L7, an MPEG-7 Query Framework

Rubén Tous and Jaime Delgado
Distributed Multimedia Applications Group (DMAG)
Universitat Politècnica de Catalunya (UPC), Dpt. d'Arquitectura de Computadors
Jordi Girona, 1-3. Building D6 Campus Nord. E-08034 Barcelona, Spain
{rtous, jaime.delgado}@ac.upc.edu

Abstract

Many vendors of multimedia databases and many providers of multimedia search and retrieval services already offer advanced indexing and retrieval techniques for multimedia contents. However, the solutions differ and do not interoperate. The already completed ISO/IEC IS 15938, formally called the Multimedia Content Description Interface (better known as MPEG-7), provides a rich model for multimedia content description. We propose a unified query framework for multimedia repositories based on MPEG-7. The framework consists of a query language and a service interface. The query language is based on XML and includes four complementary parts defining a relational-like algebra, an expression language, an input query structure and the output response format. The service interface acts as a set of views over the query language and is based in two different bindings: An extensible Web Services-based (SOAP) binding and an alternative and simple HTTP (REST-like) interface.

1 Introduction

1.1 Unified multimedia querying languages and interfaces

The value of the huge amount of multimedia information that is available today depends on how easily we can find, retrieve and access it. Search and retrieval of audiovisual materials by its content and metadata is a very relevant and challenging task. User information needs could be expressed in terms of metadata, text keywords, content-appearance relations, spatio-temporal relations, explicit low-level features and also semantic relationships. In addition to search conditions, multimedia content providers can offer other functionalities, like search personalization (e.g. recommendation), usage information (e.g. ratings) or even user collection management (insertion, updating and deletion of user contents).

Many vendors of multimedia databases and many providers of multimedia search and retrieval services already offer advanced indexing and retrieval techniques for multimedia contents. However, their databases and service interfaces are proprietary and therefore the solutions differ and do not interoperate. The definition of unified languages and interfaces to accept and respond to requests for multimedia contents searches would facilitate multimedia repositories interoperability. Interoperability would ease the access to repositories by users and applications, and would allow the deployment of distributed search and aggregation services.

1.2 Multimedia information retrieval vs. multimedia data retrieval

The aim of an Information Retrieval (IR) system is to facilitate the user access to the information in which he is interested. However, the *user information needs* cannot be easily formalized, and they must be translated into a query processable by the retrieval system. Given the user query, the IR system aims to retrieve information which might be relevant to the user.

On the other hand, a data retrieval system aims to determine which objects of a collection satisfy clearly defined conditions as those in a relational algebra expression. For a data retrieval system, like a database, a single erroneous object among a thousands retrieved means a total failure. So, data retrieval deals with well defined data models, expressive query languages and performance issues, while information retrieval faces the problem of interpreting the contents of the information items to decide their relevance.

In general, but specially in the context of multimedia search and retrieval, these two concepts are not isolated. Today digital contents are often composite of the contents themself and also machine understandable metadata. Querying such complex objects can imply the combination of data retrieval-like conditions -referred to a well defined data model- and also information retrieval-like conditions (e.g. keywords or query-by-example).

1.3 MPEG-7 query framework

The already completed ISO/IEC IS 15938, formally called the Multimedia Content Description Interface (better known as MPEG-7), provides a rich model for multimedia content description. The usage of MPEG-7 should guarantee interoperability at data retrieval level. MPEG-7 descriptions link to standard schemas and should be interoperable with respect to XML query languages like W3C's XQuery [11].

However, data retrieval is not the only feature of a modern multimedia database. Some content-based search and retrieval constraints (e.g. query-by-example or confidence management) must be also combined with the use of MPEG-7 descriptions. Furthermore, some MPEG-7 descriptors encapsulate low-level features (e.g. a colour histogram) which contain data, but these data are only useful if they can be queried and processed with specific content-based functionalities (e.g. various similarity functions). In this sense, XQuery does not provide specific normative functionalities for query-by-example or in general for content similarity measurement. A language which combines the expressive power of XQuery with content-based functionalities (e.g. an extension of XQuery) could be a good candidate for an unified query language for MPEG-7 multimedia databases. Standardizing such language could raise interesting benefits from the multimedia applications development point of view, in a similar way in which SQL did for applications based on relational data -platform independence, reusability, maintainability, etc.-.

However, such kind of language, heavily coupled with low-level details and very expressive, wouldn't be probably suited as a language for interconnecting multiple multimedia repositories in a distributed search or aggregation scenarios. Experiences from the Digital Libraries discipline, like the Protocol for Metadata Harvesting [6] of the Open Archives Initiative (OAI), or the SRW/SRU languages of the Z39.50 Maintenance Agency, show that simple and low-barrier interfaces are best suited for such contexts.

At the user level, information needs are not expressed in terms of query algebras or XML paths. Users express their needs without knowing the existence of a well defined data model, outside the realm of data retrieval. Many providers of multimedia search and retrieval services already offer interfaces for transmitting user information needs. Because standardized interfaces are not defined, each repository offers its own search interface, which prevents clients experiencing aggregated services from various MPEG-7 databases.

So, multimedia search and retrieval systems offer different layers of functionality which are suitable of being standardized (see figure 1).

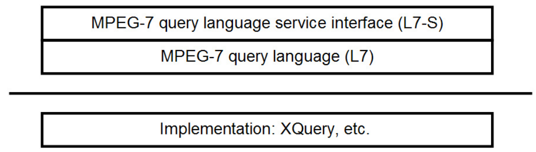

Figure 1. MPEG-7 querying layers

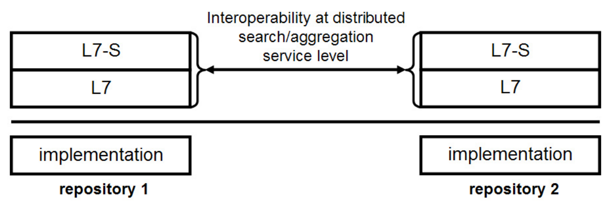

Figure 2. Proposed MPEG-7 querying layers

This work focuses in the first two layers, related the definition of an unified query language and a service interface to accept and respond to user requests in a distributed multimedia search and retrieval scenario (see figure 2). The local request processing, by its translation to a specific database language or by other ways, is out of the scope of the work.

2 Related work

Querying of audiovisual contents is not a new research problem; we can find old works facing this challenge from different approaches. In the past, when there was no metadata linked to multimedia contents, systems used to rely only on the analysis of low-level features and IR-like approaches. From the users point of view, in a pure IR approach queries are formulated using sample documents (Query-By-Example or QBE), rough sketches, or component features (outline of objects, colour, texture, shape, layout).

On the other side, and also with a long tradition, we find works based on high level query languages. These works are based on the definition of data models for multimedia (e.g. relational and object-oriented).

Today, these kind of works are based on MPEG-7 descriptors and the MPEG-7 data model (actually it is a data schema under the XML data model). Some simply defend the use of XQuery or some extensions of it like [4, 9]. Others define a more high-level and user-oriented approach like [3, 5, 2].

Now, the issue has been refuelled by the initiative of standardization of an MPEG-7 Query Format taking place within the MPEG context(ISO/IEC JTC1/SC29/WG11). The process started in July 2006 with the release of a "Call for Proposals on MPEG-7 Query Format" (ISO/IEC JTC1/SC29/WG11 N8220). A proposal containing parts of the research work presented in this paper has been submitted in January 2007 as a response to the Call.

With respect to the previous works, the language presented here is based on MPEG-7 as [4, 9, 3, 5, 2]. It outperforms XQuery-based approaches like [4, 9] because, while offering the same level of expressiveness, it offers multiple content-based search functionalities (QBE, query-by-keywords) and other IR-like features (e.g. paging, relevance feedback, personalization). It differs from high level approaches like [3, 5, 2] because it keeps working over the data model defined by MPEG-7, and does not attempt to define a higher-level data model.

3 MPEG-7 query Language (L7)

The L7 (a Language for MPEG-7 querying) is a language we are proposing for expressing user information needs and data retrieval constraints over an MPEG-7 multimedia repository. The L7 can be serialized in XML using four XML schemas which correspond to four complementary parts of the language:

1. Algebra language, which allows applying relational-like operators to sequences of description or content items.
2. Condition expression language, which allows to filter a sequence of items by combining different types of conditions and arithmetic operations. It is also applied to a pair of sequences of items in the "join" operator.
3. Input query structure, which allows specifying constraints about the format of the results, like ordering, grouping, paging, etc.
4. Output response structure, which defines the valid structure for an L7 response instance.

Following are outlined the different parts of the language. The XML schemas are not included here to save space.

3.1 L7 data model

The L7 data model defines all permissible values of expressions in the L7 language. A language is closed with respect to a data model if the value of every expression in the language is guaranteed to be in the data model. L7 is closed with respect to its data model.

L7 operates over MPEG-7 and its data model, but the L7 data model can represent various values including not only the input and the output of a query, but all values of expressions used during the intermediate calculations. Examples include the result of an arithmetic expression (represented as an atomic value), an expression resulting in a sequence of items, etc. Definitions 1, 2 and 3 express this idea formally.

Definition 1. *Every instance of the data model $\mathcal{M}$ is a sequence $\mathcal{S}$. A sequence $\mathcal{S}$ is an ordered collection of zero or more items $\mathcal{I}$.*

Definition 2. *An item $\mathcal{I}$ can be an MPEG-7 document, any part of an MPEG-7 document, the composition of parts of different MPEG-7 documents or a literal value compliant with the datatypes defined for XML Schema and MPEG-7 DDL. A single item $\mathcal{I}$ appearing on its own is modelled as a sequence $\mathcal{S}$ containing one item. An item cannot be a sequence.*

Definition 3. *An MPEG-7 multimedia repository $\mathcal{R}$ is modelled by one sequence $\mathcal{S}$ containing items which correspond to MPEG-7 documents (documents beginning with the "Mpeg7" element or documents containing description units).*

3.2 L7 algebra language

The L7 algebra describes the basic operators which operate over the L7 data model. Definition 4 enumerates the available operators.

Definition 4. *The available operators of L7 algebra are "projection" (π), "selection" (σ), "join" ($\bowtie$), "union" ($\cup$) and "difference" ($-$). An operator $\mathcal{O}$ operates on one or more sequences $< S_1, S_2, ..., S_N >$ to yield a sequence $\mathcal{S}$.*

Note that the selected operators are inspired on the relational algebra operators and, though they have similar semantics, they operate over a different data model.

The L7 algebra allows combining five basic operators to filter multimedia contents and metadata. The usage of the algebra allows filtering the MPEG-7 descriptions of a repository (and its related contents) in a flexible way. For example the same sequence of descriptions can be filtered by two different selection methods and then combined to obtain a joined result. Without the definition of the algebra (e.g. defining only an condition expression language) one could only apply a one-pass filter to the multimedia objects of a repository.

The five primitive operators chosen for the algebra are the selection, the projection, the join (which can be a Cartesian product if no condition is specified), the set union and the set difference. Other operators can be defined in terms of the primitive ones, as the set intersection or the different kinds of joins and outer-joins. These five operators are fundamental in the sense that none of them can be omitted without losing expressive power. However, the choice of which operators are primitive and which derived, is somewhat arbitrary, as it is the choice in logic of AND, OR and NOT.

3.2.1 L7 algebra in BNF

The complete algebra language in BNF (except for the booleanExpression rule, which is listed in the next section) can be find in Code 1.

Code 1 L7 algebra BNF

⟨operation⟩→⟨selection⟩| ⟨projection⟩| ⟨join⟩| ⟨union⟩ | ⟨difference⟩
⟨selection⟩→⟨operationInput⟩ ⟨booleanExpression⟩
⟨join⟩→⟨operationInput⟩ ⟨operationInput⟩ ⟨booleanExpression⟩
⟨union⟩→⟨operationInput⟩ ⟨operationInput⟩
⟨difference⟩→⟨operationInput⟩ ⟨operationInput⟩
⟨projection⟩→⟨operationInput⟩ ⟨path⟩
⟨operationInput⟩→⟨operation⟩| ⟨source⟩
⟨source⟩→⟨description⟩| ⟨variableReference⟩| **all**
⟨description⟩→!!Any part of an MPEG-7 description
⟨variableReference⟩→!!The id of a variable
⟨booleanExpression⟩→!!See the expression language BNF
⟨path⟩→!!XPath expression or an abbreviated path

3.2.2 A simple example

The XML serialization of the L7 algebra allows expressing algebra trees in XML. The following XML fragment is just a simplified example showing the structure of a possible L7 algebra instance for the tree in Figure 3. This example query serves to search for videos whose audio content is similar to the audio contents authored by ”Cohen” and present in the repository:

3.2.3 Sources, paths and abbreviated paths

In the L7 algebra, a valid input for an operation is the output of another operation or a ”source”. The keyword ”all” defines a source name referred to the complete multimedia repository. It can be seen as a huge XML document with a sequence of ”mpeg7” tags containing descriptions or description units. The projection operation allows selecting a subset of the repository by using paths (expressed with the XPath language) or abbreviated paths. Abbreviated paths are simply names of elements or schema types which are used instead of (//name or //@xsi:type=”name”).

