

# An Ontology-Driven Framework for the Management of Semantic Metadata Describing Audiovisual Information

Chrisa Tsinaraki, Eleni Fatourou, and Stavros Christodoulakis

MUSIC/TUC, Technical University of Crete Campus, 73100 Kounoupidiana, Crete, Greece  
{chrisa, eleni, stavros}@ced.tuc.gr

**Abstract.** In this paper, we describe a framework for the management of semantic metadata that describe audiovisual content. The main objectives of this work are to provide a framework allowing enhanced retrieval performance and better user interfaces for audiovisual data, based on extensions of the well accepted international standards for the digital media, digital broadcasting and TV-Anytime domains like TV-Anytime and MPEG-7, as well as to provide an interoperability framework between TV-Anytime and MPEG-7 for the rapidly expanding set of interoperating digital library and interactive TV applications. The framework was developed in the context of the UP-TV project, which investigates TV-Anytime architectures and services. Our approach is ontology-driven, in the sense that we provide ontology based domain-specific extensions of the standards for describing multimedia content. Several aspects of the framework are based on the existence of ontologies. Our framework allows the creation of more efficient user interfaces for accessing content, better retrieval performance and better support in the indexing phase. It also provides compatibility in indexing between the two standards and interoperability for TV-Anytime and digital library applications supported by TV-Anytime servers. We have developed a domain-specific ontology (football games) to test our framework ideas and implementation. The developed ontology uses as primitive modelling elements those provided by the MPEG-7 metadata model. The mapping of the ontology to TV-Anytime metadata primitives is straightforward. We show in detail the concrete implementation of the framework on top of relational databases and its interfaces with the other system components.

## 1 Introduction

We are witnessing the last few years increasing cooperation (and sometimes merging) of computer, home electronics, software, broadcasting and audiovisual information provision industries. Some of the more popular scenarios offered to users include digital audiovisual libraries [8], [11], [6], video on demand services [3] [16] [17] [19] and recently TV-Anytime services [15]. The simpler TV-Anytime scenario foresees that future TVs will be equipped with a processor and large, inexpensive disk storage devices that will be capable of selecting on a continuous basis broadcasted digital TV programs for the user to see on his spare time, according to his personal interests (profile). More recently, the TV-Anytime consortium emphasizes also an open home TV platform with Internet connection capabilities, as well as high-speed access to

TV-Anytime Servers, which have expanded program-capturing capabilities. The server may also offer additional digital library services, thus integrating TV-Anytime, digital library and video on demand services.

As we move from the simpler traditional multimedia services to the more advanced ones, content-based retrieval, either directly from the user or in a way transparent to him, becomes a dominant factor for the success of such services, as it is needed in the different levels of interaction of the service provision system with the end-users. For example, a sports reporter might search among football match videos in a video library in order to retrieve videos of football matches where a certain player, e.g. Ronaldo, scores. On the other hand, an end-user might denote in his preference profile in a personalized TV environment that he would like all the goals of a football match to be recorded for him. While viewing them he may request to see some extracts of the goals of Ronaldo against the same team two years ago from the digital library of the server, as well as articles, comments, critics from reporters, pictures or other audiovisual material about him. In those cases, appropriate retrieval capabilities on the content of the audiovisual information are needed.

The requirements of the above use cases can be satisfied in a system, which stores semantic descriptions for the audiovisual information and provides semantic-based search capabilities on the information content. The semantic descriptions are in fact semantic metadata for the content of the audiovisual information – that is, data describing the content of the audiovisual information. The descriptors may refer to the whole multimedia content or to parts of the content (segments). To guarantee interoperability in complex application environments as the ones above, it is of outmost importance to adhere to international standards.

Widely accepted standards for audiovisual applications and content descriptions include *MPEG-7* [18] and *TV-Anytime* [22]. MPEG-7 is a standard used for the description of multimedia and video metadata that defines a set of *Description Schemes (DSs)*, essentially complex data types, which will be used to describe audiovisual content. MPEG-7 provides a set of description schemes rooted on the *SemanticBase DS*, which contain all the constructs needed for semantic content description. The TV-Anytime forum on the other hand, works on the specification of advanced TV broadcasting servers and TV clients that support personalized ubiquitous services. The TV-Anytime forum has been working on a metadata series, where the most visible parts of metadata are the attractors/descriptors or hyperlinks used in electronic program guides, or in Web pages. This information will be used from the end-users in order to decide whether or not to acquire a particular piece of content. In complex application environments as the above, the content of specific multimedia segments may have been described in either standard.

Before the establishment of standards, research on the semantic description of audiovisual data has been carried out in several models [1] [2] [9] [14] [6] some of which have been implemented in either prototype or real-world systems. During standard definition some research groups used the draft standards and fed their experience back to the standard definition process [20] [21] [12]. Some other systems have been based on specific standards, but there exists no interoperability support [13] [15]. It is very important that we base our research and development efforts on top of the existing standardization efforts (MPEG-7 and TV-Anytime) and that we seek to

expand the retrieval functionality of these frameworks for application specific areas providing at the same time for the compatibility of the enhanced functionality.

Our approach to providing enhanced content-based retrieval in advanced audiovisual applications as the above is ontology<sup>1</sup>-driven. The modelling of the metadata for the description of audiovisual information is based on domain-specific ontologies and several aspects of the framework are based on these ontologies. The use of domain-specific ontologies significantly enhances retrieval performance in an open environment, since it guides the users on what is the language they can use to specify their content preferences, and it also guides the indexers and/or the automatic indexing mechanisms to provide legal content descriptions for specific content. The specification of the legal content descriptions in specific domains not only enhances the retrieval performance in the recall/precision sense, but also allows for better interactive interfaces between naive users and the system. It also enables the use of a better user interface metaphor for interactive TV which is based on selection from a list of allowed constructs and values, instead of inserting text with the help of the remote control. The systematic definition of domain-specific ontologies and their integration to MPEG-7 and TV-Anytime frameworks provides a standardized way of enhancing the functionality of those frameworks.

Newer Interactive TV trends as outlined also in the TV-Anytime forum foresee that the TV sets will be equipped not only with a large disk, but they will also have Internet connections which will allow enhanced TV-Anytime functionality through the connection of home TV-Anytime servers with last mile connections and also allow functionality for users on the move. This expanded TV-Anytime functionality imposes interoperability requirements for TV-Anytime with other audiovisual and digital library standards, like MPEG-7. For example, last mile servers may provide additional archival audiovisual information in MPEG-7, which has been produced by other information providers (Web etc.). In our framework, compatibility and interoperability between the standards is maintained since the same ontology is mapped to be used for the extension of both standards. The retrieval or filtering mechanisms can easily map the preferences of the user (which are also described using the ontology) to both the MPEG-7 and the TV-Anytime metadata representations and achieve enhanced retrieval.

In this paper we describe the framework we have developed for the management of semantic metadata used for the content-based description of audiovisual information. The design of the framework is based on existing, widely accepted standards in the Digital TV and TV broadcasting domains, namely MPEG-7 and TV-Anytime. The framework provides “dual” support, in the sense that it provides pure TV-Anytime functionality as well as pure MPEG-7 functionality, on top of two different repositories, populated consistently using the same annotation tool. This work has been done in the context of the UP-TV project<sup>2</sup>. The proposed approach can be used in any environment where the need for a detailed description of audiovisual material exists.

---

<sup>1</sup> An *ontology* is a set of entities and relationships among them that capture the knowledge on a specific application domain [10].

<sup>2</sup> UP-TV (Ubiquitous & Personalized TV Services) is a European IST project, in which MUSIC takes part and is responsible for its metadata management system.

The rest of this paper is organized as follows: In section 2 we provide an outline of our framework, while in section 3 we refer to our model for the representation of content-description metadata and the ontology we have developed for the annotation of football match videos. The semantic base, where semantic metadata are stored, is described in section 4, while the metadata annotation component we have developed and the API for the support of advanced semantic-based queries are described in sections 5 and 6 respectively. In section 7 we present the conclusions of our work and a brief description of our future research in this area.

## 2 Framework Description

In this section we provide an outline of the framework we developed for the management of semantic metadata for audiovisual information. The architecture of our framework is depicted in Fig. 1.

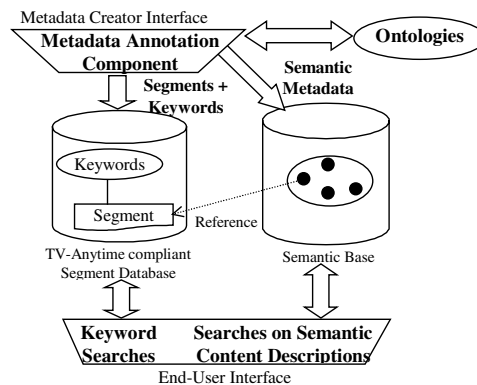


Fig. 1. Framework Architecture

The framework includes two different repositories: A *TV-Anytime compliant Segment Database* containing segmentation metadata represented as TV-Anytime segments, and an *MPEG-7 compliant Semantic Base* where semantic metadata are stored. Capabilities for the population of the Semantic Base are provided, as well as advanced information retrieval capabilities based on the semantic metadata. The framework development includes:

1. The establishment of a *two-layered model for the annotation of audiovisual content*: A set of core metadata entities defined in the Semantic part of MPEG-7 that can be used for the annotation of any video program comprise the first layer of our model, while for specific categories of video content appropriate ontologies are defined. The ontology entities and the relationships among them comprise the second layer of our model [20], [21].
2. The definition of *ontologies for specific subject domains* that implement the second layer of our model. The ontologies are represented using MPEG-7 syntax and constructs. We have developed and integrated an ontology for football matches in

order to test our framework ideas. We will expand the capabilities of our system with other domain-specific ontologies in the future (e.g. news).

3. The implementation of a *Semantic Base* (on top of a relational database) for the metadata information management. The database schema encapsulates our core model as well as the supported ontologies. In fact, the database schema is MPEG-7 compliant.
4. The implementation of a *Metadata Annotation Component* for the population of the semantic base. The component has been integrated with the segmentation tool developed by MUSIC/TUC in the context of the UP-TV project. For each video program, as well as for the video segments defined in a video program, the annotation component is used for the definition of the corresponding metadata items. Then, the metadata items are stored in the semantic base. The semantic base also stores references from the metadata items to the corresponding programs and/or segments. In addition, sets of keywords that are automatically inferred from the metadata items are produced. These keywords are used for the annotation of the video programs and the video segments defined as their parts and are stored in a TV-Anytime compliant database together with the segmentation metadata.
5. The implementation of an *API (Application Programmer's Interface)* that permits, in addition to the simple keyword search capabilities provided by the TV-Anytime compliant database, *advanced semantic-based queries*. The API methods make use of the semantic metadata stored in the semantic base, thus the queries posed can be more accurate as far as it concerns the desired properties of the query results.

From the above, it is obvious that the end-user may search among the programs and segments in our system either via simple keyword (or keyword phrase) searches or using more complex, accurate and expressive semantic searches. As the later is difficult to be supported in a TV environment, the searches will be posed using certain search templates that will be easily filled with a TV interaction device (e.g. remote control). We must note here that the semantic search interfaces are based on the ontologies present in the system.

All the framework components have been implemented in the Linux OS environment, using the MySQL database for the Semantic Base and the TV-Anytime compliant database. The Semantic Annotation Component has been implemented in java.swing. The ontology for the description of football matches is expressed using MPEG-7 syntax and constructs. The advanced search functions made available through the search API are implemented using JDBC on top of a relational database.

### 3 Semantic Metadata Model

In this section, we describe the model for the representation of content-description metadata for audiovisual information. As we have already mentioned it is compatible with the MPEG-7 and TV-Anytime standards. Both standards provide syntax and constructs for the definition of metadata items for the description of audiovisual content. We will provide an overview of the metadata models for both standards next.

