

The POLYSEMA MPEG-7 Video Annotator

George Valkanas, Vassileios Tsetsos and Stathes Hadjiefthymiades
 Pervasive Computing Research Group, Department of Informatics and Telecommunications,
 University of Athens, Panepistimiopolis, Ilissia, 15784, Greece
 {g.valkanas, b.tsetsos, shadj}@di.uoa.gr

Abstract— Annotation of multimedia content with metadata is regarded as a key issue for efficiently handling such content. MPEG-7 is a major standard that has been developed for this reason, supporting the description of both low- and high-level multimedia content features. Low-level feature annotation may be automated in many ways, contrary to high-level annotation which is a more complex task that requires human intervention. Therefore, high-level semantic annotation needs to be simplified as a process, mainly through annotation tools. In this paper we describe such a tool, which facilitates expressive video annotation process by combining MPEG-7 metadata and semantic web ontologies.

Index Terms— video annotation tool, MPEG-7, ontologies

I. INTRODUCTION

The recent popularity of multimedia-based applications (e.g., Web 2.0, interactive TV), calls for more advanced and efficient techniques for searching and managing multimedia content. A standard means to achieve this is through multimedia metadata. A lot of metadata standards have been proposed for describing functional and non-functional multimedia semantics. One of the most complete standards is the MPEG-7 [1], formally known as “*Multimedia Content Description Interface*”, developed by the Moving Picture Experts Group. The standard supports the description of both low-level features and high-level semantics, for different aspects of the content, using XML to store the metadata. However, mainly due to the complexity of MPEG-7, only a few tools have been developed for the creation of MPEG-7 multimedia descriptions. Another current trend is using Semantic Web technologies and ontologies for semantic multimedia annotations [8].

In the area of audiovisual content annotation, which this paper focuses on, two of the most well-known solutions are VideoAnnex [5] and Frameline 47 [6]. Both tools adopt MPEG-7 for representing metadata, but none of them provides adequate support for describing high-level semantics nor support for ontology-based annotation. In this paper, we briefly describe a new video annotation tool that relies on MPEG-7 and exploits Semantic Web ontologies as controlled annotation vocabularies. This tool has been developed in the context of the POLYSEMA platform for semantics-based interactive TV services [4].

II. POLYSEMA VIDEO ANNOTATOR

As implied by its name, the POLYSEMA MPEG-7 Video Annotator performs MPEG-7-based annotation of video content through an easy-to-use Graphical User Interface (GUI).

Specifically, it supports high-level annotation of the multimedia, based on Part 5 (Multimedia Description Scheme, MDS) of the MPEG-7 standard. High-level annotation refers to metadata that describe high-level concepts instead of structural, visual and audio features. Currently, the tool takes as input an existing MPEG-7 document, that is assumed to contain basic information about the video segments (i.e., start time and duration), possibly obtained from a shot segmentation tool or algorithm.

The tool supports annotation both for the entire video and for discrete video segments. Based on the type of annotation, different kinds of information may be stored. When annotating the entire video, basic information (e.g., title, abstract), and creation information (e.g., creation date and place, contributors, genre, parental rating) may be inserted (see Figure 1). In order to further increase the expressiveness and formality of the annotation, the tool uses RDF(S)/OWL domain ontologies as controlled vocabularies. We provide some predefined ontologies (e.g., TV-Anytime genre and parental rating classification schemes), but any domain ontology may be used in order to facilitate the annotation process.

The screenshot shows the POLYSEMA MPEG-7 Annotator GUI in the 'Entire Video Annotation' mode. The interface is organized into several functional areas:

- General Video Information:** Includes fields for Title (The Best of Mr. Bean), Abstract (The Best of Mr. Bean), Free Text (This is a comedy show, where the best films from Mr. Bean (Rowan Atkinson) are compiled into one! The compilation is hilarious and contains indeed some of the best of Mr. Bean's shows. Recommended for people of all ages and ethnicities!), Structured, and Keyword.
- Media creation information:** Includes Date (18 / 8 / 2003), Place (London), and Country (United Kingdom).
- Creators List:** Lists Rowan Atkinson and Ben Elton with a Role field and buttons for Add and Remove.
- Movie Classification Information:**
 - Genre:** main: Humor, secondary: Pure_entertainment.
 - Subject:** This is a humor, pure entertainment show, with Mr. Bean. It's for the whole family, and for people of all ages. The show is hilarious, and makes you laugh a lot, since the best from "Mr. Bean"'s show have been put into one. Enjoy!
 - Release Date:** 16 / 8 / 2004, Country: United Kingdom.
 - Parental Guidance:** Parental rating (set to General_audiences) and Minimum Age.
 - Countries:** Afghanistan, Albania, Algeria, Andorra, Angola, Antigua and Barbuda.
- Related Material:** Includes a URL (http://www.imdb.com) and a list of Material Items (actor biography, general info, actor biography).

Figure 1: Entire Video Annotation

In the “video segment annotation” mode (see Figure 2), the human annotator is allowed to describe segments with one of the supported ways: *FreeText*, *Structured*, and *Keyword* annotation. Controlled vocabularies are supported here as well, through domain ontologies (in the case of Structured annotation). The use of such vocabularies is recommended by the current trends in multimedia semantics annotation [8], but it is not mandatory. The tool’s Structured annotation supports

all the elements specified in MPEG-7: *Who*, *WhatObject*, *WhatAction*, *Where*, *When*, *Why*, *How*.