Other possible ”sources” are inline descriptions and variable references. Inline descriptions allow including example descriptions and also example contents (using the MPEG-7 MediaURI or the InlineMedia elements). Variable references allow reusing source and operation definitions.

Code 2 L7 algebra example

```
<operation type="tJoin">
  <operation type="tXPathProjection" id="a">
    <source type="tAll" />
    <path>//mpeg7:Video</path>
  </operation>
  <operation type="tSelection" id="b">
    <operation type="tXPathProjection">
      <source type="tAll" />
      <path>//mpeg7:Audio</path>
    </operation>
    <booleanExpression
      type="tStringEqualToExpression">
      <path type="tPath">//Creator//FamilyName</path>
      <stringExpression type="tStringLiteral">
        <stringLiteral>Cohen</stringLiteral>
      </stringExpression>
    </booleanExpression>
  </operation>
  <booleanExpression type="tMediaSimilarityExpression">
    <path type="tPath" idref="a">
     //mpeg7:MediaLocator
    </path>
    <path type="tPath" idref="b">
     //mpeg7:MediaLocator
    </path>
  </booleanExpression>
</operation>
```

3.3 L7 condition expression language

The selection and join operators evaluate an expression which returns a boolean value for each one of the items (or pair of items in the case of join) of the input sequence/s. This expression must return a boolean value, so it must be a comparison expression or the combination of comparison expressions with boolean operators. The language offer the four basic boolean operators (AND, OR, NOT and XOR).

Comparison expressions allow two values to be compared. The L7 expression language provides multiple types of comparison expressions, depending on the involved in-

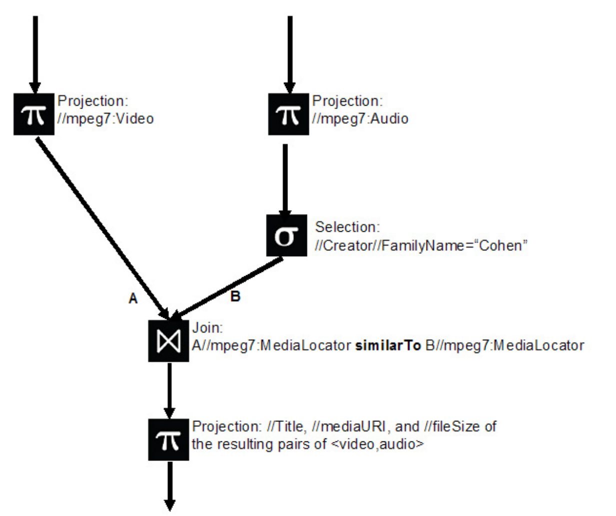

Figure 3. L7 algebra example

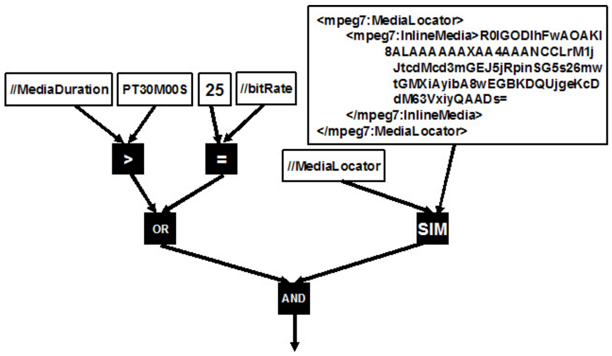

Figure 4. L7 expression language example

put values: Arithmetic comparison expressions ($>$, $<$, $\geq$, $\leq$ and $\neq$), string comparison expressions ($equals$, $equalsIgnoreCase$, $match$ and $contains$), keywords-based comparison expressions ($containsKeywords$) and content-based comparison expressions ($similarTo$). The values which participate in a comparison can be single values, the result of an arithmetic expressions ($+$, $-$, $*$, $/$, abs, $ceiling$, $floor$, $round$, $indexOf$, $sizeOf$) or string expressions ($substring$, $concat$).

