A Logic Based Approach for the Multimedia Data Representation and Retrieval

Samira Hammiche, Salima Benbernou
LIRIS – University Claude Bernard Lyon 1
43, bld du 11 Novembre 1918
69622 Villeurbanne - France
{shammich, sbenbern}@liris.univ-lyon1.fr

Athena Vakali
Department of Informatics
Aristotle University of Thessaloniki
54006 Thessaloniki – Greece
avakali@csd.auth.gr

Abstract

Nowadays, the amount of multimedia data is increasing rapidly, and hence, there is an increasing need for efficient methods to manage the multimedia content. This paper proposes a framework for the description and retrieval of multimedia data. The data are represented at both the syntactic (structure, metadata and low level features) and semantic (the meaning of the data) levels. We use the MPEG-7 standard, which provides a set of tools to describe multimedia content from different viewpoints, to represent the syntactic level. However, due to its XML Schema based representation, MPEG-7 is not suitable to represent the semantic aspect of the data in a formal and concise way. Moreover, inferential mechanisms are not provided. To alleviate these limitations, we propose to extend MPEG-7 with a domain ontology, formalized using a logical formalism. Then, the semantic aspect of the data is described using the ontology's vocabulary, as a set of logical expressions. We enhance the ontology by a rules layer, to describe more complex constraints between domain concepts and relations.

User's queries may concern the syntactic and/or semantic features. The syntactic constraints are expressed using XQuery language and evaluated using an XML query engine; whereas the semantic query constraints are expressed using a rules language and evaluated using a specific resolution mechanism.

1. Introduction

Usually, the multimedia content is studied in two parallel dimensions : the syntax (structure and low level features) and the semantic (the meaning of the data). In multimedia data representation and retrieval models, these two dimensions have not only to be taken into account, but also require that each of them be dealt by the the most appropriate tools to it, and these sets of tools have to be integrated in a complimentary way in order to respond user's queries.

The Moving Picture Experts Group (MPEG) published the standard MPEG-7 [1], in February 2002 which aims to describe the multimedia content from different viewpoints. This standard is based on XML Schema [2] and focuses on the structural and low-level features, then it is suitable for the representation of the syntactic dimension of the multimedia content. However, XML Schema displays some limitations when there is a need to extract implicit information. This is due to the fact that there is neither formal semantics nor inference capabilities associated with the elements of an XML document. Consequently, MPEG-7 cannot represent the meaning (semantics) of the multimedia content. Then, another formalism has to be considered for the semantic dimension, where the following issues are required:

- formally define the semantics of MPEG-7 terms and multimedia content;
- express these definitions in a machine understandable language;
- and provide inferential mechanisms to allow reasoning over the multimedia content.

Logical formalisms, which are rather appropriate for the above requirements, are considered in our framework.

To take into account the two dimensions (syntactic and semantic) in the same framework and to alleviate the limitation of MPEG-7 description, we propose to extend MPEG-7 with domain ontology to provide a shared and formal vocabulary for the specification of the semantics. As description logic languages are usual to represent ontologies, we use a specific description logic (DL) language to describe the semantic of the multimedia content. However, if most of the DL languages are very expressive to describe the concepts (i.e. conceptual entities of the domain), they don't pay so much attention to the role expressions (relations between concepts). Moreover, DLs formalisms have rather weak query languages. The simple and existing ones are mostly

[1] http://www.chiariglione.org/mpeg/standards/mpeg-7/mpeg-7.htm

based on concepts satisfiability. This technique cannot be used directly for queries involving roles and variables. To cope with these limitations, we enhance the ontology by a rules layer (represented using Datalog rules) which provide a powerful query langauge. Therefore, a unified and adequate formalism integrating rules and DLs is used to represent the semantic dimension of the multimedia data.

The semantic query processing is done in two steps, first the query is rewritten using the Datalog rules and the DL reasoning mechanism, then an evaluation strategy is applied to the rewritten queries, which is based on first order logic resolution algorithm.

To summarize, we present a formal framework where the standard MPEG-7 and an hybrid language (DL and rules) are used to capture the syntactic and semantic dimensions respectively, of the multimedia data. These two dimensions are related using XPath expressions. For the retrieval, we use XQuery [4] and a resolution algorithm to answer user's queries in a complimentary way.

This paper is organized as follows: in Section 2, we give a case study scenario and some preliminaries concerning the MPEG-7 standard and the hybrid language (DL and Datalog rules languages). Section 3 presents the logical based framework architecture used for the representation and retrieval of the multimedia data. Section 4 introduces an application example. In Section 5, we relate our work to the existing ones and we conclude in Section 6.

2. Preliminaries

2.1 A case study scenario

In our proposal, we consider soccer game videos. Then, the data dealt with are video data and the ontology is a soccer game ontology. The soccer game ontology involves many concepts and relations. An extract of the soccer game ontology and the rules is given in Table 1. Our DL knowledge base is a set of axioms using two forms. The first is $C \sqsubseteq D$, which states that every instance of the concept C must be an instance of the concept D; and the second form $R \sqsubseteq C \times D$ states that the role R has the domain C and the range D.

We introduce a Datalog rules layer on the top of the DL to cope with the DL limitations about roles and query language. Then, it is possible to define n-ary predicates such as the predicate introduced in the rule R_1, which captures an event in a video segment and its actors (for example, the goal event has a specific player as agent and a goalkeeper as patient); and the predicate of the rule R_2, which gives the conditions to have two players in the same team.

Semantic queries are posed in term of the domain ontology vocabulary and the ordinary predicates defined using Datalog rules. For example, if a query is to obtain the video segment where the player "$Ronaldo$'" scores a goal against the goalkeeper "$Khan$". This query can be expressed in the Datalog query language as follows:
$Q(x) \longleftarrow Eventactors(Ronaldo, Khan, goal, x)$.

To answer this query, we should first rewrite it using the ontology vocabulary and the intentional Datalog rules, then evaluate the resulting queries. The rewriting is done in two steps: the first is to expand the Datalog intentional rule.
Q will be rewritten into the query Q' as follows:
$Q'(x) \longleftarrow Person(Ronaldo), Person(Khan), Event(goal), Video(x), AgentOf(Ronaldo, goal), PatientOf(Khan, goal), EventOccur(x, goal)$.

The second step is to reason over the domain ontology to infer concepts that are subsumed by the concepts in the query. Then, the query Q' will be rewritten into a disjunction of queries $Q_1, ..., Q_n$. Each of these queries is semantically equivalent to the initial query Q'.

Table 1. Extract of Soccer game ontology

Ontology Concepts
Actor roles hierarchy
$Person \leq \top$, $Player \sqsubseteq Person \sqcap \exists play.Team$
$AreaPlayer \sqsubseteq Player$, $Goalkeeper \sqsubseteq AreaPlayer$
$Substitute \sqsubseteq Player$, $AreaPlayer \sqcap Substitute \sqsubseteq \perp$
Game event hierarchy
$Event \leq \top$
$RefereeAction \sqsubseteq Event \sqcap \exists r\_action.Referee$
$GameAction \sqsubseteq Event \sqcap \exists g\_action.Player$
$Miscellneous \sqsubseteq Event \sqcap \exists m\_action \neg(Referee \sqcap Player)$
The concept Video
$Video \sqsubseteq (\geq 1 EventOccur.Event) \sqcap (\geq 1 Appear.Person)$
Ontology roles
$AgentOf \sqsubseteq Person \times Event$
$PatientOf \sqsubseteq Person \times Event$
$MemberOf \sqsubseteq Person \times Team$
$ResultOf \sqsubseteq Event \times Event$
$play \sqsubseteq Player \times Match$
$EventOccur \sqsubseteq Video \times Event$
$Appear \sqsubseteq Video \times Person$
Datalog rules
$R_1 : Eventactors(P_1, P_2, E, X) \longleftarrow Person(P_1), Person(P_2), Event(E), Video(X), AgentOf(P_1, E), PatientOf(P_2, E), EventOccur(X, E)$
$R_2 : SameTeam(P_1, P_2, T) \longleftarrow Player(P_1), Player(P_2), Team(T), MemeberOf(P_1, T), MemberOf(P_2, T)$

2.2 The MPEG-7 standard

MPEG-7, formally called "*Multimedia Content Description Interface*", specifies a standard set of description tools, which are applicable for all kinds of multimedia content independent of its format and coding. These tools are [15]:

- A set of *Descriptors* (Ds), to represent the features of audio-visual material, (e.g. color histogram).
- A set of *Description Schemes* (DSs), which define the structure and the semantics of the relationships between elements, which include Ds and DSs. An example is the hierarchical structure of a video represented by the *VideoSegment DS*.
- A *Description Definition Language* (DDL), which provides a syntax to combine, express, extend and refine Ds as well as DSs. It is XML Schema based with some MPEG-7 specific extensions.
- *System Tools*, to support efficient binary encoding multiplexing, synchronization and transmission of the descriptions.

```
<Mpeg7> <Description xsi:type="ContentEntityType">
 <MultimediaContent xsi:type="VideoType">
 <video id="Video">
  <!-- Creation Information -->
  <CreationInformation>
  <Creation> <Title>Spain vs Sweeden (July 199) </Title>
             <Creator> BBC </Creator> </Creation>
  <Classification> <Genre Type="main"> Sport </Genre> </Classification>
  </CreationInformation>
  <!-- Media Information -->
 <MediaLocator><MediaUri>soccer.mpg</MediaUri></MediaLocator>
 <MediaTime><MediaTimePoint>T00:00:00</MediaTimePoint>
        <MediaDuration>PT1M30S</MediaDuration> </MediaTime>
  <!-- FreeTextAnnotation -->
  <!-- Structural Decomposition of the Video -->
 <TemporalDecomposition gap="false" overlap="false">
  <VideoSegment id="ID-V1">
   <!-- MediaTime  & MediaLocation Informaiton -->
   <SpatioTemporalDecomposition gap="false" overlap="false">
    <!-- Three Moving Regions: "Player1-MR, Player2-MR,  Ball-MR" and
     a still  region "Net-SR" after the spatio-temporal decomposition -->
   </SpatioTemporalDecomposition>
   <Semantic><Label>
        <Name>Foul event and players relationships </Name></Label>
      <SemanticBase xsi:type="EventType" id="Foul-EV">
       <Label><Name>Foul</Name></Label>
       <Relation type="cs:SemanticRelationCS:2001:agent"
             source="#Player1-MR" target="#Player2-MR"/>
      </SemanticBase>
      <Graph><Relation type="cs:SpatialRelationCS:2001:left"
              source="#Player1-MR" target="#Player2-MR"/>
         <Relation type="cs:SpatialRelationCS:2001:touch"
              source="#Player2-MR" target="#Ball-MR"/>
    </Graph></Semantic></VideoSegment>
  <VideoSegment id="ID-V2">
     <!-- Decomposition  and Semantic Information -->
  </VideoSegment></TemporalDecomposition>
 </Video></MultimediaContent></Description></Mpeg7>
```

Figure 1. Example of an MPEG-7 description

Figure 1 is an MPEG-7 description of a soccer game video extract. This description includes *Creation Information* (e.g. title, creator, classification), *Media Information* (e.g. media time, media location), *Structural Aspects* (segment decomposition) and *Semantic Aspects* (e.g. text annotation, semantic entities such as objects and events, semantic attributes and semantic relations). The two first features, supplied by *Usage Information* (e.g. access rights, distributor), address information related to the management of the content (content management information), whereas the last two are mainly devoted to the description of perceivable information in term of structure and semantics (content description information).

The semantic descriptions have advantages over syntactic descriptions, because of their proximity to human understanding of multimedia information. They need a formalism which provides flexible representation (graph structure rather than the XML tree structure), formal and shared vocabulary, methods for expressing rules (the rule of inverse property for example), constraints and axioms as well as support of inferential mechanisms. The combination of DL and rules languages provides an interesting angle to deal such requirements.

2.3 Ontologies, Description Logic and Rules

As mentioned in Section 1, we extend the MPEG-7 standard with domain ontology supplied by a rules layer. In the following, we present the three notions implied: ontologies, DL and rules.

a) Ontologies

An ontology is defined as an explicit specification of a conceptualization [10] and it consists by several components of which the most important are concepts, relations and attributes, instances (or individuals) and axioms. It defines precisely and formally the semantics of concepts and their inter-relationships for a specific domain, and thus, it provides a shared and common understanding of a domain that can be communicated between people and application systems.

b) Description Logic : Syntax, Semantic and Knowledge Base

The use of description logics (DLs) as ontological language has been highlighted by the recent explosion of interest in the so-called "Semantic Web". OWL-DL [5] is a W3C recommendation language for ontology representation in the Semantic Web. It is a syntactic variant of the $\mathcal{SHOIN}(\mathcal{D})$ DL, offering a high level of expressivity while still being decidable. We use the DL $\mathcal{SHIQ}$ (a subset of the $\mathcal{SHOIN}(\mathcal{D})$ DL), which is enough expressive to represent the domain ontology that we consider (soccer game ontology). The syntax of the concepts and roles expressions is given in the following [13].

Let N_R be the set of role names. The set of $\mathcal{SHIQ}$ roles is the set $N_R \cup \{R^- | R \in N_R\}$. For $R \in N_R$, let $Inv(R)$ denote R^- and let $Inv(R^-)$ denote R. An $RBox$ $\mathcal{R}$ over N_R is a finite set of transitivity axioms $Trans(R)$ and role inclusion axioms $R \sqsubseteq S$, where R and S are roles, such that if $R \sqsubseteq S \in \mathcal{R}$, then $Inv(R) \sqsubseteq Inv(S) \in \mathcal{R}$ as well. The reflexive transitive closure of the role inclusion relationship is denoted with $\sqsubseteq^*$. A role not having transitive subroles is called *simple* role.

Let N_C be a set of atomic concepts names. The set of $\mathcal{SHIQ}$ concepts over N_C and N_R is defined inductively as the smallest set for which the following holds: $\top$ and $\bot$ are concepts, each atomic concept name $A \in N_C$ is a concept, if C and D are concepts and R is a role, then $C \sqcap D$ (concept conjunction), $C \sqcup D$ (concept disjunction), $\neg C$ (concept negation), $\forall R.C$ (universal quantification), $\exists R.C$ (existential quantification) are also concepts, and if C is a concept, R a simple role and n an integer, then $\geq nR.C$ and $\leq nR.C$ (number restriction) are concepts also. A $TBox$ $\mathcal{T}$ over N_C and $\mathcal{R}$ is a finite set of concept inclusion $C \sqsubseteq D$ or concept equivalence axioms $C \equiv D$, where C and D are concepts terms.

Let N_I be a set of individual (or instance) names. An $ABox$ $\mathcal{A}$ is a set of assertions of concepts and role membership axioms $C(a)$ and $R(a, b)$, where C is concept term, R is a role, and a and b are individual symbols.

A $\mathcal{SHIQ}$ knowledge base KB is a triple of the form $(KB_{\mathcal{R}}, KB_{\mathcal{T}}, KB_{\mathcal{A}})$, where $KB_{\mathcal{R}}$ is an $RBox$, $KB_{\mathcal{T}}$ is a $TBox$, and $KB_{\mathcal{A}}$ is an $ABox$.

For the semantic of the $\mathcal{SHIQ}$ DL, we utilize the recent research results of [13] about the transformation of OWL-DL KB into disjunctive Datalog KB. The transformation is based on the fact that OWL-DL is a subset of first order logic. Then, the semantic of $\mathcal{SHIQ}$ KB is given by a mapping π which transforms the KB axioms into a set of first order logic (FOL) formulae, so that the $\mathcal{SHIQ}$ DL KB is satisfiable if $\pi(KB)$ is satisfiable.

c) Datalog Rules

Let N_P be a set of predicate symbols such that $N_C \cup N_R \subseteq N_P$. A term is either a constant (denoted by a,b,c) or variable (denoted by x,y,z). An atom has the form $P(s_1, ..., s_n)$, where P is a predicate symbol and s_i are terms; a rule has the form $H \longleftarrow B_1, ..., B_n$, where H and B_i are atoms; H is called the *rule head*, and the set of all B_i is called the *rule body*. A Datalog *program* P is a finite set of horn rules with the following restrictions: Datalog allows only safe rules, i.e. every variable occurring in the head of a rule must also occur in the body of the rule and Datalog disallows the use of function symbols. For the semantics, the rule $H \longleftarrow B_1, ..., B_n$ is equivalent to the clause $H \vee \neg B_1 \vee .. \vee \neg B_n$.

3 Multimedia Framework

3.1 Architecture of the framework

Figure 2 illustrates the multimedia framework architecture, where the levels of data description and data retrieval are represented. The data description level uses metadata annotation component, ontologies, rules, the MPEG-7 database and the knowledge base; and the retrieval level uses

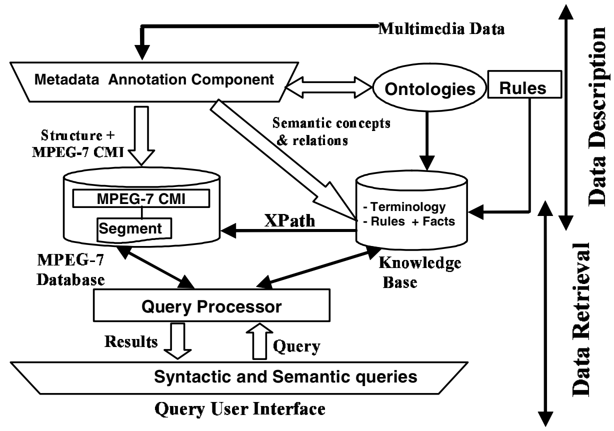

Figure 2. Framework Architecture

the query interface, query processor and the storage system composed of a knowledge base and the MPEG-7 database. In the sequel, we detail these different components and their inter-relationships.

3.2 Metadata Annotation Component

The metadata annotation component is used to annotate the multimedia data and produce the corresponding MPEG-7 description. It contains the structure (segment decomposition) of the multimedia data, the Content Management Information (CMI) as well as the low level information. For example, the annotation of a video data is done as follows: first the video is segmented with respect to the MPEG-7 $VideSegment\ DS$ (we obtain a video segment hierarchy), then; media time, creation, location and annotation information are associated to each structural unit (for example : video, video segment, moving region and still image). For the semantic part, we associate a set of logical assertions, expressed in term of the ontology vocabulary, to each structural unit considered as interesting and meaningful to the user. To deal with spatial and temporal relation between the different objects and/or media, we use a predefined set, mainly based on the MPEG-7 $SpatialRelation\ CS$ and $TemporalRelation\ CS$, and use this set in the spatio-temporal annotation of the data.

Since, the identification attribute (**ID**) is required in each MPEG-7 segment, we exploit it to connect the semantic description to the MPEG-7 description, which provide links to the raw data segment (using the MPEG-7 **MediaLocator** element). Therefore, the coupling of ontologies and rules based knowledge with metadata enriched MPEG-7 multimedia files is achieved and then, multimedia retrieval techniques should be considered for this hybrid representation.

3.3 The MPEG-7 Database

The MPEG-7 database is a set of MPEG-7 descriptions which represent the segment decomposition of the described data, the metatdata information (creator, location, duration) and may contain also a set of low level features of the data (color histogram, dominant color, texture, shape, etc.). This base is used to respond the syntactic queries.

3.4 Ontologies, rules and knowledge base components

The domain ontology is expressed using the DL $\mathcal{SHIQ}$. Then, according to the approach of [13], the $\mathcal{SHIQ}$ KB is translated into disjunctive Datalog KB using the following steps: elimination of the transitivity axioms, translation of $\mathcal{SHIQ}$ KB axioms into the clausal form, saturation of the $TBox$ and $RBox$, elimination of function symbols, elimination of irrelevant rules and conversion to disjunctive Datalog rules.

The approach uses the DL-SafeRules formalism where the following conditions have to be satisfied: concepts and roles are allowed to occur in the antecedent and consequent of the Datalog rules and each variable in a rule is required to occur in a non-DL atom in the rule body. After the translation of $\mathcal{SHIQ}$ DL KB to a Disjunctive Datalog KB, the rules KB is added to this DL $\mathcal{SHIQ}$ KB. The algorithm developed for query answering in this approach is decidable for $\mathcal{SHOIN}(\mathcal{D})$ DL without nominals ($\mathcal{SHIQ}(\mathcal{D})$).

