Knowledge-Driven Video Information Retrieval with LOD: From Semi-Structured to Structured Video Metadata

Leslie F. Sikos
CSEM Centre for Knowledge and Interaction Technology
Flinders University, Tonsley Park
GPO Box 2100 Adelaide SA 5001
AUSTRALIA

David M. W. Powers
CSEM Centre for Knowledge and Interaction Technology
Flinders University, Tonsley Park
GPO Box 2100 Adelaide SA 5001
AUSTRALIA

ABSTRACT

In parallel with the tremendously increasing number of video contents on the Web, many technical specifications and standards have been introduced to store technical details and describe the content of, and add subtitles to, online videos. Some of these specifications are based on unstructured data with limited machine-processability, data reuse, and interoperability, while others are XML-based, representing semi-structured data. While low-level video features can be derived automatically, high-level features are mainly related to a particular knowledge domain and heavily rely on human experience, judgment, and background. One of the approaches to solve this problem is to map standard, often semi-structured, vocabularies, such as that of MPEG-7, to machine-interpretable ontologies. Another approach is to introduce new multimedia ontologies. While video contents can be annotated efficiently with terms defined by structured LOD datasets, such as DBpedia, ontology standardization would be desired in the video production and distribution domains. This paper compares the state-of-the-art video annotations in terms of descriptor level and machine-readability, highlights the limitations of the different approaches, and makes suggestions towards standard video annotations.

Categories and Subject Descriptors

H.3.1 [**Information Storage and Retrieval**]: Content Analysis and Indexing – *Dictionaries*, *Indexing methods*;
I.2.4 [**Knowledge Representation Formalisms and Methods**]: *Representation languages*.

General Terms

Management, Performance, Design, Standardization, Languages.

Keywords

Video annotation, ontology, MPEG-7, Linked Open Data

1. INTRODUCTION

In recent years, the amount of videos being produced and shared has grown exponentially, mainly because high-quality video recording became more affordable, while storage costs have decreased, and more options have appeared for free online video sharing. Video contents are, however, difficult to process automatically due to the lack of semantics that would make the content "understandable" to software agents. Search engines, for example, still heavily rely on textual context and manually added key-phrases for processing high-level, domain-specific concepts [17], because there is a "semantic gap" between what computers and humans understand [18]. Beyond the unstructured proprietary tags embedded in video files, video metadata specifications have been standardized over the years for generic video metadata as well as for annotating regions derived from the spatial, temporal, and spatiotemporal segmentation of video contents (see Table 1).

Table 1. Core Standards of Online Video Annotation

Common Name	Standard	Description
Dublin Core	ISO 15836-2003	A set of general-purpose descriptors to annotate the title, creator, file format, language, etc., of a resource.
MPEG-7	ISO/IEC 15938 [14]	XML metadata to be attached to the timecode of MPEG-1, MPEG-2, and MPEG-4 contents (e.g., synchronized lyrics for a music video).
MPEG-21	ISO/IEC 21000 [15]	Machine-readable licensing information for MPEG contents in XML.
NewsML	IPTC NewsML-G2 [12]	A media-independent, structural framework for multimedia news.
TTML	W3C TTML1 [7]	Timed Text Markup Language.
TV-Anytime	ETSI TS 102 822 [8]	Controlled delivery of personalized multimedia content to consumer platforms [20].

A common feature of these standards is that they are based on XML or XML Schema (XSD), which are machine-readable, but not machine-interpretable, making the previous standards inefficient for automated content access, sharing, and reuse. The problem can be solved by using Semantic Web standards, rather than XML, for video annotations, in particular the Resource Description Framework (RDF), the RDF Schema (RDFS), and the Web Ontology Language (OWL). Among the previous standards, MPEG-7 and TV-Anytime have already been mapped to Semantic Web standards.

