INTEGRATION OF STRUCTURAL AND SEMANTIC MODELS FOR MULTIMEDIA METADATA MANAGEMENT

Suzanne Little[*†], *Massimo Martinelli, Ovidio Salvetti*
ISTI-CNR
Pisa, Italy
Firstname.Lastname@isti.cnr.it

Uğur Güdükbay, Özgür Ulusoy
Bilkent University
Dept of Computer Engineering
Bilkent, Ankara, Turkey

Gaël de Chalendar, Gregory Grefenstette
CEA List
Centre de Fontenay-aux-Roses
France

ABSTRACT

The management and exchange of multimedia data is challenging due to the variety of formats, standards and intended applications. In addition, production of multimedia data is rapidly increasing due to the availability of off-the-shelf, modern digital devices that can be used by even inexperienced users. It is likely that this volume of information will only increase in the future. A key goal of the MUSCLE (Multimedia Understanding through Semantics, Computation and Learning) network is to develop tools, technologies and standards to facilitate the interoperability of multimedia content and support the exchange of such data. One approach for achieving this was the creation of a specific "E-Team", composed of the authors, to discuss core questions and practical issues based on the participant's individual work. In this paper, we present the relevant points of view with regards to sharing experiences and to extracting and integrating multimedia data and metadata from different modes (text, images, video).

1. INTRODUCTION

The management and exchange of multimedia data is a challenging area of research due to the variety of data and the diversity of intended applications. Many research groups are investigating and developing solutions or standards to promote the interoperability of multimedia data within and between groups, organisations and application domains. The challenge lies in producing multimedia metadata to support interoperability, exchange and enable sophisticated semantic search and retrieval.

Within MUSCLE research is focusing on standards, technologies and techniques for integrating, exchanging and enhancing the use of multimedia within a variety of research areas. "E-Teams" have been organised to collaborate, discuss and combine research and expertise. This article describes work being undertaken by participants in the E-Team titled "Integration of Structural and Semantic Models for Multimedia Metadata Management" and discusses how this work addresses the issue of multimedia integration and exchange.

To utilise the diverse areas of interest and expertise within the E-Team we plan to discuss the difficulties in extracting and integrating multimedia data and metadata from different media and modes. Through this we aim to achieve a better understanding of the semantic models used within this group and the requirements for integration and dissemination of media.

The broad questions we intend to investigate are:

1. What are the different requirements for recording and storing media?
2. What are the outcomes/outputs from analysing different media?
3. What is the analysis process/workflow used for the media?
4. What standards are used? What are their limitations or strengths?
5. How are annotations defined and used? Specifically, what type of annotations and how are they captured or extracted?

In the remainder of this paper, section 2 looks at related work in the field of multimedia metadata interoperability and exchange, summarising briefly some of the most relevant standards and technologies and discussing other similar frameworks and architectures. Section 3 examines the individual projects of the E-Team participants focussing on their classification according to media type, outcome, intended use and the standards and technologies applied. Section 4 discusses the challenges of integrating multimodal, multimedia data and metadata using the projects within the E-Team to investigate the limitations and possible approaches. Section 5 outlines the future plans of the activity and the

[*]Correspondence author at *Suzanne.Little@isti.cnr.it*

[†]This work was carried out during the tenure of a MUSCLE Internal Fellowship.

outcomes of this virtual collaboration.

2. RELATED WORK

Collections of multimedia data can be used for many different purposes. Therefore systems which manage multimedia and its metadata need to support a variety of functionalities. These include: high-level semantic searching; low-level feature and statistical analysis; semantic grouping; semi-automatic identification of semantic relationships within and between media and capture of provenance and bibliographic metadata about the media. Multimedia metadata models are therefore multi-layered and have both syntactic and semantic facets (Figure 1). This section describes standards and other projects that model multimedia data to support some of these functionalities.

