Semantic Multimedia Content Retrieval and Filtering

Chrisa Tsinaraki
Technical University of Crete, Greece

Stavros Christodoulakis
Technical University of Crete, Greece

INTRODUCTION

Several consumer electronic devices that allow capturing digital multimedia content (like mp3 recorders, digital cameras, DVD camcorders, smart phones etc.) are available today. These devices have allowed both the amateur and the professional users to produce large volumes of digital multimedia material, which, together with the traditional media objects digitized recently (using scanners, audio and video digitization devices) form a huge distributed multimedia information source. The multimedia material that is available today is usually organized in independent multimedia information sources, developed on top of different software platforms.

The Internet, the emergence of advanced network infrastructures that allow for the fast, efficient and reliable transmission of multimedia content and the development of digital multimedia content services on top of them form an open multimedia consumption environment. In this environment, the users access the multimedia material either through computers or through cheap consumer electronic devices that allow the consumption and management of multimedia content. The users of such an open environment need to be able to access the services offered by the different vendors in a transparent way and to be able to compose the different atomic services (like, for example, multimedia content filtering) into new, composite ones. In order to fulfill this requirement, *interoperability* between the multimedia content services offered is necessary.

Interoperability is achieved, at the syntactic level, through the adoption of standards. At the semantic level, interoperability is achieved through the integration of domain knowledge expressed in the form of domain *ontologies*. An ontology is a logical theory accounting for the intended meaning of a formal vocabulary, i.e. its ontological commitment to a particular conceptualization of the world (Guarino, 1998).

The standard that dominates in multimedia content description is the *MPEG-7* (Salembier, 2001), formally known as *Multimedia Content Description Interface*. It supports multimedia content description from several points of view, including media information, creation information, structure, usage information, textual annotations, media semantics, and low-level visual and audio features. Since the MPEG-7 allows the structured description of the multimedia content semantics, rich and accurate semantic descriptions can be created and powerful *semantic retrieval and filtering* services can be built on top of them.

It has been shown, in our previous research (Tsinaraki, Fatourou and Christodoulakis, 2003), that domain ontologies capturing domain knowledge can be expressed using pure MPEG-7 constructs. This way, domain knowledge can be integrated in the MPEG-7 semantic descriptions. The domain knowledge is subsequently utilized for supporting semantic personalization, retrieval and filtering and has been shown to enhance the retrieval precision (Tsinaraki, Polydoros and Christodoulakis, 2007).

Although multimedia content description is now standardized through the adoption of the MPEG-7 and semantic multimedia content annotation is possible, multimedia content retrieval and filtering (especially *semantic* multimedia content retrieval and filtering), which form the basis of the multimedia content services, are far from being successfully standardized.

We focus in this chapter on MPEG-7 based semantic multimedia retrieval and filtering and we introduce the *MPEG-7 Query Language (MP7QL)* and its compatible user preference model, which aim to provide standardized support to such services. The rest of the chapter is structured as follows: In the *Background*

section we present the state of the art in MPEG-7 based multimedia content retrieval and filtering; In the *Main Focus* section we introduce the MP7QL query language and its compliant user preference model; In the *Future Trends* section we outline the future research directions in semantic multimedia content retrieval and filtering and we conclude in the *Conclusions* section.

BACKGROUND

In this section we present the state of the art in MPEG-7 based multimedia retrieval and filtering. It has been mentioned in the introduction that MPEG-7 allows the creation of rich multimedia content descriptions, based on the different aspects of the content. Powerful retrieval and filtering capabilities can be built on top of these descriptions.

Several research groups have been working on MPEG-7 based multimedia content retrieval and filtering, exploiting different features of the MPEG-7 descriptions. The systems offering MPEG-7 based multimedia content retrieval and filtering are classified in three categories:

1. Systems that exploit the textual annotations together with the media-related elements of the MPEG-7 descriptors for retrieval and filtering support (Graves and Lalmas, 2002; Rogers, Hunter and Kosovic, 2003; Tseng, Lin and Smith, 2004). Multimedia content filtering in these systems utilizes the MPEG-7 *Filtering and Search Preferences (FASP)*, which allow the users to specify their preferences regarding multimedia content retrieval and filtering.
2. Systems that exploit the *MPEG-7 Visual* (ISO/IEC, 2001a) and *Audio* (ISO/IEC, 2001a) *Descriptors* in order to support multimedia content retrieval based on the low-level MPEG-7 features (Eidenberger and Breiteneder, 2003; Bertini, del Bimbo and Nunziati 2006). These systems cannot be transparently integrated with the MPEG-7 FASP for multimedia content filtering, because the MPEG-7 FASP do not allow the expression of the user preferences regarding the MPEG-7 low-level visual and audio features.
3. Systems that exploit the semantic metadata descriptions formed according to the Semantic DS of the MPEG-7 *Multimedia Description Schemes (MDS)* (ISO/IEC, 2003a) for semantic multimedia retrieval and filtering (Hammiche, Lopez, Benbernou, Hacid and Vakali, 2006; Tsinaraki, Fatourou and Christodoulakis, 2003; Tsinaraki, Polydoros and Christodoulakis, 2007). These systems cannot be fully exploited using the MPEG-7 FASP for multimedia content filtering, because the MPEG-7 FASP allow only the keyword-based expression of the user preferences regarding the multimedia content semantics.

The major limitation of the above research efforts is that each of them exploits some of the features of the MPEG-7 multimedia content descriptions, but none of them provides a uniform and transparent MPEG-7 retrieval and filtering framework. The most important efforts in the direction of MPEG-7 based retrieval and filtering that transparently exploits all the features of the MPEG-7 descriptions are the following:

- The use of plain XQuery (Chamberlin et al., 2005) on top of an XML repository for MPEG-7 based multimedia content retrieval (Lee et al., 2003). This approach does not take into account the peculiarities of the MPEG-7 description elements. Thus, the different MPEG-7 metadata description elements cannot be fully exploited. This happens because both the MPEG-7 semantic model and the domain knowledge integrated in the semantic MPEG-7 descriptions are expressed in an involved way and cannot be successfully exploited if they are accessed in the same way with the textual and the media-related elements of the MPEG-7 metadata descriptions. The low-level visual and audio features also need special treatment. It is difficult for the average user to express, using plain XQuery, query conditions on the semantics and/or the low-level features and even more difficult to combine such query conditions with textual and media-related query conditions. Finally, XQuery does not support queries with preference values, which allow the users to state which query conditions are more important for them.
- The use of the existing MPEG-7 FASP in order to allow multimedia content filtering and retrieval. The limitations of this approach are the following: