularies. These vocabularies should be suitable for private use but, at the same time, be sufficiently flexible for extending to more complex and professional annotation tasks.

Multimedia semantic annotation

We can formalize, represent, analyze, and process the information a multimedia docu-

Editor's Note

Multimedia in all forms (images, video, graphics, music, speech) is exploding on the Web. The content needs to be annotated and indexed to enable effective search and retrieval. However, recent standards and best practices for multimedia metadata don't provide semantically rich descriptions of multimedia content. On the other hand, the World Wide Web Consortium's (W3C's) Semantic Web effort has been making great progress in advancing techniques for annotating semantics of Web resources. To bridge this gap, a new W3C task force has been created to investigate multimedia annotations on the Semantic Web. This article examines the problems of semantically annotating multimedia and describes the integration of multimedia metadata with the Semantic Web.

—*John R. Smith*

Related Standardization Activities

Several standardization activities are relevant to multimedia document annotation. We can use the standards to serialize metadata descriptions containing several levels of information: descriptive for describing and identifying information, structural for navigation and presentation, administrative for management and processing, amd so forth. This is a list of the most important standardization activities:

- Dublin Core Community, "Dublin Core ElementSet, Version 1.1," 2003, ISO Standard 15836-2003, 2003; http://www.niso.org/international/SC4/n515.pdf (it provides a small set of descriptors that quickly drew global interest from various information providers in the arts, sciences, education, business, and government sectors).

- ISO/IEC, "Overview of the MPEG-7 Standard; Version 6.0," ISO/IECJTC1/SC29/WG11/N4980, 2001 (it's a standard for describing the multimedia content data that supports some degree of interpretation of the information meaning, which can be passed onto, or accessed by, a device or a computer code).

- "MPEG-21 Overview v.5," ISO/IEC JTC1/SC29/WG11/N5231, Shanghai, 2002; http://www.chiariglione.org/mpeg/standards/mpeg-21/mpeg-21.htm (it aims at defining a normative open framework for multimedia delivery and consumption for use by all the players in the delivery and consumption chain).

- NewsML; http://www.newsml.org/ (it's designed to provide a media-independent, structural framework for multimedia news).

- TV-Anytime Forum; http://www.tv-anytime.org/ (it aims to develop specifications to enable audio–visual and other services based on mass-market high volume digital storage in consumer platforms—simply referred to as local storage).

- Virtual Resource Association, VRA Core 3.0; http://www.vraweb.org/vracore3.htm (it consists of a single element set that can be applied as many times as necessary to create records to describe works of visual culture and the images that document them).
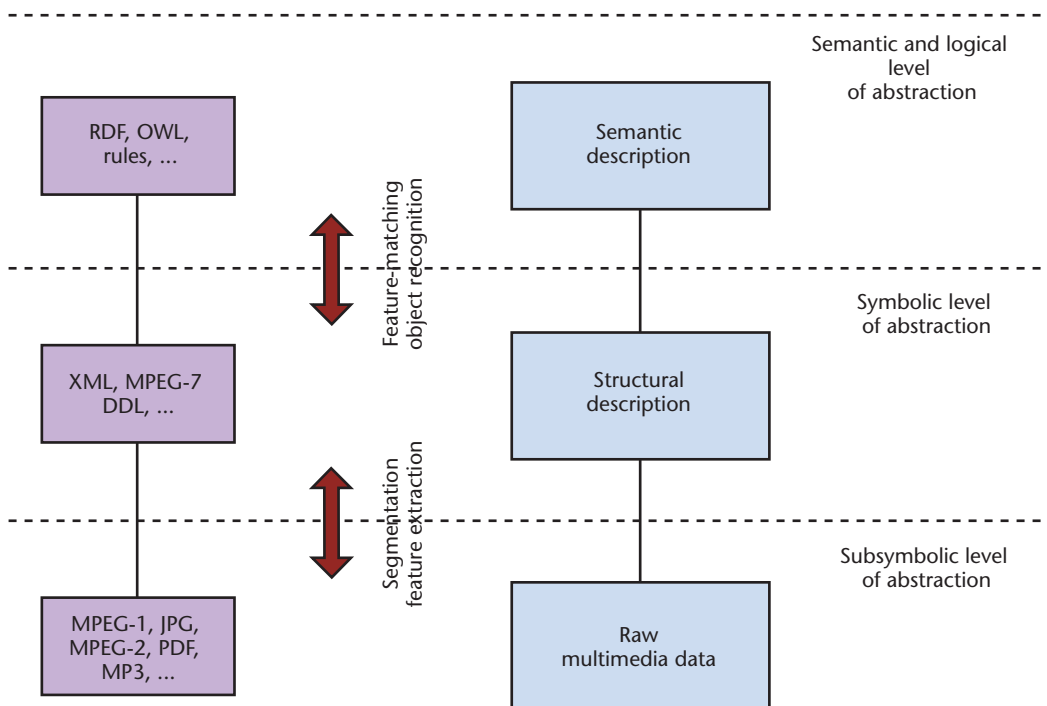