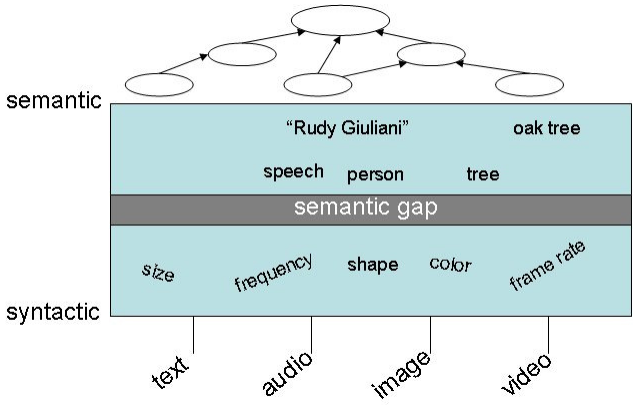

Fig. 1. Correlation of Data and Semantics

The simplest type of syntactic interoperability of multimedia metadata can be achieved through the use of a number of standards and protocols such as Dublin Core [1], MPEG-7 [2], MPEG-21 [3], CIDOC-CRM [4] etc. These often originate with in the digital libraries domain and aim to define a syntax either through a high-level model or though specific schema in formats such as XML. The application of such schema can be found within projects such as the DELOS project [5] and protocols such as the Open Archives Initiative (OAI) [6, 7]. OAI uses a simple Dublin Core based syntax and proscribes a protocol for making descriptions of digital media representations available for harvesting. This enables aggregation services to query distributed collections of multimedia metadata.

Higher level but still relatively generic semantic models may be based upon the models defined by standards such as MPEG-7. MPEG-7 based ontologies enable higher level models of multimedia types (Image, Video, Audio etc.), structures (Segment, StillRegion etc.) and features (DominantColor, ColorHistogram etc.) to be applied within systems. Previous work by Hunter [8], by Tsinaraki et al. [9] and by Garcia et al. [10] has provided direct translations of segments of the MPEG-7 standard into semantic web formats such as OWL.

The semantic gap (marked on Figure 1) is defined as "the discrepancy between the information that one can extract from the visual data, and the interpretation that the same data has for a user" [11]. Many projects have aimed to overcome or mitigate this gap in multimedia data. Some of these have specifically used multimedia or semantic models while others have focused on lower-level, machine-learning based techniques to identify patterns or relationships. Hollink et al. [12] describes an analysis process for labelling art works. While work by Hollink, Little et al. [13] presents an evaluation of a technique called semantic inferencing rules that explicitly relate low-level MPEG-7 features to semantic terms from a domain ontology for scientific images. Dorado et al. [14] combine features such as color, texture and shape with keyword mining technique to perform semantic labelling of images. Recently, Hare et al. [15] and Vembu et al. [16] have presented broad approaches for bridging the semantic gap using ontologies.

Beyond bridging the semantic gap, many projects have used and applied multimedia models to enable richer semantic search, discovery and exchange of media data. These projects often propose a multimedia semantic framework to organise, analysis, combine and manage multimedia data and provide advance semantic querying functionalities among others. Recent work includes [17, 18, 19, 20, 21].

The topic of exploiting multimedia content within the semantic web has also been the focus of research with the chartering of a W3C Incubator Group [22] to discuss issues relating to multimedia integration using semantic web technologies. In addition Van Ossenbruggen et al. [23, 24] discuss some of the specific requirements for integrating and applying multimedia within a semantic web infrastructure. Stamou et al. [25] summarises techniques and standards for integrating multimedia on the semantic web.

These projects provide a range of functionality and support interoperability through the use of standards and semantic models. However the systems are generally presented independently although they are often intended to support integration. Within this activity we aim to explore how the different technical, syntactic and semantic requirements of independent systems for multimedia metadata management and analysis effect their integration.

3. PARTICIPANT'S CONTRIBUTIONS

As part of this project we will discuss the different syntactic and semantic models used by each of the participants. We aim to establish the different modelling requirements for each project, the approaches used and how these models

can interact and relate to form wider networks of multimedia and metadata.

This section describes the individual work and focuses on the following questions:

- What type of media does the project use and what is the main domain of evaluation?
- What is the outcome or product of the project? (E.g., architecture, standards, web-based app, stand alone app etc.)
- What is the goal of the project? What is the main service it aims to provide or support? (E.g., semantic annotation, tagging, search and retrieval, analysis, archival etc.)
- What technology and standards does the project use?