In MPEG-7 a set of description schemes provides the complex data types needed for the semantic description of audiovisual content. The more important among the description schemes are:

- *SemanticBase DS*: The abstract type *SemanticBaseType* is defined here. *SemanticBaseType* is the base type extended by other description schemes according to the needs for the description of semantic entities of specific types in a narrative world.
- *SemanticBag DS* and *Semantic DS*: Description schemes used for the description of collections of semantic entities. *SemanticBagType* is an abstract type, defined in the *SemanticBag DS*, extending *SemanticBaseType*, while *SemanticType* is defined in *Semantic DS*. *SemanticType* is concrete type, thus its instances are used for the representation of semantic entity collections.
- *AgentObject DS*: The actors that appear in an audiovisual segment are described here using the extensions of the abstract type *AgentObjectType* that extends *SemanticBaseType*. Actors in general are represented using the *AgentType*, an abstract type extending *AgentObjectType* defined in the *Agent DS*. *PersonType*, *OrganizationType* and *PersonGroupType* extend *AgentType*, are defined correspondingly in the *Person DS*, the *Organization DS* and the *PersonGroup DS* and are used for the representation of persons (e.g. a football player), organizations (e.g. a football team) and groups of persons.
- *Object DS*: The *ObjectType* defined here extends *SemanticBaseType* and is used for the description of objects and object abstractions in the material world (e.g. the ball in a football match).
- *Event DS*: The *EventType* defined here extends *SemanticBaseType* and is used for the description of events that take place in a semantic world (e.g. a goal).
- *ConceptDS*: The *ConceptType* defined here extends *SemanticBaseType* and is used for the description of concepts present in an audiovisual segment (e.g. co-operation).
- *SemanticState DS*: The *SemanticStateType* defined here extends *SemanticBaseType* and is used for the description of a state of the world described in an audiovisual segment and the parametric description of its features (e.g. the score in a football match before and after a goal).
- *SemanticPlace DS*: The *SemanticPlaceType* defined here extends *SemanticBaseType* and is used for the description of a place in a semantic world (e.g. Athens).
- *SemanticTime DS*: The *SemanticTimeType* defined here extends *SemanticBaseType* and is used for the description of semantic time (e.g. Easter).

An important attribute, useful for the support of ontologies, is defined in *SemanticBaseType* and is inherited in all its extensions: *AbstractionLevel*. *AbstractionLevel* is of type *AbstractionLevelType*. It has only one attribute, *Dimension*, of non-negative integer type. When *AbstractionLevel* is not present in a semantic description, the description refers to specific audiovisual material. When *AbstractionLevel* is present, abstraction exists in the description. When the *Dimension* of *AbstractionLevel* is 0, there exists a description of a semantic entity (e.g. the football player Ronaldo) that is related to every audiovisual segment where the entity appears. When *AbstractionLevel* has a non-zero *Dimension*, it specifies classes for the description of abstract semantic entities. The bigger the *Dimension* of *AbstractionLevel*, the more extended the abstraction in the description (e.g. Ronaldo is an instance of the “football player” class with *AbstractionLevel.Dimension*=1,

while “football player” is an instance of the “person” class with `AbstractionLevel.Dimension=2`).

The TV-Anytime standard provides, through the TV-Anytime metadata series, the TV-Anytime metadata system for enabling the consumer to find, navigate and manage content from a variety of internal and external sources including, for example, enhanced broadcast, interactive TV, Internet and local storage. It defines a standard way to describe consumer profiles including search preferences to facilitate automatic filtering and acquisition of content by agents on behalf of the consumer.

As far as it concerns semantic metadata describing audiovisual programs and segments, they are related to the types `BasicContentDescriptionType` and `BasicSegmentDescriptionType` respectively. These types include the following common description elements:

- *Title*: A title given to the segment or program. There can be multiple titles for a segment or program.
- *Synopsis*: A textual description of the segment or program content.
- *Keywords*: Important words or phrases, describing the content of the segment or program.
- *RelatedMaterial*: References to external material, related to the content of the segment or program.

It is obvious from the above that the main mechanism in the TV-Anytime metadata system that could be used for semantic content description is the use of keywords or keyword phrases. Taking advantage of this mechanism, it would be possible to support a keyword-based information retrieval mechanism, similar to a Web search engine. The experience from the search engines shows that such a mechanism has limited performance (precision/recall).

In the digital TV environment we have additional problems due to the interface limitations (lack of keyboard, limited resolution). Instead of an interface that allows the user to select letters to form keywords, an interface that provides a set of application-specific search templates has been implemented.

In our framework we define domain-specific ontologies compatible with the standards. We integrate these ontologies in our databases. The indexers (people responsible for providing metadata for indexing the segments of the audiovisual content) use terms and relationships from the controlled vocabulary with the precise semantics that the ontology contains to do the indexing. In this work they may use the PC interfaces. The users on the other hand, utilize TV-based interfaces. However these interfaces guide the users to select from an existing set of items and relationships through templates and selection mechanisms that guarantee faster interaction and better retrieval performance.

We have dual support, for both MPEG-7 and TV-Anytime in our framework:

- We maintain a Semantic Base where semantic metadata for the description of the content of audiovisual information are stored. We have chosen to define the core entities as restrictions of the corresponding MPEG-7 entities, as the MPEG-7 definitions cannot be fully exploited in a TV-Anytime like environment, while they make the system much more complicated. Thus, the Semantic Base is MPEG-7 compliant. The Semantic Base is described in detail in section 4.
- A TV-Anytime compliant database is used for the storage of segmentation information annotated with keywords that describe the segment contents. The

keywords in the TV-Anytime database have been inferred from the semantic metadata stored in the Semantic Base. The inference is based on the knowledge encapsulated in domain-specific ontologies and a set of rules stored in the Semantic Base. Each rule is a template, which produces a keyword phrase for each relationship (e.g. for a goal scored by Ronaldo is produced the keyword phrase “Ronaldo scores”). In addition, independent keywords corresponding to the metadata items related to the video segment are stored in the TV-Anytime database (e.g. “Ronaldo” and “goal” for a goal scored by Ronaldo).