Figure 1. The different levels of multimedia information and the type of annotation provided for each level. RDF = Resource Description Framework, OWL = Web Ontology Language, XML = Extensible Markup Language, and MPEG-7 DDL = MPEG-7 Description Definition Language.

ment conveys with three different levels of abstraction: subsymbolic, symbolic, and logical (see Figure 1).

The subsymbolic abstraction level covers the raw multimedia information represented in well-known formats for video, image, audio, text, metadata, and so forth. These are typically binary formats, optimized for compression and streaming delivery. They aren't necessarily well suited for further processing that uses, for exam-
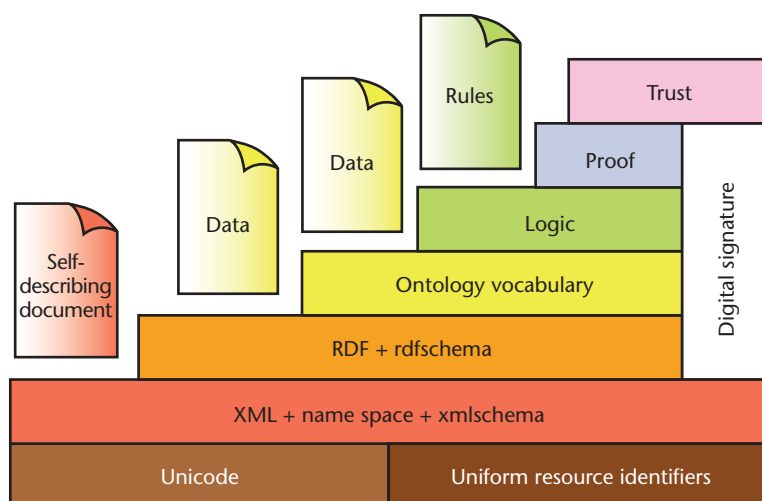
Figure 2. The Semantic Web stack, as presented by Tim Berners-Lee.

ple, the internal structure or other specific features of the media stream.

To address this issue, we can introduce a symbolic abstraction level, like the middle layer in Figure 1, which provides this information. This is the MPEG-7 approach, which lets us use feature detectors' output, (multicue) segmentation algorithms, and so on to provide a structural layer on top of the binary media stream. Information on this level is typically serialized in XML. The standards that have been proposed and partly used in the literature for the representation of multimedia document descriptions (Dublin Core, MPEG-7, MPEG-21, Visual Resource Association [VRA], International Press Telecomunications Council [IPTC], and so on) mainly operate in this middle layer of Figure 1.

The problem with this structural approach is that the semantics of the information encoded in the XML are only specified within each standard's framework (using that standard's structure and terminology). For example, if we use the MPEG-7 standard, then it's hard to reuse this data in environments that aren't based on MPEG-7 or to integrate non-MPEG metadata in an MPEG-7 application. This conflicts with the interoperability that's so crucial to Web-based applications.

To address this, we could simply replace the middle layer with another open one that has formal, machine-processable semantics by using a more appropriate, semantically enriched language like the Resource Description Framework (RDF). However, this wouldn't take advantage of existing XML-based metadata, and more importantly, it ignores the advantages of an XML-based

structural layer. Rather than changing the middle layer, a possible solution is to add a third layer (the logical abstraction level) that provides the semantics for the middle one, actually defining mappings between the structured information sources and the domain's formal knowledge representation. An example of this is the Web Ontology Language (OWL). In this layer, we can make the implicit knowledge of the multimedia document description explicit and reason with it—for example, to derive new knowledge not explicitly present in the middle layer.

Providing interoperable semantic multimedia annotations

The stack of RDF-based languages and technologies provided by the World Wide Web Consortium community (see Figure 2)[1] is well suited to the formal, semantic descriptions of the terms in a multimedia document's annotation. However, because they lack the structural advantages of the XML-based approach and the work on multimedia document annotation already done within the framework of other standards, a combination of the existing standards seems to be the most promising path for multimedia document description in the near future.[2-6]

For these reasons, the World Wide Web Consortium (W3C) has started a Multimedia Annotation on the Semantic Web Task Force (http://www.w3.org/2001/sw/BestPractices/MM/) as part of the Semantic Web Best Practices and Deployment Working Group (see the sidebar for details). The new task force operates within the framework of the W3C Semantic Web Activity group (http://www.w3.org/2001/sw/). The task force's target audience includes institutions and organizations with research and standardization activities in multimedia, including professional (museums, libraries, audiovisual archives, media production and broadcast industry, and image and video banks) and nonprofessional (end users) multimedia annotators. One goal is to provide guidelines for using Semantic Web languages and technologies to create, store, manipulate, interchange, and process image metadata. Another is to study interoperability issues between multimedia annotation standardization frameworks (such as MPEG-7, Dublin Core, and VRA) and RDF- and OWL-based approaches.

Hopefully, this effort will provide a unified framework of good practices for constructing interoperable multimedia annotations. In its first stage, the task force will focus on image annota-

tion, while later on, it will cover other types of multimedia data such as video. More specifically, the task force has several main objectives.

Review and discuss vocabularies

Choosing which vocabularies to use when annotating an image is a key decision in an annotation project. Typically, we need more than a single vocabulary to cover the images' different relevant aspects. Many of the relevant vocabularies were developed prior to the Semantic Web, but the task force's main focus will be on translations of such vocabularies to RDF or OWL. Most notably, the key international standard in this area, MPEG-7, is defined using XML Schema. At this time, no commonly accepted mapping exists from the XML Schema definitions in the standard to RDF or OWL. The task force will examine several alternative mappings that have been developed so far.

Another relevant vocabulary is the VRA Core. Where the Dublin Core specifies a small and commonly used vocabulary for online resources in general, VRA Core defines a similar set targeted especially at visual resources. Dublin Core and VRA Core both refer to terms in their vocabularies as elements and similarly use qualifiers to refine elements. VRA Core's more general elements have direct mappings to comparable fields in Dublin Core. Furthermore, both vocabularies are defined in a way that abstracts from implementation issues and underlying serialization languages. A key difference, however, is that a commonly accepted mapping to RDF exists for Dublin Core, along with the associated schema. At this time, this isn't the case for VRA Core. The task force will discuss the pros and cons of the alternative mappings.

Review and test existing tools

Besides the hundreds of tools used for image archiving and annotation, many tools also exist for semantic image annotation. It's the task force's goal to describe the semantic annotation tools and their characteristics to provide some guidelines for their proper use. Using these characteristics as criteria, we can categorize the tools so that users who want to annotate their multimedia content could choose the most appropriate for their application.

Provide use cases examples

The task force will develop use cases as a representative set of examples that can later be used to

---

---

discuss the vocabularies and tools that are relevant for image annotation on the Semantic Web. Moreover, some example solutions will guide users on how to use the tools and vocabularies. The use cases are organized in four categories (world images, cultural images, scientific images, and media), which reflect the topics the images depict. These topics often determine the tools and vocabularies used in the annotation process.