ESAIR'15, October 23, 2015, Melbourne, VIC, Australia

ACM 978-1-4503-3790-8/15/10…$15.00
DOI: http://dx.doi.org/10.1145/2810133.2810141

2. CONTROLLED VOCABULARIES

Dublin Core has been widely deployed on the conventional Web both as metadata embedded to multimedia files and as attached semi-structured data in XHTML, while the Semantic Web deploys Dublin Core as structured data. Creative Commons became the *de facto* standard for general-purpose licensing metadata, which is also suitable for video licensing. Schema.org provides the *de facto* standards for annotating video objects on the Web, such as `schema:video` and `schema:VideoObject` for generic video metadata, `schema:Movie`, `schema:MovieSeries`, `schema:CreativeWorkSeason`, and `schema:Episode` for specific video objects, all of which are indexed by search engines.

3. RDFS AND OWL ONTOLOGIES

Formal ontologies have proved to be efficient in the representation of video content and scene models with contextual information, suitable for advanced interpretation of video scenes [10]. Domain-specific multimedia ontologies are used to interlink machine-readable definitions of real-world persons and objects depicted in video contents from Linked Open Data datasets, most of which segregate domain knowledge from administrative and technical metadata. The structured video metadata can efficiently modelled in RDF and implemented 1) in the markup as RDFa, JSON-LD, and in (X)HTML5, also as HTML5 Microdata; 2) as a separate file, or 3) in an LOD dataset as part of the Linked Open Data cloud. Ontologies can be used not only for the semantic enrichment of video databases and embedded video objects on the Web, but also for streaming videos on video sharing portals, such as YouTube [3].

For over a decade, the semantic limitations of MPEG video content descriptions were addressed by mapping MPEG-7 concepts to OWL [5]. Another approach, the introduction of domain ontologies, produced ontologies that can be used for movie annotation, scene model representation, and video concept detection [21]. The research community focuses on MPEG-7 for its capability to describe dominant colors, texture, shape, motion, as well as spatial and temporal multimedia segment relations, and hierarchical structures for multimedia segment decomposition, regardless that these low-level descriptors are not ideal for content description, as pointed out by Boll et al. with the famous remark "color distribution feature values of an image for red, black, and yellow still do not allow the conclusion that the image shows a sunset" [2]. More importantly, MPEG-7 provides machine-readable syntactic metadata only, rather than machine-interpretable semantics. To address this limitation, researchers proposed multimedia ontologies based on the MPEG-7 standard. The first attempt to model the core parts of MPEG-7 in OWL Full was Hunter's MPEG-7 ontology [11]. The approach used RDF to formalize the MPEG-7 core, and incorporated DAML+OIL constructs to further detail the semantics. Tsinaraki et al. extended this ontology by including the full Multimedia Description Scheme part of MPEG-7 in OWL [22]. Isaac and Troncy proposed a core audio-visual ontology considering MPEG-7, ProgramGuideML, and TV-Anytime [13]. The first complete MPEG-7 ontology, *MPEG-7Ontos*, was generated using an XSD to OWL mapping in combination with a transparent mapping from XML to RDF [9]. Blöhdorn et al. proposed the *Visual Descriptor Ontology* (VDO) for image and video analysis, based on the visual components of MPEG-7, written in OWL DL [1]. Dasiopoulou et al. developed two OWL DL ontologies, the *Multimedia Content Ontology* (MCO) and the *Multimedia Descriptors Ontology* (MDO), covering structural descriptions of the MPEG-7 Multimedia Description Scheme (MDS), as well as the visual and audio parts of MPEG-7 [6]. Oberle et al. created an ontological framework to formally model the MPEG-7 descriptions and export them into OWL and RDF, covering the structural, localization, media, and low-level description tools [16].