This approach is compliant with the TV-Anytime forum policies, as the forum moves to the support of servers that provide added value services together with the standard TV-Anytime functionality.

### 3.1 Ontology for the Description of Football Matches

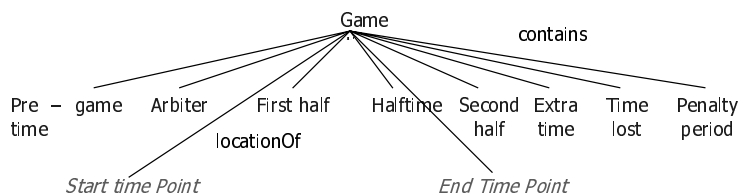
In order to evaluate our framework, we have defined an ontology for the description of football matches. The entities of the ontology and the relationships among them comprise the second layer of our video metadata model. The definition of the ontology is based on the rules of FIFA (the international organization responsible for the definition of football rules) [7].

We defined our ontology using MPEG-7 syntax and constructs. Thus, we use the mechanism provided by MPEG-7 for the definition of abstract semantic entities, the use of AbstractionLevel. An MPEG-7 compliant XML document contains the entities of interest that have been defined.

In the following paragraphs, the representation of time periods, places, actor roles and events in a football match are described.

**Time.** Abstract time entities, referring to the time periods in a football match are defined in our ontology. The part-of hierarchy of the ontology terms for the time periods of a football match is shown in Fig. 3.

As shown in Fig. 3, there exist the *Pre-game Time*, the *Arbiter*, the *First half*, the *Halftime*, the *Second half*, the *Extra time*, the *Time lost* and the *Penalty period*. The *Start Time Point* is the beginning of the First half, while the *End Time Point* is the end of the Second half.

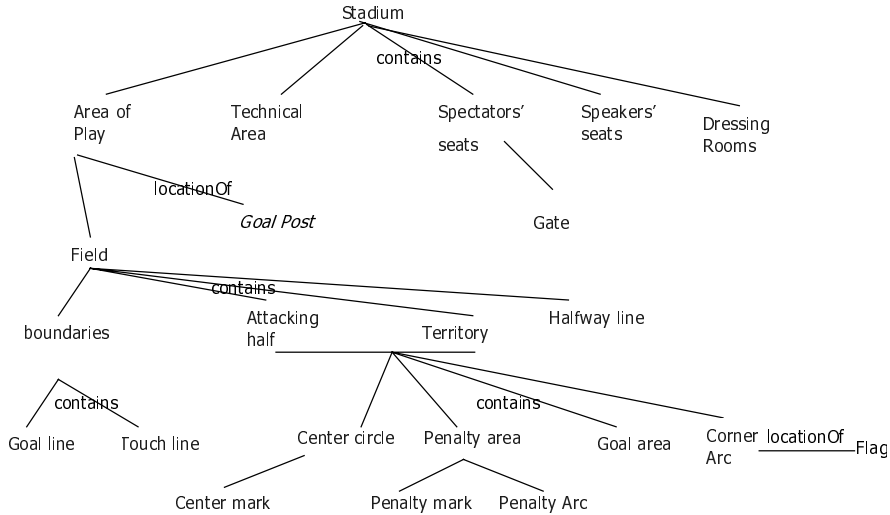


**Fig. 2.** The ontology terms for time periods in a football match

**Places.** Football matches usually take place in a stadium. In order to make possible the support of more accurate queries (e.g. events near the halfway line), we have defined a part-of hierarchy of the ontology terms for the stadium regions, as shown in Fig. 2. According to Fig. 2, the ontology terms for the stadium regions are:

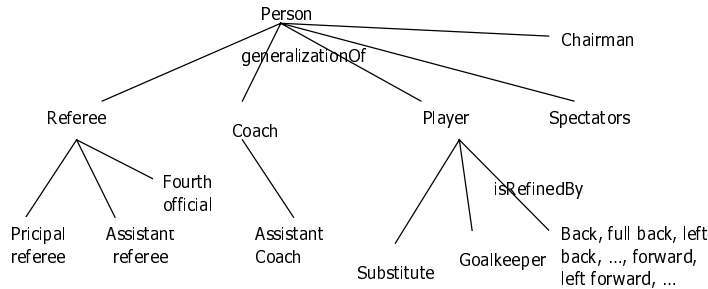


1. The *Area of Play*. It contains the *Field* and the *Goal Posts*. The field's *Boundaries* are the *Goal Line* and the *Touch Line*. There exist the team *Territories* and their *Attacking Halves* (the territory of a team is the opponent's attacking half). In the field exist the *Center Circle* (that contains the *Center Mark*), the *Penalty Areas* (attached to the *Penalty Arcs*) and the *Corner Arcs* (where the *Corner Flags* are located).
2. The *Technical Area*.
3. The *Spectator's Seats*, including the fans' *Gates*.
4. The *Speaker's Seats*.
5. The *Dressing Rooms*.



**Fig. 3.** The part-of hierarchy of the ontology terms for regions of the stadium

**Actor Roles.** The roles of the football match actors (e.g. teams, players etc.) are abstract semantic entities contained in our ontology. The ontology entities for the representation of person roles in a football match are shown in Fig. 4, while the entities and relationships for the description of teams are depicted in Fig. 5.



**Fig. 4.** The ontology terms for person roles in a football match

According to Fig. 4, the roles in a football match are *Referee*, *Coach*, *Player*, *Spectator* and *Chairman*. A referee instance may be the *Principal referee*, an *Assistant referee* or a *Fourth official*. A Player instance may be a *Goalkeeper*, a *Substitute* or may be in a specific place (e.g. *Left Forward*), which refines his role. The roles are represented by abstract MPEG-7 Person entities (AbstractionLevel=2).