Study the syntactic and semantic interoperability problem

Another objective is to study the problem of syntactic and semantic interoperability between Semantic Web technologies and existing image annotation standards (such as MPEG-7, Dublin Core, VRA, and IPTC). For example, the problem of aligning Semantic-Web-based approaches with MPEG-7 is a major issue. (A thorough comparison and list of the open issues in integrating the MPEG-7 and Semantic Web approaches are available elsewhere.[7,8]) The task force's framework will cover this by studying the way to define mappings between different multimedia annotations that take advantage of the MPEG-7 standard using OWL. Moreover, in the short term, we need to show how RDF-based software can take advantage of popular existing, non-RDF metadata. In the long term, integrating Semantic Web technologies in major multimedia tools is essential.

Study possible transformation issues

The task force will study possible transformation issues between Semantic Web technologies and existing image annotation standards by connecting the problem with the definition of map-

pings and ontology alignment techniques. Moreover, the task force will provide a feasibility report that clarifies the levels of interoperability that could be achieved in each case. Using this study, it will describe the main ideas for constructing interoperable image annotations. The main goal is to provide clear and concrete suggestions for annotators and tool developers.

Conclusions

Solving the problems we've discussed here obviously requires effort from both the multimedia and the Semantic Web communities. It's important, however, that the W3C community make a special effort to explain the Semantic Web technologies' added value to convince multimedia content owners to go beyond the current, XML-based approaches. From this perspective, this W3C Multimedia Annotation on the Semantic Web Task Force is a call to arms to both the Semantic Web and the multimedia community to address these issues.

All the documents the task force produces are available for public review. Contributions in any possible form to this effort are not only welcome but a necessary condition for achieving our goals.

Even when all the relevant issues have been solved, multimedia annotation will remain a difficult, time-consuming, and expensive process. The question is whether we can develop the required standards in a way that reduces, and not adds, to the task's complexity and develop the tools and applications with an added value to make multimedia annotation payoff in practice.   **MM**

Acknowledgments

References

1. T. Berners-Lee, *Weaving the Web*, Orion Business Books, 1999.
2. J. Hunter, J. Drennan, and S. Little, "Realizing the Hydrogen Economy through Semantic Web Technologies," *IEEE Intelligent Systems*, vol. 19, no. 1, Jan./Feb. 2004, pp. 40-47.
3. A. Jaimes and J.R. Smith, "Semiautomatic, Data-Driven Construction of Multimedia Ontologies," *Proc. IEEE Int'l Conf. Multimedia and Expo* (ICME), IEEE CS Press, 2003.
4. G. Stamou and S. Kollias, eds., *Multimedia Content and the Semantic Web: Methods, Standards, and Tools,* John Wiley & Sons, 2005.
5. R. Troncy, *Integrating Structure and Semantics into Audio-Visual Documents*, Springer-Verlag, pp. 566-581; http://www.springerlink.com/media/6cc6l524yndvyn5jte27/Contributions/U/%3/T/X/U3TXQY8BR03TE7RG.pdf.
6. C. Tsinaraki, P. Polydoros, and S. Christodoulakis, "Interoperability Support for Ontology-Based Video Retrieval Applications," *Proc. 3rd Int'l Conf. Image and Video Retrieval* (CIVR), 2004, pp. 582-591.
7. J. van Ossenbruggen, F. Nack, and L. Hardman, "That Obscure Object of Desire: Multimedia Metadata on the Web (Part I)," *IEEE MultiMedia*, vol. 11, no. 4, Oct.– Dec. 2004, pp. 38-48.
8. F. Nack, J. van Ossenbruggen, and L. Hardman, "That Obscure Object of Desire: Multimedia Metadata on the Web (Part II)," *IEEE MultiMedia*, vol. 12, no. 1, Jan.– Mar. 2005, pp. 54-63.

Readers may contact Giorgos Stamou at gstam@softlab.ntua.gr.

Contact Standards department editor John R. Smith at jsmith@us.ibm.com.