3.1. Bilkent University

Project Summary: At Bilkent, a prototype video database management system, called BilVideo is developed [26]. The system architecture of BilVideo is original in that it provides full support for spatio-temporal queries that contain any combination of spatial, temporal, object appearance, external predicate, trajectory projection, and similarity-based object trajectory conditions by a rule-based system built on a knowledge-base, while utilizing an object-relational database to respond to semantic (keyword, event/activity, and category-based), color, shape, and texture queries. The knowledge-base of BilVideo contains a fact-base and a comprehensive set of rules implemented in Prolog. The rules in the knowledge-base significantly reduce the number of facts that need to be stored for spatio-temporal querying of video data.

A Web-based visual query interface is currently being used to query videos[1]. BilVideo can handle multiple requests over the Internet via a graphical query interface [27]. An NLP-based interface also exists to allow users to formulate queries as sentences in English [28].

Media Type: Video

Intended outcome or product: A prototype video DBMS

Services provided or functions supported: BilVideo supports spatio-temporal, semantic, color, shape, and texture queries in an integrated manner.

Technology and standards: MPEG-7

3.2. CEA List

Project Summary: The CEA LIST is involved within MUSCLE and within a national project called WebContent[2] in creating tools for adding semantic annotation to raw data. In a platform providing a means to combine various semantic web technologies, we are developing web services to build and enrich OWL ontologies from text corpora, to annotate texts with concepts and relations from ontologies and finally to navigate through these semantically annotated documents.

Media Type: Text, and then images and text.

Intended outcome or product: General applications involving Watch (Technology Watch, Strategic Watch, Event Watch, etc). A Watch system adds additional markup to certain watch specific entities and events in a flow of data, inplace or out as RDF annotations. The identified information can also be extracted from the input stream and presented in tabular form.

Services provided or functions supported: Given an input ontology, describing the objects and events of interest in an application domain, the CEA LIST technology will watch streams of text and identify those ontology-related items in the input text. Depending on the client application, the identified items will be added as XML-interpretable semantic markup to the input stream or they will annotate the document through RDF triples or they will produce new individuals added to the input ontology.

Technology and standards: The technologies used are natural language processing tools (tokenization, morphological analysis, syntactic analysis, semantic annotation) and semantic web (OWL, RDF, Web Services).

3.3. ISTI

Project Summary: At CNR ISTI, we are developing an infrastructure for MultiMedia Metadata Management (4M) [29] to support the integration of media from different sources. This infrastructure enables the collection, analysis and integration of media for semantic annotation, search and retrieval. The challenge is to provide an infrastructure that enables disparate groups to combine and disseminate multimedia research data. The achievement of this goal requires the use of standards and the development of tools to assist in the extraction and conversion of multimedia metadata.

Media Type: images, audio, video (partial support)

Intended outcome or product: architecture, prototype

Services provided or functions supported: automatic standardised analysis to produce MPEG-7 descriptions in XML format; similarity search based on MPEG-7 features

Technology and standards: MPEG-7, XML, eXist database (extensions for access control)

4. INTEGRATION AND INTEROPERABILITY OF MULTIMODAL MULTIMEDIA DATA

Figure 2 shows a possible amalgamation of the three individual multimedia systems described in the previous sections to enable a single, integrated querying interface. Each of the systems use different media modalities and provide

[1] BilVideo Web Client is available at `http://www.cs.bilkent.edu.tr/~bilmdg/bilvideo`

[2] `http://www.webcontent.fr`

different but related functionality. The type of metadata produced by each system is also quite different. ISTI's system produces very low-level feature analysis metadata, CEA identifies general concepts (e.g., person, place) within a text stream and BilVideo provides interfaces that support high-level semantic queries. Enabling the systems to be accessed in a standard, transparent fashion while still retaining their strengths and independence will enable advanced semantic functionalities and the interoperability of multimedia data.