While annotating video segments, the user may simultaneously playback the video, in the “Video Section” area. The *Java Media Framework* (JMF) [2] has been used for media presentation. If both an MPEG-7 document with segmentation information and the corresponding video file are loaded, the user may view the video snapshots in the “Sections” area. Each snapshot is the first frame, and, thus, signifies the beginning, of a video segment. In order to facilitate the navigation between video segments, “next” and “previous” buttons exist, while direct access to a segment’s metadata is also possible, either by inserting the video segment number in the designated area or by selecting the corresponding snapshot.

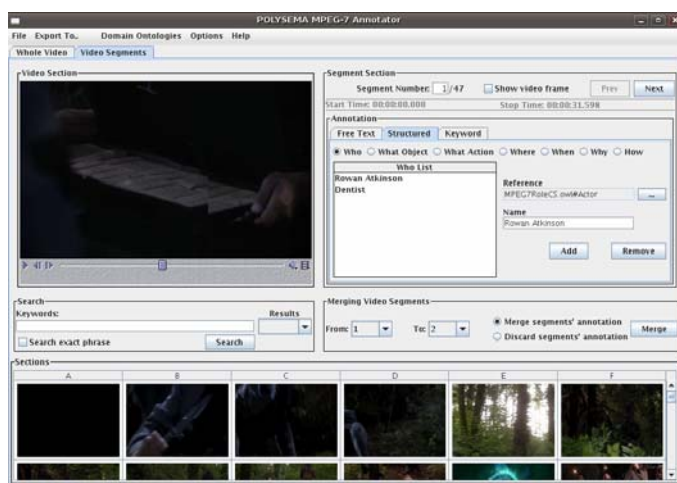


Figure 2: Video Segment Annotation

Apart from the annotation functionality of the tool, two more functions are supported: searching and merging video segments. The tool allows the user to perform a search based on keywords or phrases. Once the user has inserted the keywords or phrase, the tool searches among the video segment annotations and presents the user with the video segment numbers that are annotated with these keywords.

The second function, allows the merging of multiple sequential video segments into one logical segment. The user may select whether to merge the annotations as well, or to discard them. In the “Merging Video Segments” GUI section, the user selects the beginning/ending video segments and whether the annotations will be also merged, and the tool performs the action. This functionality is important, because, as already mentioned, the input file is a result of another video segmentation tool or algorithm, which may create segments based on visual features rather than content semantics.

Once the user has finished annotating the video, they may serialize the descriptions to an MPEG-7 XML file. Moreover, the tool allows the user to use the metadata for populating an MPEG-7 OWL ontology developed in the context of the project POLYSEMA. This ontology is based on [9] and contains a subset of the MPEG-7 MDS elements. Through

such ontological representation of MPEG-7 metadata, they can be more easily exploited in other knowledge-based applications, such as rule-based personalization and recommendation systems.

III. CONCLUSION AND DISCUSSION

We described a new annotation tool that addresses some limitations of existing video annotation tools. The main contributions of this work can be summarized to:

- Integration of third-party Semantic Web ontologies into the video annotation process
- Support for the most practical and useful elements of the MPEG-7 MDS specification
- A user-friendly GUI that reduces the annotation effort
- Exportation of created MPEG-7 metadata both in XML and OWL formats.

As future work we plan to integrate unsupervised tools for semantics extraction (e.g., [7]) as well as techniques for knowledge extraction from Web resources, e.g., Internet Movie Database (IMDB).

Finally, we should note that the current version of the tool does not support the most expressive high-level semantic annotation defined in the MPEG-7 Semantic Description Scheme (expressed through graphs similar to semantic networks). This is a design decision taken for two main reasons:

a) There are no commonly agreed guidelines regarding *what* the semantic graphs should model or *how* should they be structured. Therefore, their exploitation in real-world applications (e.g., multimedia retrieval systems) is limited.

b) Manual creation of expressive descriptions is a very demanding process that deters most human annotators from performing it.

ACKNOWLEDGEMENT

This work was partially funded by the Greek General Secretariat for Research and Technology (GSRT) under the Operational Program “Information Society”, co-funded by the EU.

REFERENCES

- [1] Manjunath, B.S., Salembier, P., and Sikora, T. *Introduction to MPEG-7, Multimedia Content Description Interface*, John Wiley & Sons Ltd, 2002
- [2] Java Media Framework, <http://java.sun.com/products/java-media/jmf/index.jsp>
- [3] <http://en.wikipedia.org/wiki/MPEG-7>
- [4] Papadimitriou A., Anagnostopoulos C., Tsetsos V., Paskalis S., Hadjiefthymiades S. "A Semantics-aware Platform for Interactive TV Services", I-MEDIA '07 Conference, Graz, Austria, September 2007
- [5] IBM VideoAnnex tool, <http://www.research.ibm.com/VideoAnnEx/>
- [6] Frameline 47, <http://www.frameline.tv/>
- [7] Giannakopoulos T., Pikrakis A., Theodoridis S. "A Multi-class Audio Classification Method with Respect to Violent Content in Movies using Bayesian Networks", IEEE International Workshop on Multimedia Signal Processing, Crete, Greece, October, 2007
- [8] W3C Multimedia Semantics Incubator Group, <http://www.w3.org/2005/Incubator/mmsm/>
- [9] Tsinarakis C., Polydoros P., Christodoulakis S. "Interoperability Support for Ontology-Based Video Retrieval Applications", CIVR 2004