As shown in Fig. 5, a team, resented as an MPEG-7 Organization instance, may be *Visiting Team* or *Home Team*. Team players comprise PersonGroups.

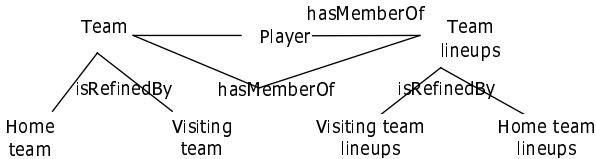


Fig. 5. The ontology terms for the teams participating in a football match

**Events.** The events that take place in a football match are specified in the FIFA rules [7]. An event may fall in one of 3 categories: *Referee Actions*, *Game Actions* and *Miscellaneous*. We implement them in our model as MPEG-7 Event instances (AbstractionLevel=2). As the events are too many (123) to fit in a diagram, their names and the category that each one falls are listed in Table 1.

Table 1. The ontology terms for the events taking place in a football match

<b>Referee Action</b>	Principal Referee Action, Whistle, Red Card, Book a player, Yellow Card, Advantage Rule, Warning, Assistant Referee Action, Raise flag, Fourth Official Action
<b>Game action</b>	Interact with a player, Hit the ball, Pass, Shoot, Trap, Player action, Ball action, Clear, Illegal Action, Foul, Off-side, Dangerous play, Half volley, Heel, Toed Ball, Juggling, Deadening the ball, Trap, chest, Trap, foot, Kick, Trap, thigh, Diving, First time, Flick, Bicycle kick, Scissor kick, Backswing, Chip shot, Banana kick, In swinger, Out Swinger, Missile, Pass, back, Pass, front, Power kick, Outlet, Assist, Cross, Through Ball, Pass, drop, Early Service, Pass, forward, Pass, hospital, 50/50 ball, Pass, lead, Nutmeg, Pass, push, Send, Hacking, Hand-to-ball, Obstruction, Carrying the ball, Header, Instep drive, Volley, Pass, square, Pass, hopped, Goal, Golden goal, Sudden death, Clinical goal, Substitution, Switch, Fake, Feint, Close down, Re-start, kick-off, Penalty kick, Free kick, Dropped Ball, Indirect free kick, Direct free kick, Corner kick, Goal kick, Throw-in, Retake, Injury, Jump, Fall, Ball to Hand, Reflection
<b>Miscellaneous</b>	Reaction of the spectators, Applause, Boo, Encourage, Sing, Wave flags, Throw items, Give instructions, Medical treatment, Talk, Fight, Hit, Shake hands, Cheer

4 Semantic Base

In this section we describe the Semantic Base maintained in our framework for the storage of the metadata describing audiovisual information. The Semantic Base is implemented on top of a MySQL relational database. The schema of the underlying database depicts our core model and is MPEG-7 compliant. Thus, the entities represented in the Semantic Base are the MPEG-7 types described in section 3. The database tables that represent these entities are shown in Fig. 6.

In addition, in the Semantic Base information related to the audiovisual segments but not contained in them (e.g. the age of a player) is stored. This information is

stored in the tables Person, Organization, PersonGroup, State, Place and Time. These tables are shown in brief in Fig. 7.

Due to its semantic content, the Semantic Base can support much more powerful queries than those supported by a keyword only based annotation system.