In spite of the number of contributions to the field, mapping MPEG-7 concepts to RDFS or OWL does not address many of the issues inherited from MPEG-7, including conceptual ambiguity, syntactic interoperability, and structural complexity. Hence, multimedia ontologies have been introduced with MPEG-7 alignment, or as complementary, domain-specific ontologies, such as *LinkedMDB* [4]. However, multimedia ontologies consisting of a terminological part (TBox) and an assertional part (ABox), also known as *knowledge bases* (KB), do not have role definitions (RBox axioms), and are limited in terms of rule definitions and Description Logic expressivity. For example, the *Large Scale Concept Ontology for Multimedia* (LSCOM) and the *LinkedMDB* ontology correspond to the base Attribute Language ($\mathcal{AL}$), defining only atomic negation, concept intersection, universal restrictions, and limited existential quantification for their concepts. The *Multimedia Metadata Ontology* (M3O) corresponds to $\mathcal{SHIQ}^{(\mathcal{D})}$, and W3C's *Ontology for Media Resources* to $\mathcal{ALCHI}^{(\mathcal{D})}$. While the *Core Ontology for Multimedia* (COMM) uses $\mathcal{SHOIN}^{(\mathcal{D})}$, the most expressive DL of the time of its release (in OWL DL), the current standard, OWL 2 DL, supports far more constructs, such as complex role inclusion axioms, reflexive, asymmetric, irreflexive, and disjoint roles, universal role, self-constructs, negated role assertions in Aboxes, and qualified number restrictions. The first and, to our knowledge, the only multimedia ontology that exploits the full range of OWL 2 constructs ($\mathcal{SROIQ}^{(\mathcal{D})}$ DL) is the VidOnt ontology, which covers the professional video production and broadcasting domains [19].

4. CONCLUSIONS

Text-based video metadata annotations provide unstructured data with limitations in automated processing. Manually added metadata, such as YouTube tags, have problems with tag variation, polysemy, misspelling, compound tags, ambiguity, subjectivity, and the relevance of the tag corresponding to the content. With the growing number of multimedia ontologies, the machine-readable structured video annotations still have open issues. While the direct mapping of application-specific XML schemas to domain-specific machine-interpretable metadata terms seems promising, the MPEG-7 XSD to OWL mappings released so far use OWL DL, rather than the more expressive OWL 2 DL. The lack of role box definitions prevents advanced inference and reasoning. The structural complexity issues of MPEG-based ontologies can be addressed by introducing comprehensive yet transparent, open access, and well-documented multimedia ontologies with advanced reasoning support. The explicit formalization of the intended semantics is required to address semantic interoperability issues. The semantically equivalent descriptors representing the same information should be defined with one class or entity only. Addressing these issues could provide the formal semantics in video annotations necessary for efficient semantic searches and factual information retrieval from the Google Knowledge Graph, or programmatic access via SPARQL endpoints, as well as reasoning over audiovisual contents. To improve semantic video indexing, the standardization and global adaption of RDF-powered multimedia annotators in Content Management Systems and video editing tools would be desirable.

5. REFERENCES

[1] Blöhdorn, S., Petridis, K., Saathoff, C., Simou, N., Tzouvaras, V., Avrithis, Y., Handschuh, S., Kompatsiaris, Y., Staab, S., and Strintzis, M. Semantic Annotation of Images and Videos for Multimedia Analysis. *Lect Notes Comput Sc*, *3532*. 592-607. DOI= http://dx.doi.org/10.1007/11431053_40.

[2] Boll, S., Klas, W., Sheth, A. Overview on using metadata to manage multimedia data. In: Sheth, A., Klas, W. (eds.) *Multimedia data management: Using metadata to integrate and apply digital media*. McGraw-Hill, New York, 1998, 3.

[3] Choudhury, S., Breslin, J. G., and Passant, A. Enrichment and Ranking of the YouTube Tag Space and Integration with the Linked Data Cloud. *Lect Notes Comput Sc*, *5823*. 747–762. DOI=http://dx.doi.org/10.1007/978-3-642-04930-9_47.

[4] Consens, M. P., Hassanzadeh, O., and Teisanu, A. M., 2014. Linked Movie Database. Retrieved 12 June 2015, from http://www.linkedmdb.org.

[5] Dasiopoulou, S., Tzouvaras, V., Kompatsiaris, I., and Strintzis, M. G. Enquiring MPEG-7 based multimedia ontologies. *Multimed Tools Appl*, *46*. 331–370. DOI=http://dx.doi.org/10.1007/s11042-009-0387-4.

[6] Dasiopoulou, S., Tzouvaras, V., Kompatsiaris, I., and Strintzis, M. Capturing MPEG-7 semantics. In *2nd International conference on metadata and semantics*, (Corfu, Greece, 2007).