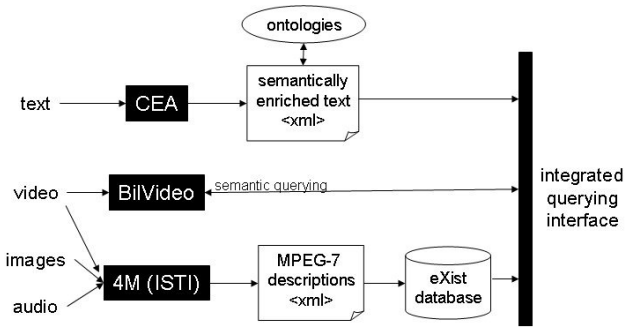

Fig. 2. Possible Integration of Multimedia Systems

This section discusses the challenges faced by the E-Team when discussing how these independent systems and approaches can be combined and networked to provide richer interactions over a broad range of media types. Section 4.2 presents some possible approaches to exploiting standards to integrate the systems both syntactically and semantically.

4.1. Challenges

There are three main challenges that need to be addressed by the E-Team's participants. Firstly the syntactic integration of the metadata produced by the independent projects. This involves, for example, converting between MPEG-7 formatted data produced by ISTI's 4M architecture and the internal knowledge-format used within BilVideo. Bilkent plans to develop an automatic MPEG-7 feature (Color, Shape, and Texture) extraction tool for videos. The output of this tool should also be converted into BilVideo knowledge-base format. This is necessary to make feature-based querying of videos and integrate all available metadata.

Secondly, the construction of an integrated querying interface to exploit and relate all of the media and metadata produced by the systems. This raises technical challenges based on the compatibility of formats and systems (relational vs xml databases) and the network architecture (centralised vs de-centralised). This interface would enable queries to be conducted across all systems and modes and would be useful for identifying relationships between media objects.

For example, a news report about former New York mayor, Rudy Giuliani, could be analysed by CEA which identifies the person "Rudy Giuliani" and the place "New York" and associates some representative images with the terms. This information could then be used to query collections managed by BilVideo and analysed by the 4M architecture to find further related media objects.

Thirdly, the largest challenge is determining and applying techniques for overcoming the semantic gap between the low-level feature data produced by the 4M (ISTI) system, the mid-level semantic enrichment provided by CEA's system and the high-level semantic querying capabilities supported by BilVideo and required for the general interface. Addressing this issue will enable more sophisticated semantic functionalities to be supported and improve the general applicability of the systems.

4.2. Possible Approaches

While this work is still in a preliminary stage, some possible approaches and relevant technologies for addressing the challenges have been discussed.

Syntactic interoperability between the systems can be achieved through the use of standards such as XML and RDF from the semantic web domain. This will facilitate the development of convertors and interfaces between the systems and the various metadata output formats used. At the semantic level ontologies that define similarities and relationships between terms can be useful to convert between the BilVideo knowledge-base, CEA's markup and 4M's MPEG-7 descriptions.

A key initial step is to implement tools that support the transformation of or provide wrapper interfaces to media and metadata from each of the systems. This will enable information to be more easily exchanged and analysed and facilitate the development of a general interface. The use of standards, such as MPEG-7, will also be investigated. By applying MPEG-7 across these systems and the media modes and domains supported we hope to evaluate its suitability for use within a general multimedia integration system.

Finally, techniques for identifying low-level patterns (color, shape modelling, feature descriptors) within different media types, describing these patterns and linking them with semantic terms will be investigated (e.g., [30, 31]). This may involve exploring the use of analysis algorithms in conjunction with semantic models described in domain ontologies. Additionally the interfaces, knowledge management and reasoning capabilities provided by BilVideo could be exploited to provide feature sets or correlations to be applied to data from CEA or 4M.

5. FUTURE WORK AND CONCLUSIONS

Section 1 presented a list of five general questions that are being addressed within this activity. We aim to identify the different and diverse requirements, outcomes, processes, standards and purposes of multimedia metadata management and analysis systems using the integration of our independent systems.

This paper has described the current work being explored by this MUSCLE E-Team. It has presented an outline of each of the systems, focussing on the media profiles, supported functionalities and technologies used. The main challenges faced when aiming to integrate the syntactic and semantic models used in these diverse applications have been discussed and possible approaches to these challenges have been presented.