The single values which participate in a comparison expression or in an arithmetic/string expressions can be literals, paths within the items (e.g. a path to the CreationDate) or description fragments (e.g. a MediaLocator). Because in a join operation two different sequences of items can be accessed, the paths can include a sequence identifier to clarify to which sequence they belong.

The graphical example in Figure 4 ilustrates the usage of the expression language.

The complete expression language in BNF (except for the literals and paths) can be find in Code 3:

3.4 The complete L7 language

3.4.1 Input query format

The L7 language defines the format of queries (input) and responses (output) to be interchanged between an L7 client and an L7 server (one with a total support of language functionality). The L7 includes the algebra language and also the expression language defined above, but it also adds some parameters related to client preferences (e.g. paging, relevance feedback, etc.).

The additional parameters available for the language are the *inserted-from* (datestamp to harvest only those records that were inserted after the specified date), *inserted-until* (datestamp to harvest only those records that were inserted before the specified date), *maxItems* (maximum number of items within the same response), *numPage* (number of the response page related to this request), *timeout*, *answer-id*

Code 3 L7 expression language BNF

```
⟨booleanExpression⟩→
⟨arithmeticComparisonExpression⟩|
⟨stringComparisonExpression⟩|
⟨keywordsComparisonExpression⟩|
⟨mediaComparisonExpression⟩|
⟨booleanExpression⟩ ⟨binaryBooleanOperator⟩
⟨booleanExpression⟩|
⟨unaryBooleanOperator⟩ ⟨booleanExpression⟩
⟨binaryBooleanOperator⟩→and|or|xor
⟨unaryBooleanOperator⟩→not
⟨arithmeticComparisonExpression⟩→
⟨artihmeticExpression⟩ ⟨arithmeticComparisonOperator⟩
⟨artihmeticExpression⟩
⟨arithmeticComparisonOperator⟩→=|>|<|≥|≤|≠
⟨artihmeticExpression⟩→⟨numericValue⟩|
⟨stringExpression⟩ indexOf ⟨stringExpression⟩|
⟨stringExpression⟩ sizeOf ⟨stringExpression⟩|
⟨artihmeticExpression⟩ ⟨artihmeticBinaryOperator⟩
⟨artihmeticExpression⟩|
⟨artihmeticUnaryOperator⟩ ⟨artihmeticExpression⟩
⟨artihmeticBinaryOperator⟩→+|-|*|/
⟨artihmeticUnaryOperator⟩→abs|ceiling|floor|round
⟨stringComparisonExpression⟩→⟨stringExpression⟩
⟨stringComparisonOperator⟩ ⟨stringExpression⟩
⟨stringExpression⟩→⟨stringValue⟩| ⟨stringExpression⟩
substring ⟨artihmeticExpression⟩ ⟨artihmeticExpression⟩|
⟨stringExpression⟩ concat ⟨stringExpression⟩
⟨stringComparisonOperator⟩→match|contains|equals|
equalsIgnoreCase
⟨numericValue⟩→⟨numericLiteral⟩|⟨path⟩
⟨stringValue⟩→⟨stringLiteral⟩|⟨path⟩
⟨mediaComparisonExpression⟩→⟨mediaContent⟩
⟨similarityOperator⟩ ⟨mediaContent⟩
⟨mediaContent⟩→⟨mediaLocator⟩|⟨path⟩
⟨similarityOperator⟩→ sim ⟨artihmeticExpression⟩
⟨keywordsComparisonExpression⟩→⟨path⟩ containsKey-
words ⟨stringExpression⟩
```

(answer-id of a previous response for paging), *fromPreviousResponse* (answer-id of a previous response for searching within previous results), *order-by*, *group-by*, *personalization* (personalized/biased result), *relevanceFeedbackAnswerId* (feedback about a previous answer), *relevanceFeedbackRecordNum* (relevant record from a previous answer), *resultsFormat* (the default response format is XML, but other formats can be specified).

Example in Code 4 contains a complete example query, which makes use of some of the additional input parameters. The query tries to find videos which contain one or more keyframes similar to a given image. The additional parameters signal the server that the client wants responses (pages) with no more than 10 records, they also signal that the client wants the second page and that he wants the records to be ordered by mpeg7:Title and mpeg7:CreationDate.