[7] Adams, G. (ed.), Dolan, M., Freed, G., Hayes, S., Hodge, E., Kirby, D., Michel, T., Singer, D. Timed Text Markup Language 1. Retrieved 12 June 2015, from World Wide Web Consortium: http://www.w3.org/TR/ttaf1-dfxp/.

[8] ETSI, 2015. TV Anytime. Retrieved 12 June 2015, from European Telecommunications Standards Institute: http://www.etsi.org/technologies-clusters/technologies/broadcast/tv-anytime.

[9] García, R. and Celma, O. Semantic Integration and Retrieval of Multimedia Metadata. In *5th Int. Workshop on Knowledge Markup and Semantic Annotation*, (Galway, Ireland, 2005).

[10] Gómez-Romero, J., Patricio, M. A., García, J., and Molina, J. M. Ontology-based context representation and reasoning for object tracking and scene interpretation in video. *Expert Syst Appl*, *38*(6). 7494–7510. DOI=http://dx.doi.org/10.1016/j.eswa.2010.12.118.

[11] Hunter, J. Adding Multimedia to the Semantic Web—Building an MPEG-7 Ontology. In *1st International Semantic Web Working Symposium*, (Stanford, USA, 2001), 261–281.

[12] IPTC, 2015. NewsML-G2. Retrieved 12 June 2015, from International Press Telecommunications Council: https://iptc.org/standards/newsml-g2/.

[13] Isaac, A. and Troncy, R. Designing and using an Audio-Visual Description Core Ontology. *Workshop on Core Ontologies in Ontology Engineering*, (Northamptonshire, UK, 2004).

[14] ISO, 2013. MPEG-7. ISO/IEC 15938. Retrieved 12 June 2015, from International Organization for Standardization: http://www.iso.org/iso/home/store/catalogue_tc/catalogue_detail.htm?csnumber=34230.

[15] ISO, 2013. MPEG-21. ISO/IEC 21000. Retrieved 12 June 2015, from International Organization for Standardization: http://www.iso.org/iso/home/store/catalogue_tc/catalogue_detail.htm?csnumber=35367.

[16] Oberle, D., Ankolekar, A., Hitzler, P., Cimiano, P., Sintek, M., Kiesel, M., Mougouie, B., Baumann, S., Vembu, S., and Romanelli, M. DOLCE ergo SUMO: on foundational and domain models in the SmartWeb integrated ontology (SWIntO). *J Web Semantics*, *5*(3). 156–174. DOI=http://dx.doi.org/10.1016/j.websem.2007.06.002.

[17] Sikos, L. F. Advanced (X)HTML5 metadata and semantics for Web 3.0 videos. *DESIDOC Libr Inf Technol*, *31*(4). 247–252. DOI=http://dx.doi.org/10.14429/djlit.31.4.1105.

[18] Sikos, L. F. *Mastering Structured Data on the Semantic Web: From HTML5 Microdata to Linked Open Data.* Apress Media, New York, 2015.

[19] Sikos, L. F., 2015. VidOnt: The Video Production and Broadcasting Ontology. Retrieved 12 June 2015, from http://vidont.org.

[20] Solla, A. G. and Bovino, R. G. S. *TV-Anytime: Paving the Way for Personalized TV.* Springer Berlin Heidelberg, Berlin, 2013. DOI=http://dx.doi.org/10.1007/978-3-642-36766-3.

[21] Suárez-Figueroa, M. C., Atemezing, G. A., and Corcho, O. The landscape of multimedia ontologies in the last decade. *Multimed Tools Appl*, *62*(2). 377–399. DOI=http://dx.doi.org/10.1007/s11042-011-0905-z.

[22] Tsinaraki, C., Polydoros, P., Moumoutzis, N., and Christodoulakis, S. Integration of OWL ontologies in MPEG-7 and TV-Anytime compliant Semantic Indexing. *Lect Notes Comput Sc*, *3084*. 398–413. DOI=http://dx.doi.org/10.1007/978-3-540-25975-6_29.