This MUSCLE E-Team is in a unique position. The participants have a broad range of expertise and the systems each use media of different modes and provide distinctive functionalities. The syntactic and semantic integration of these systems raises and aims to address significant issues in the interoperability and exchange of multimedia content within different organisations.

6. ACKNOWLEDGEMENTS

This work has been supported by EU MUSCLE Network of Excellence.

7. REFERENCES

[1] Dublin Core Community. Dublin core metadata initiative. http://dublincore.org/.

[2] ISO/IEC 15938-5 FDIS Information Technology. Mpeg-7: Multimedia content description. See http://www.chiariglione.org/mpeg/standards/mpeg-7/mpeg-7.htm.

[3] ISO/IEC 15938-5 FDIS Information Technology. Mpeg-21. See http://www.chiariglione.org/mpeg/standards/mpeg-21/mpeg-21.htm.

[4] ICOM-CIDOC Data Model Working Group. Conceptual reference model, cidoc/crm. http://cidoc.ics.forth.gr/.

[5] DELOS - Network of Excellence on Digital Libraries. See http://www.delos.info/.

[6] Open Archives Initiative Community. Open archives initiative. http://www.openarchives.org/.

[7] Open Archives Initiative Community. Oai protocol for metadata harvesting 2.0. http://www.openarchives.org/OAI/openarchivesprotocol.html.

[8] Jane Hunter. Adding multimedia to the semantic web – building an mpeg-7 ontology. In *International Semantic Web Working Symposium (SWWS)*, Stanford, 2001.

[9] C. Tsinaraki, P. Polydoros, and S. Christodoulakis. Interoperability support for Ontology-based Video Retrieval Applications. In *Proc. of 3rd International Conference on Image and Video Retrieval (CIVR 2004)*, Dublin, Ireland, July 2004.

[10] R. Garcia and O. Celma. Semantic integration and retrieval of multimedia metadata. In *Proc. of the 5th International Workshop on Knowledge Markup and Semantic Annotation (SemAnnot 2005)*, Galway, Ireland, November 2005.

[11] A.W.M. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain. Content-based image retrieval at the end of the early years. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(12), 2000.

[12] L. Hollink, A.Th. Schreiber, J. Wielemaker, and B. Wielinga. Semantic annotation of image collections. In *KCAP'03 Workshop on Knowledge Capture and Semantic Annotation*, Florida, USA, 2003.

[13] Laura Hollink, Suzanne Little, and Jane Hunter. Evaluating the application of semantic inferencing rules to image annotation. In *Proceedings of the Third International Conference on Knowledge Capture, KCAP05*, Banff, Canada, 2005.

[14] A. Dorado and E. Izquierdo. Semantic labeling of images combining color, texture and keywords. In *Proceedings of the International Conference on Image Processing (ICIP2003)*, volume 3, September 2003.

[15] J. S. Hare, P. A. S. Sinclair, P. H. Lewis, K. Martinez, P. G. B. Enser, and C. J. Sandom. Bridging the semantic gap in multimedia information retrieval: Top-down and bottom-up approaches. In P. Bouquet, R. Brunelli, J. P. Chanod, C. Niedere, and H. Stoermer, editors, *Proceedings of Mastering the Gap: From Information Extraction to Semantic Representation / 3rd European Semantic Web Conference*, Budva, Montenegro, 2006.

[16] Shankar Vembu, Malte Kiesel, Michael Sintek, and Stephan Bauman. Towards bridging the semantic gap in multimedia annotation and retrieval. In *First International Workshop on Semantic Web Annotations for Multimedia (SWAMM2006)*, 2006.

[17] Stephan Bloehdorn, Kosmas Petridis, Carsten Saathoff, Nikos Simou, Vassilis Tzouvaras, Yannis Avrithis, Siegfried Handschuh, Yiannis Kompatsiaris, Steffen Staab, and Michael G. Strintzis. Semantic annotation of images and videos for multimedia analysis. In Asuncin Gmez-Prez and Jrme Euzenat, editors, *The Semantic Web: Research and Applications: Proceedings of the Second European Semantic Web Conference, ESWC 2005, Heraklion, Crete, Greece, May 29-June 1, 2005*, volume 3532 of *Lecture Notes in Computer Science*, pages 592–607. Springer, MAY 2005.

[18] S. Dasiopoulou, V. K. Papastathis, V. Mezaris, I. Kompatsiaris, and M. G. Strintzis. An ontology framework for knowledge-assisted semantic video analysis and annotation. In *Proc. 4th International Workshop on Knowledge Markup and Semantic Annotation (SemAnnot 2004) at the 3rd International Semantic Web Conference (ISWC 2004)*, November 2004.

[19] Jane Hunter and Suzanne Little. A framework to enable the semantic inferencing and querying of multimedia content. *International Journal of Web Engineering and Technology – Special Issue on the Semantic Web*, 2(2/3):pp 264–286, December 2005.

[20] Kosmas Petridis, Stephan Bloehdorn, Carsten Saathoff, Nikos Simou, Stamatia Dasiopoulou, Vassilis Tzouvaras, Siegfried Handschuh, Yannis Avrithis, Yiannis Kompatsiaris, and Steffen Staab. Knowledge representation and semantic annotation of multimedia content. *IEEE Proceedings on Vision, Image and Signal Processing - Special issue on the Integration of Knowledge, Semantics and Digital Media Technology*, 153(3):255–262, JUN 2006.

[21] Yves Raimond, Samer A. Abdallah, Mark Sandler, and Mounia Lalmas. A scalable framework for multimedia knowledge management. In *Proceedings of the First international conference on Semantics And digital Media Technology (SAMT2006)*, pages 11–25, December 2006.

[22] W3C. Incubator group on multimedia semantics. See `http://www.w3.org/2005/Incubator/mmsem/`, January 2007.

[23] Jacco van Ossenbruggen, Frank Nack, and Lynda Hardman. That obscure object of desire: Multimedia metadata on the web, part 1. *IEEE MultiMedia*, 11(4):38–48, 2004.

[24] Frank Nack, Jacco van Ossenbruggen, and Lynda Hardman. That obscure object of desire: Multimedia metadata on the web, part 2. *IEEE MultiMedia*, 12(1):54–63, 2005.

[25] Giorgos Stamou, Jacco van Ossenbruggen, Jeff Z. Pan, and Guus Schreiber. Multimedia annotations on the semantic web. *IEEE MultiMedia*, 13(1):86–90, 2006.

[26] M. E. Dönderler, E. Şaykol, U. Arslan, Ö. Ulusoy, and U. Güdükbay. BilVideo: Design and Implementation of a Video Database Management System. *Multimedia Tools and Applications*, 27(1):pp.79–104, 2005.

[27] E. Şaykol, U. Güdükbay, and Ö. Ulusoy. Integrated Querying of Images by Color, Shape, and Texture Content of Salient Objects. In Tatyana Yakhno, editor, *Lecture Notes in Computer Science (LNCS), Vol. 3261, Advances in Information Sciences (ADVIS'2004)*, pages pp. 363–371. Springer-Verlag, 2004.

[28] O. Küçüktunc, U. Güdükbay, and Ö. Ulusoy. A Natural Language Based Interface for Querying a Video Database. *IEEE Multimedia*, 14(1):pp.83–89, 2007.

[29] Patrizia Asirelli, Suzanne Little, Massimo Martinelli, and Ovidio Salvetti. MultiMedia Metadata Management: a Proposal for an Infrastructure. In *Workshop on Semantic Web Applications and Perspecitives (SWAP06)*, Pisa, December 2006.

[30] P Perner and S Jnichen. *Structural, Syntactic and Statistical Pattern Recognition*, chapter Learning of Form Models from Exemplars, pages 153–161. Springer Verlag, 2004.

[31] P Perner, H Perner, and B Mller. *Advances in Case-Based Reasoning*, chapter Similarity Guided Learning of the Case Description and Improvement of the System Performance in an Image Classification System, pages 604–612. Springer Verlag, 2002.