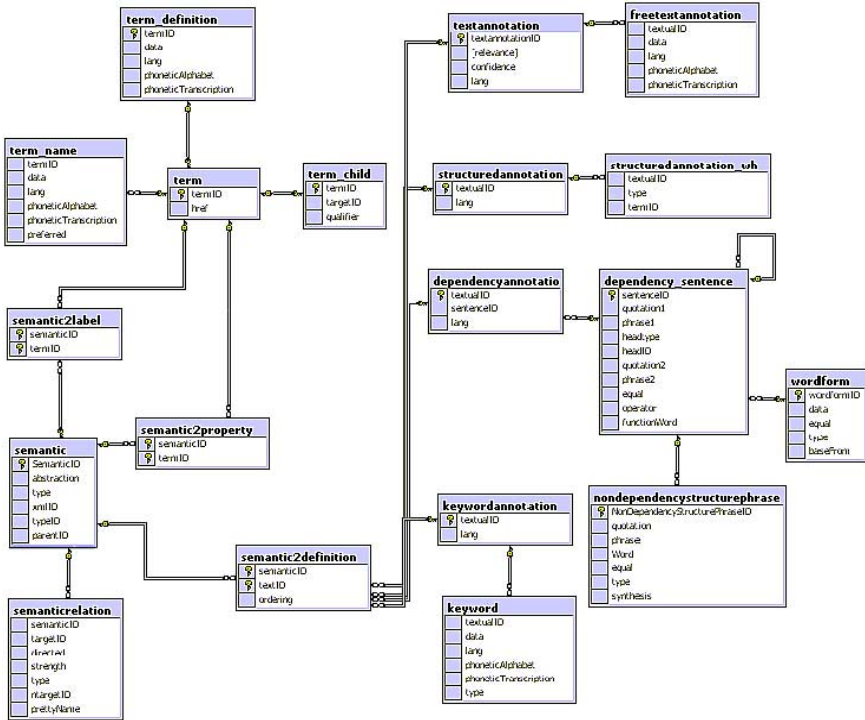


Fig. 6. The MPEG-7 compliant part of the Semantic Base schema

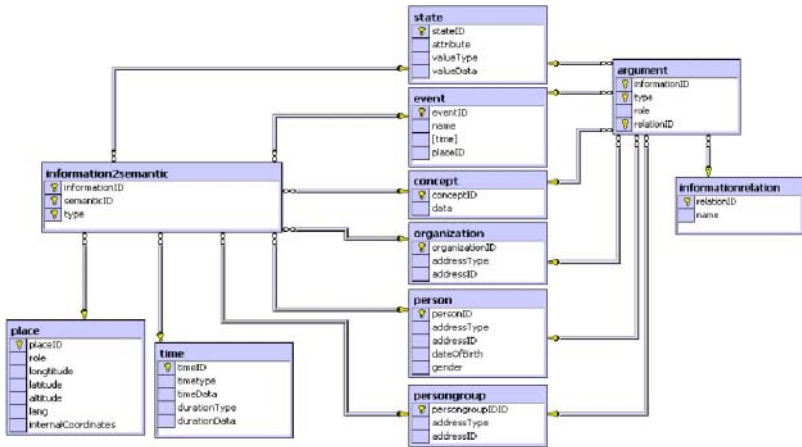


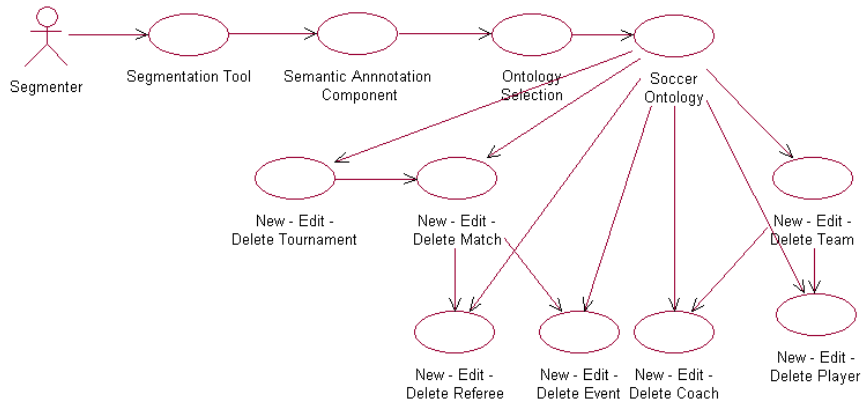
Fig. 7. Representation of the additional information

## 5 Semantic Annotation Component

The semantic metadata stored in the Semantic Base are created during the segmentation process and are related to certain audiovisual programs and segments. Thus, the segmentation process includes a *metadata creation process*. For this reason, we have implemented in java.swing the *Semantic Annotation Component*, which is used, as a component of a segmentation tool, by the metadata creators, who are responsible for the appropriate (manual or semiautomatic) segmentation and metadata description of the programs.

When the metadata creation finishes, a set of segments, annotated with appropriate keywords or keyword phrases is inserted in the TV-Anytime compliant segment base and a set of metadata items describing them is inserted in the semantic base. The metadata items in the Semantic Base reference the corresponding programs and/or segments. The keywords and the keyword phrases for a segment are produced in an automated way from the metadata that correspond to the segment in order to achieve TV-Anytime compatibility. The metadata creation process is guided, for each kind of audiovisual program, by the appropriate ontology.

The use case diagram of Fig. 8 depicts the usage scenario of the semantic annotation component for the annotation of football match audiovisual programs and segments. As is shown there, the person responsible for the segmentation process (Segmenter) uses a segmentation tool in order to define the video program segments. Segment definition includes segment annotation, thus the semantic annotation component is used as part of the segmentation tool<sup>3</sup>.



**Fig. 8.** Use Case diagram of the Semantic Annotation Component

The first task of the segmenter is the selection of the appropriate ontology for the segment annotation (the “soccer ontology” is selected for football matches). Then, the tournament to which the current match belongs is selected. If the tournament doesn’t

<sup>3</sup> The metadata items defined for each audiovisual segment have their own Semantic Time that allows the expression of the game time of the event. This permits the support of multiple cameras, as video time may be different from the game time (e.g. 40<sup>th</sup> video minute may correspond to the 36<sup>th</sup> game minute, as a goal has been shown 4 times of duration 1 minute).

exist, a new tournament is defined and selected. After tournament selection, the football match is selected. Again, if the match doesn't exist it is first defined. The match is associated with the participating teams and each team is associated with its coach. All the non-existent entities are defined. Then, for each notable event in the segment an appropriate event instance is created. The event is associated with the appropriate metadata items (e.g. a goal is associated with the scorer and the goalkeeper) after instantiating the non-existent ones. In Fig. 9 a goal event is defined.

The image shows two overlapping windows from a software application. The background window is titled 'Keywords' and contains a tree-like structure for defining a goal event. It has tabs for 'Segment Title', '1', '2', and '3'. The 'Segment Title' tab is active, showing a hierarchy: 'Category' (Soccer games), 'Season' (2002), 'Tournament' (2002 FIFA World Cup Japan/Korea), 'Match' (Germany - Ireland Republic), and 'Event' (Goal). Below this, there are buttons for 'OK' and 'CANCEL'. The foreground window is titled 'Goal' and contains fields for defining a goal event. It has a 'Scored by' field with a dropdown menu set to 'Player' and a text field containing 'Linke Thomas'. Below this is a 'Lost by' field with a dropdown menu set to 'Goalkeeper' and a text field containing 'Kahn Oliver'. There are two 'At' fields: one with a dropdown set to 'Field' and another with a dropdown set to 'Game time'. Both have associated text input fields and units ('meters away' and 'minute' respectively). At the bottom are buttons for 'OK', 'CANCEL', and 'RESET'.

Fig. 9. Goal definition

The semantic annotation component also permits the definition of semantic information for time points contained in a segment, thus enhancing the information retrieval capabilities provided.

The semantic information defined through the usage of the semantic annotation tool is in the form of an XML document. The XML document is then transformed, through the usage of an XSLT document [5], in a set of SQL queries that insert in the semantic base the instances corresponding to the appropriate document elements.