Code 4 L7 language example query

```
<query>
    <maxItems>10</maxItems>
    <numPage>2</numPage>
    <orderBy>
        <path type="tPath">Title</path>
        <path type="tPath">CreationDate</path>
    </orderBy>
    <operation type="tSelection">
      <operation type="tXPathProjection">
        <source type="tAll" />
        <path>Video</path>
      </operation>
      <booleanExpression type="tMediaSimilarityExpression">
          <path type="tPath">MediaLocator</path>
          <mpeg7:MediaLocator>
            <mpeg7:InlineMedia>R0lGODlhFwAOAKIAAP////Dw8O
                         Dg4MDAwIqN1oCAgAAAAAAAACH5BAQUAP
                         8ALAAAAAAXAA4AAANCCLrM1jAqY8iTkp
                         JtcdMcd3mGEJ5jRpinSG5s26mwTKUfF1
                         tGMXiAyibA8wEGBKDQUjgeKcDg5vGDRi
                         dM63VxiyQAADs=
            </mpeg7:InlineMedia>
        </mpeg7:MediaLocator>
      </booleanExpression>
    </operation>
</query>
```

3.4.2 L7 Output format

The responses from an L7 compliant server will always include a sequence of records (under the "record" tag), and within each record the client can find a complete MPEG-7 description, a fragment of an MPEG-7 description, a combined fragment of different descriptions or even a literal value. Additionally, the server always includes other useful information fields, which are the *answerId*, *expirationDate*, *numPage* (number of the page related to the response within the same response), *numPages* (total Number of resulting response), *numItems* (total number of items of the response), *requestId*, *statusCode* (exceptions and warnings) and the *recordNumber* (identifier of each record).

Example in Code 5 contains a complete example response. The response contains the second page of an answer to a query which asks for images.

Code 5 L7 language example response

```
<response>
    <answerId>mp7qf:1394123432</m:answerId>
    <expirationDate>
      2007-11-30T16:25:00.000-05:00
    </expirationDate>
    <numPage>2</numPage>
    <numPages>2</numPages>
    <numItems>12</numItems>
    <requestId>mp7qf:3394143532</m:requestId>
    <statusCode>00000</statusCode>
    <record id="urn:foo:di:X9072663/I001">
         <mpeg7:Title>Item 01</mpeg7:Title>
         <mpeg7:MediaURI>
           http://www.someprovider.com/items/item01.jpg
         </mpeg7:MediaURI>
     </record>
    <record id="urn:foo:di:X9072663/I002">
         <mpeg7:Title>Item 02</mpeg7:Title>
         <mpeg7:MediaURI>
           http://www.someprovider.com/items/item02.jpg
         </mpeg7:MediaURI>
    </record>
</response>
```

4 MPEG-7 service interface (L7-S)

The definition of an MPEG-7 query language is just one of the steps to achieve a real interoperability between multimedia repositories. It defines a unified way to query MPEG-7 multimedia databases in the most flexible way, without any constraints or assumptions about the target scenarios. However, because not all repositories need to be interested or need to be capable of exposing all possible query functionalities, a service interface layer is needed. This higher level layer will allow selective service level adoption, and will also simplify the task of aggregators and distributed search engines.

If we want users experiencing aggregated services from various MPEG-7 databases we need to define a unified and low-barrier service interface over our query language as it exists in the Digital Libraries discipline. Experiences like the Protocol for Metadata Harvesting (PMH) of the Open Archives Initiative (OAI) [6], or the SRW/SRU languages of the Z39.50 Maintenance Agency, show that simple and low-barrier interfaces are best suited for a distributed search or aggregation scenario. So, we need to define a low-barrier and scalable interface that allows new and old repositories to expose their contents by selecting a subset of query functionalities, which could range from simple metadata aggregation to complex query-by-example searches. We suggest:

1. The use of an extensible Web Services-based (SOAP) [8] interface for querying and managing audiovisual

contents described with MPEG-7. Each exposed functionality will be represented by an operation (in web services terminology). Operations will act as views over the L7 language, limiting the expressivity of the L7 queries for a specific use case.

2. An alternative and simple HTTP (REST-like [7]) interface (being the request a simple URI and the response an XML file) when possible and just for operations involving searches compliant with the SOAP 1.2 web-method recommendation (idempotent, free of side effects, etc.).

4.1 SOAP and HTTP bindings

The service is defined through two different bindings, the W3C's SOAP protocol (Simple Object Access Protocol) [8] which is currently at version 1.2, and a raw usage of the HTTP protocol and XML.