In the sequel, we give two translation examples of the DL axioms $\sqsubseteq$ (for concepts and roles) and the transitive property of roles to first order logic formulae. The translation is done using the mapping function π.
$\pi(C \sqsubseteq D) = \forall x : \pi_y(C, x) \longrightarrow \pi_y(D, x)$, where C and D are concepts
$\pi(R \sqsubseteq S) = \forall x, y : R(x, y) \longrightarrow S(x, y)$, where R and S are roles.
$\pi(Trans(R)) = \forall x, y, z : R(x, y) \wedge R(y, z) \longrightarrow R(x, z)$
For the inverse role property, the mapping of the $\mathcal{SHIQ}$ KB to first order formulae add automatically the clause $\forall x, y : R(x, y) \longleftrightarrow R^{-}(y, x)$ to the Datalog KB. This means that all the inverse of relations are deduced and added to the Datalog KB. For example, if we have a predicate $left(x, y)$ then, $right(x, y)$ is deduced and added to the Datalog KB.

3.5 Query Interface and Query Processor

The user submit his queries through a query interface. The query constraints may be semantic and/or syntactic. Then, the query processor separates the query conditions (syntactic constraints) that could be answered by the MPEG-7 database from those that could be answered by the knowledge base (semantic constraints). The query processor is also responsible for retrieving and responding the user's queries.

We use XQuery to specify the syntactic constraint of queries. XQuery is defined as a superset of XPath [3]. The overall design of XQuery is based on a language proposal called Quilt [1]. XQuery is defined to deal with the features of XML (hierarchical, ordered, textual and potentially schema less structure). It consists of $FLWR$ (For-Let-Where-Return) expressions that support iteration and binding of variables to intermediate results. Therefore, users must understand and know the XML schema and the $FLWR$ expression to write queries in XQuery.

Semantic queries are expressed using Datalog query language i.e single horn rule which has the following form: $Q(x) \longleftarrow E_1(x_1), E_2(x_2), ..., E_n(x_n)$ where $E_1, ..., E_n$ are concepts, roles and/or ordinary predicates, x_i represent a tuple of variables and constant. The atom $Q(x)$ is the head of the query, and the result of the query is a set of tuples, each giving a binding for every variable in x.
The evaluation of a Datalog query can be done *bottom-up* (called also forward chaining); or *top-down* (called also backward chaining). The semantic queries are processed in two steps:

1. Rewrite the query, where the Datalog intentional rule are expanded to the body concepts and roles. Then, reason on domain ontology to infer concepts that are subsumed by the concepts in the query.

2. Evaluate the rewritten query with respect to the Disjunctive Datalog KB. We choose to use the *top-down* (backward chaining) approach to evaluate the semantic queries which based on resolution algorithm.

Syntactic queries, which may concern content management and/or low level information, are expressed using XQuery [4] language which is the standard language, defined by W3C, to deal with the features of XML documents. To answer a user query, the semantic part is used to search the portion of the MPEG-7 description (the XPath expression) and the variable values that satisfy the semantic specifications. From the MPEG-7 description portion, we can access the raw data (using the MPEG-7 MediaLocator descriptor) and since we get the MPEG-7 description, XQuery can be used to filter the data with respect to the syntactic constraints such as duration, creator and metadata.

4 Example Application

4.1 Domain ontology and individuals

We consider the domain ontology introduced in the Section 2.1. We recall that we consider the concept video as

a concept where at least one event takes place and at least one person appears (player, referee, etc.). We suppose that the video segmentation is driven by events. The name of the concept *Video* is the XPath expression that allow to capture the video segment in the MPEG-7 description.
For spatial and temporal relation between the different concepts, due to the space limitation, we consider just the spatial relation "right" (that will be used in the illustration example). Its inverse role is "left". These two relation are defined in the DL component as follows:
$right \sqsubseteq (Person \sqcup Object) \times (Person \sqcup Object)$;
$left \sqsubseteq (Person \sqcup Object) \times (Person \sqcup Object)$;

The semantic description of the data is done by the instantiation of the different predicates (concepts, roles and the predicates). It can be seen as a conjunction of concept and roles individuals. We suppose that the MPEG-7 description introduced in Figure 1 is stored as an XML document called "soccer.xml". We denote by $Path_1$ the XPath expression which allows to accede the first video segment ($document(soccer.xml)//VideoSegment[@id$ = "ID-V1"]) . Then the semantic part of the document will be represented with the following assertion on individuals A (notice that some concepts and roles are not defined in Section 2.1 since we gave just an extract.
$A = \{Video(Path_1), Event(Foul),$
$EventOccur(Path_1, Foul), Player(Player1),$
$Appear(Path_1, Player1), AgentOf(Player1, Foul),$
$Player(Player2), Appear(Path_1, Player2),$
$PatientOf(Player2, foul), left(Player1, Player2),$
$Object(Ball), Touch(Player2, Ball)\}$

4.2 Query evaluation

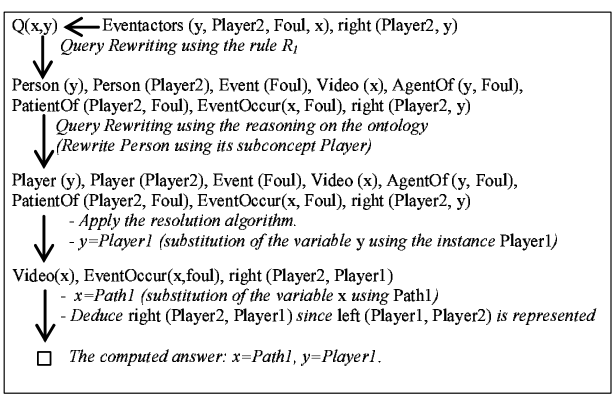

Figure 3. Query rewriting and resolution steps

Let's consider the following query: "find video segment where there is a foul on the Player2, which appear in the right of another player". This may be expressed with the intensional Datalog predicate "Eventactors" as follows: $Q(x, y) \longleftarrow Eventactors(y, Player2, Foul, x), right(Player2, y)$. The steps of the query answering are illustrated in Figure 3. We recall that we use the assertion on individuals A presented in the Section 4.1. The relation $right(Player2, Player1)$ was already added to the Datalog KB since we have a horn rule in the KB which express the inverse property of the roles $left$ and $right$.

The result of the query is the substitution $\{x/Path_1, y/Player1\}$. $Path1$ is the XPath expression which allow to accede the MPEG-7 description from where we can have the raw data (the location of video segment) and other information like metadata and content management information.

5 Related Work

Our work is related with two main topics: querying MPEG-7 description and adding semantic to the multimedia data.

There are few querying techniques proposed for multimedia content retrieval described by MPEG-7. Some approaches use directly XQuery as the retrieval language. Others define new query languages or adapt the existing ones. In [9], access methods based on an inference network model, that allows deduction of semantic and implicit information, are developed for searching MPEG-7 video descriptions. While in [12], a specification of crucial query issues in MPEG-7 XML queries is proposed and a logic formalism called *Path Predicate Calculus* is illustrated to handle the limitations in current text-oriented XML query languages for multimedia XML queries. Spatial, temporal and visual datatypes and relationships can be described in this formalism for content retrieval. Notice that since the semantic aspects are not directly expressible in XQuery, these two approaches [9, 12] proposes their own specialized query language. However, in [8] an XQuery adaptation for MPEG-7 document retrieval called *SVQL (Sematic Views Query Language)*, was proposed. *SVQL* is a high level query language that allows users to express their requirements following the semantic view model (*PhysicalView, ProductionView, ThematicView, VisualView and AudioView*) in a concise, abstract and precise way.

Our approach is complimentary to the existing ones, since we provide and hybrid query language where we combine the power of XQuery, to search the MPEG-7 descriptions, with the Datalog query language, to retrieve the multimedia data using the semantic aspect. The combination is done in a complimentary way in order to respond the user's queries.

The idea of adding semantic to multimedia data has been studied in several works. In [11], Hunter has proposed to employ RDF, RDF Schema (RDFS) and DAML+OIL to express the semantics of the MPEG-7 DSs and Ds in order to add multimedia to the semantic web. This work doesn't consider the multimedia content and the query language for the data. In [6], a multimedia framework is developed for the support of ontology-based semantic indexing and retrieval of audiovisual content following the MPEG-7 and TV-Anytime standard specifications for metadata descriptions. The domain ontology is described using the MPEG-7 abstract tools which are not provided with reasoning mechanisms, then it's not possible to infer implicit information. In [14], a framework which combine a knowledge representation language (OWL) and structural (XML Schema) was proposed to represent audiovisual documents. The content is expressed in OWL language whereas the structure is an MPEG-7 description. In [7], an ontology based approach for the knowledge assisted domain specific semantic video analysis. In this paper, the main goal was to reduce the gap between the low level features and the object layer. They didn't exploit the reasoning mechanisms provided by the DL language used to represent the knowledge. Moreover, they didn't deal with the query language of the descriptions.

Our work is quite similar to the work of [6] and [14], but we deal the representation of the multimedia data from a new point of view. Our contribution is the combination of the MPEG-7 and OWL-DL enhanced by rules to describe the multimedia content and the use of logical conjunctive queries and XQuery query languages for the query specification. The hybrid language (OWL-DL and Datalog rules) used provides a decidable reasoning algorithm to answer conjunctive queries. These combinations have not been considered in previous works.

6 Conclusion

In this paper, a framework of multimedia enriched domain ontologies have been presented, to be used in several tasks including content description and reasoning, as well as content retrieval. Our aim was to exploit the MPEG-7 abilities to represent the syntactic dimension of the multimedia data, and introduce a logical formalism, based on OWL-DL supplied by Datalog rules and more adequate, to represent the semantic dimension. XQuery is used to specify the metadata and structural part of the queries, whereas the Datalog query language is used to specify the semantic constraints. The MPEG-7 database is used to answer XQuery queries, whereas the knowledge base is used to answer the semantic queries.

For future work, we plan to explore this framework in an a real application and study the performances of the hybrid retrieval language (recall and precision). We plan also to investigate the query relaxation and/or rewriting and consider query optimization techniques to get neighboring queries. Another direction of this work, is to formalize the structure of the MPEG-7 description in order to provide reasoning over the structure and the content of multimedia. We plan also to consider the framework in a distributed system, where the data are stored in different data sources.

References

[1] Quilt : an XML Query Language. http://www.almaden.ibm.com/cs/people/chamberlin/quilt-euro.html.

[2] XML Schema Part 1: Structures, W3C recommandation, 2 May 2001. http://www.w3.org/TR/xml-schema-1/.

[3] XPath: XML Path Language. http://www.w3.org/TR/xpath.

[4] XQuery 1.0 : An XML Query Language. W3C Working Draft 12 November 2003. http://www.w3.org/TR/xquery/.

[5] F. Baader, D. Calvanese, D. L. McGuinness, D. Nardi, and P. F. Patel-Schneider, editors. *The description logic handbook: theory, implementation, and applications*. Cambridge University Press, New York, NY, USA, 2003.

[6] S. C. Ch. Tsinaraki, E. Fatourou. An Ontology-Driven Framework for the Management of Semantic Metadata Describing Audiovisual Information. In *Proceedings of the 15th International Conference on Advanced Information Systems Engineering (CAiSE)*, Austria, June 2003.

[7] S. Dasiopoulou, V. Papastathis, V. Mezaris, I. Kompatsiaris, and M. Strintzis. An Ontology Framework for Knowledge Assisted semantic Video Analysis and Annotation. In *Proceedings 4th International Workshop on Knowledge Markup and Semantic Annotation*, November 2004.

[8] N. Fatemi, O. A. Khaled, and G. Coray. An XQuery adaptation for MPEG-7 documents retrieval. *XML conference and Exposition*, Decemeber 2003.

[9] A. Graves and M. Lalmas. Video retrieval using an MPEG-7 based inference network. In *The* 25^{th} *Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 339– 346. ACM Press, August 2002.

[10] T. Gruber. A translation approach to portable ontology specifications. *Knowl. Acquis.*, 5(2):199–220, 1993.

[11] J. Hunter. Adding multimedia to the semantic web - building an MPEG-7 ontology. In *Proceedings of the First Semantic Web Working Symposium (SWWS)*, pages 261– 281, 2001.

[12] P. Liu and L. Sushu. Queries of digital descriptions in MPEG-7 and MPEG-21 XML documents. *XML Europe 2002, Barcelona, Spain*, May 2002.

[13] B. Motik, U. Sattler, and R. Studer. Query answering for OWL-DL with rules. In *Proc. of International Semantic Web Conference ISWC 2004*, 2004.

[14] R. Troncy. *Formalisation des connaissances documentaires et des connaissances conceptuelles l'aide d'ontologies : application la description de documents audiovisuels*. Phd thesis in computer science, Joseph Fourier University, Grenoble (FR), 2004.

[15] A. Vakali, M.-S. Hacid, and A. Elmagarmid. Mpeg-7 based description schemes for multi-level video content classification. *Image Vision Comput*, 22(5):367–378, 2004.