## 6 Search API

As mentioned above, TV-Anytime allows only indexing based on keywords or keyword phrases. However, the pure TV-Anytime approach to specifying the content utilizing keywords and keyword phrases in addition to limiting the power of the retrieval language, has interoperability limitations that become apparent with the new evolutions in technology, standards and business models. The use of domain specific ontologies to limit the possible values of keywords in conjunction with the TV-Anytime metadata alleviates two significant problems:

- New TV-Anytime Forum trends emphasize the existence of servers that are connected with high capacity lines with homes and provide additional digital audiovisual channels, but also digital library functionality with multimedia content. This content may be structured according to the MPEG-7 standard to give high discriminatory power to users.
- It is highly possible that digital audiovisual information stored in the TV-Anytime server will also be used by other consumption channels, for example PCs or mobile

devices (the later is a scenario also supported by the TV-Anytime Forum). It is also possible that digital audiovisual content will be transferred between servers to support users during travelling. In all the above scenarios high precision and recall of the retrieval mechanism is needed and the limitations of the TV interface do not exist. A more powerful than keyword based retrieval mechanism presents advantages.

To overcome these limitations, we have developed a search API containing functions that allow the user to pose powerful MPEG-7 queries if he wishes (simple keyword queries to support also TV-Anytime queries are a special case).

The API does not make any assumptions on the kind of the end-user equipment and interaction capabilities. The application through which the end-user interacts with the rest of the system uses a set of SQL queries that retrieve the IDs of the entities specified in the interface (e.g. retrieve the ID of the Person named “Ronaldo”). Thus, the search API can be used in different environments (e.g. TV, Web), with different user interaction capabilities (e.g. remote control vs. mouse and keyboard): In a TV interface a search template is presented, where only the missing words need to be filled from a list presented in the TV. In this case, the proper items are selected using the remote control. On the other hand, in the Web, a Java application that permits full user interaction will be provided.

The same API and set of interfaces can be used to interact and retrieve audiovisual contents described in both the database which supports the MPEG-7 standard and the database that supports the TV-Anytime standard, taking into account that both databases are based on domain-specific extensions of the standards which use the same ontology.

The API has been implemented in Java, using JDBC on top of a relational database. It is comprised of two functions that are capable of supporting queries that contain one or more criteria combined with the logical operators AND and OR. More specifically, the API functions are:

1. *GetSegmentA((AND|OR)? (informationID informationType semanticID)+)*, where:

- The first argument denotes the operator (AND/OR) used for the combination of the different criteria expressed in the rest of the query. The argument can appear at most once and may be omitted if there is only one criterion.
- The second argument is a set of triplets that describe the query terms, that is the metadata items that must be related to the query results: *informationID* is the id of the item, *informationType* is the type of the item and *semanticID* is the role of the item. There must appear at least one triplet in a query. The “null” value should be used when we don’t want to set a criterion for any part of the triplet.

This function permits queries that retrieve audiovisual segments related to specific metadata items. No relationships among the metadata can be defined, while the query terms may be combined using the AND and OR logical operators.

In Table 2 we present examples of queries expressed using *GetSegmentA*, where we assume that the *informationID* of the person named “Ronaldo” is 1 and the *informationID* of the person named “Rivaldo” is 2.

2. *GetSegmentB(informationID informationType semanticID (AND|OR) (informationID informationType semanticID relationType)\*)*, where:

**Table 2.** Examples of queries expressed using GetSegmentA

Query	Description in Free Text
GetSegmentA(1 Person null)	“Give me the segments where Ronaldo appears” (not only as a player!)
GetSegmentA(1 Person player-obj)	“Give me the segments where the player Ronaldo appears”
GetSegmentA(AND, (1 Person player-obj) (null Event goal-event))	“Give me the segments where the player Ronaldo appears and a goal is scored” (the scorer may be any player)
GetSegmentA(OR, (1 Person player-obj) (2 Person player-obj))	“Give me the segments where the players Ronaldo or Rivaldo appear”

- The first argument is a triplet that describes the main query term that is a metadata item related to the other ones: *informationID* is the id of the item, *informationType* is the type of the item and *semanticID* is the role of the item.
- The second argument denotes the operator (AND/OR) used for the combination of the query terms expressed in the rest of the query. The argument can appear at most once and may be omitted if no query terms follow.
- The third argument is a set of quadruplets that describe the query terms that are related to the main query term: *informationID* is the id of the item, *informationType* is the type of the item, *semanticID* is the role of the item and *relationType* is the type of the relationship of the current item with the main query term. There can appear any number of quadruplets in a query. The “null” value should be used when we don’t want to set a criterion for any part of the quadruplet.

This function permits queries that retrieve audiovisual segments that relate to a specific metadata item that is the main query term. The main query term may be related, through specific relationships, to the other metadata items (including the information related to them that is not present in the described audiovisual segment, a feature of special interest in Digital Library environments).

In Table 3 we present examples of queries expressed using GetSegmentA, where we assume that the *informationID* of the person named “Ronaldo” is 1 and the *informationID* of the person named “Kahn” is 3. We also assume that the Greek citizenship has citizenshipID 1.

**Table 3.** Examples of queries expressed using GetSegmentB

Query	Description in Free Text
GetSegmentB(null null goal-event AND 1 Person player-obj hasCauserOf)	“Give me the segments where the player Ronaldo scores”
GetSegmentB(null null goal-event AND (1 Person player-obj hasCauserOf) (3 Person player-obj hasPatientOf))	“Give me the segments where the player Ronaldo scores against the goalkeeper Kahn”
GetSegmentB((null Person player-obj hasCauserOf) AND (1 citizenship person2citizenship))	“Give me the segments where a player with Greek citizenship appears”

The combination of the above functions (through the union or intersection of the results of more than one queries) can support even more complex queries. Thus, powerful queries in several platforms can be supported. The integration of the API to the end-user applications remains a challenge, especially in the Digital TV environment.

## 7 Conclusions – Future Work

We have described a framework for the management of semantic metadata used for the description of audiovisual information. Metadata modelling is based on domain-specific ontologies and several aspects of the framework are based on these ontologies. Our framework provides interoperability and “dual” support for both MPEG-7 and TV-Anytime in order to capture the needs of the evolving digital multimedia applications and architectures. This support is provided through the maintenance of two information repositories, an MPEG-7 compliant Semantic Base and a TV-Anytime compliant database.

The Semantic Base maintenance permits the support of advanced queries based on the semantic base contents, which allows better content description and retrieval performance. In addition, it allows better user interfaces, especially user interfaces based on list selection, which are more appropriate for interactive TV. This support can be exploited, through the Search API that we have implemented. The API can facilitate retrieval in TV-Anytime, digital video libraries, video on demand and integrated application environments.

Our current work in this area includes:

- Investigation of system performance and automatic adaptation oriented aspects in the ontology-driven framework.
- The deployment of the framework in order to provide automatic (or semi-automatic) abstractions of the audiovisual content, based on the corresponding semantic information stored in the semantic base and the domain specific ontologies.
- Experimentation with the retrieval performance (precision vs. recall) of queries aiming to retrieve the same audiovisual content on top of both the TV-Anytime compliant database and the Semantic Base.
- Investigation of domain specific fuzzy query and user interface aspects.
- Methodology and system development for semi-automatic creation of MPEG-7 metadata descriptions from the TV-Anytime descriptions. Since MPEG-7 descriptions encode more semantics, a single TV-Anytime description may correspond to more than one MPEG-7 descriptions.
- The definition of ontologies for other application domains (e.g. news) and the adaptation of existing ones in order to enhance the capabilities of our system. Complexity and cost for these tasks will be investigated.

**Acknowledgments.** Our thanks to the staff of MUSIC/TUC, especially Fotis Kazasis and Nektarios Moumoutzis, for their comments and suggestions during the design of the framework presented in this paper. This work was financially supported by the UP-TV project.



## References

1. Al-Khatib Q., Day F., Ghafoor A., Berra B., *Semantic Modelling and Knowledge Representation in Multimedia Databases*, IEEE Transactions on Knowledge and Data Engineering, Vol. 11, No. 1, January/February 1999
2. Analyti A., Christodoulakis S., *Multimedia Object Modeling and Content-Based Querying*, Proceedings of Advanced Course – Multimedia databases in Perspective, Netherlands 1995.
3. Christodoulakis S., Pappas N., Kyriakaki G., Maragoudakis Y., Mavraganis Y., Tsinaraki C., *The KYDONIA multimedia information server*, Proceedings of the European Conference on Multimedia Applications Services and Techniques – ECMAS 97, Milan, May 1997
4. Carnegie Mellon University, *Informedia project*, <http://www.informedia.cs.cmu.edu>
5. Clark J., *XSL Transformations (XSLT) Specification Version 1.0*, W3C Recommendation, <http://www.w3.org/TR/xslt>, November 1999
6. de Jong F., Westerveld T., *MUMIS: Multimedia Indexing and Searching*, Proceedings of the CBMI 2001 Workshop, pp. 423–425
7. Federation Internationale de Football (FIFA), *Football Game Rules*, [www.fifa.org](http://www.fifa.org), 2002
8. Gemmell D.J., Han J., Beaton R., Christodoulakis S., *Delay-Sensitive Multimedia on Disks*, IEEE Multimedia 1.3, Fall 1994.
9. Grosky W., *Managing Multimedia Information in Database Systems*, Communications of the ACM, Vol. 40, No. 12, December 1997
10. Guarino N., *Formal Ontology and Information Systems*, pp. 3-15, Proceedings of the first International Conference Formal Ontology in Information Systems (FOIS '98), June 6–8 1998, Trento, Italy, Ed. Guarino N., IOS Press
11. Hacid M-S., Declair C., Kouloumdjian J., *A Database Approach for Modeling and Querying Video Data*, IEEE Transactions on Knowledge and Data Engineering, Vol. 12, No. 5, September/October 2000
12. IBM Japan, *Video Enrichment project*, [http://www.tri.ibm.com/projects/video/index\\_e.htm](http://www.tri.ibm.com/projects/video/index_e.htm)
13. Jaimes A., Echigo T., Teraguchi M., Satoh F., *Learning Personalized Video Highlights from Detailed MPEG-7 Metadata*, in Proceedings of the IEEE International Conference on Image Processing, ICIP 2002, Sept 22–25, Rochester, NY, USA
14. Jiang H., Elmagarmid A. K., *WVTDB – A Semantic Content-Based Database System on the World Wide Web*, IEEE Transactions on Knowledge and Data Engineering, Vol. 10, No. 6, November/December 1998
15. Kazasis F., Moumoutzis N., Pappas N., *Informative Annex: Generic Architecture for handling TVA Metadata using relational database technologies*, contribution TV-Anytime Forum/Metadata WG, 2nd Implementer's workshop, Geneva July 2002
16. Kyriakaki G., *MPEG Information Management Services for Audiovisual Applications*, Master Thesis, Technical University of Crete, March 2000
17. Mavraganis Y., Maragoudakis Y., Pappas N., Kyriakaki G., *The SICMA multimedia server and the virtual museum application*, Proceedings of the European Conference on Multimedia Applications Services and Techniques – ECMAS 98, Berlin, March 1998.
18. MPEG Group, *MPEG-7 (Multimedia Content Description Interface)*, <http://mpeg.telecomitalialab.com/standards/mpeg-7/mpeg-7.htm>
19. Pappas N., Christodoulakis S., *Design & Development of a Stream Service in a Heterogeneous Client Environment*, Proceedings VLDB 2000.
20. Tsinaraki C., Papadomanolakis S., Christodoulakis S., *A Video Metadata Model supporting Personalization & Recommendation in Video-based Services*, Proceedings of the MDDE Workshop 2001 (in conjunction with ReTIS 2001), pp 104-109, Lyon (France), July 2001
21. Tsinaraki C., Papadomanolakis S., Christodoulakis S., *Towards a two-layered Video Metadata Model*, Proceedings of the DEXA Workshop 2001 – DLib, pp 937–941, 3–7 September 2001, Munich
22. TV-Anytime Forum, <http://www.tv-anytime.org/>