All operations will be defined as web services operations and described in WSDL 2.0. All operations will offer a SOAP 1.2 binding with both, request and response, expressed in XML. Additionally, some operations will offer also an HTTP binding which will allow embedding the request inside an URI. This URI-based invocation adheres to the SOAP 1.2 - Part 2 section 4.1 recommendations and follows the Representational State Transfer (REST) architectural style [7]. These two possibilities have been chosen following the model of the SRW/SRU languages of the Z39.50 Maintenance Agency, which combine some features of the ISO 8601:1988 (Z39.50) [12].

4.2 Queries within URIs

Operations that can be invoked through the HTTP binding need to include a special serialization of a limited version of L7. The L7-URI syntax is inspired on the syntax of CQL [1], the Common Query Language, in combination with some datatypes and functions defined for XQuery. CQL is a formal language for representing queries to Information Retrieval systems developed by the Z39.50 (ISO 23950) Maintenance Agency as part of its ZING initiative ("Z39.50-International: Next Generation"). CQL combines the simplicity of intuitive languages like Google searching with the expressive power of the more complex languages like SQL, W3C's XQuery or the Z39.50 Type-1 query. An example instance of L7-URI could be:

```
mediaType=VideoType and (creationDate=1899-12-30
or creatorFamilyName=Smith)
```

When sending a request through HTTP GET, with the L7-URI instance embedded within the query part of an URI, some characters must be percent-encoded as defined in RFC 3986. The previous example will be converted to:

```
http://www.someprovider.com/mp7qf/REST/
getMultimediaByTextualDescription?
where=mediaType%3DVideoType%20and(creationDate%3D1899-12-30
%20or%20creatorFamilyName%3DSmith
```

4.3 List of operations

We have partitioned the MPEG-7 Query functionality in a set of independent operations. MPEG-7 querying defines a multidimensional space of functionalities, which have 1) different levels of deployment difficulty (e.g. keywords-based querying can be easier to deploy than a query-by-example functionality), 2) different contexts and domains (e.g. audio content providers and video content providers) and 3) transversal aspects (e.g. user authentication or digital rights management). Figure 5 enumerates the list of defined operations.

OPERATION	DESCRIPTION	REST
abstractRetrieval	Abstract	No
getMultimedia	Harvesting	Yes
getMultimediaBytTextualDescription	Description	Yes
getMultimediaByKeywords	Keywords search	Yes
getImageByExampleDescription	Query by description	No
getImageByExampleImage	QBE	No
getImageByExampleAudio	QBE	No
getImageByExampleSegment	Segment-based QBE	No
getVideoByExampleDescription	Query by description	No
getVideoByExampleKeyframe	Keyframe QBE	No
getVideoByExampleAudio	Audio QBE	No
getVideoByExampleVideo	QBE	No
getVideoByExampleSegment	Segment-based QBE	No
getAudioByExampleDescription	Query by description	No
getAudioByExampleAudio	QBE	No
getAudioByExampleImage	Image QBE	No
getAudioByExampleSegment	Segment-based QBE	No
getMultimediaBySemantics	Query by semantics	Yes
abstractModify	Abstract	No
insert	Updating – insert	No
update	Updating – update	No
delete	Updating – delete	No
setUserPreferences	Updating - preferences	No
getUserPreferences	User profile	No
getUserActions	Personal usage info	No
getRecommendations	Recommended search	Yes
getRepositoryInformation	Adminsitrative	Yes

Figure 5. Service interface operations list

5 Conclusions and future work

The presented work describes a unified query framework for multimedia repositories based on MPEG-7. Its purpose is to define a unified way to query multimedia databases and also to allow clients experiencing aggregated services from various MPEG-7 repositories.

The framework consists of a query language and a service interface. We have defined the four parts of the query language, which are a relational-like algebra, an expression language, an input query structure and the output response format. An XML schema is provided for each one of the four parts of the language in order to allow the serialization of language instances in XML.

The work also defines a service interface which acts as a set of views over the query language layer. This interface offers two different bindings: An extensible Web Services-based (SOAP) [8] binding and an alternative and simple HTTP (REST-like [7]) interface.

Currently we have a partial implementation of a reference query processor and we are working to release a complete reference implementation for both, a query processor and the service interface (with the SOAP and HTTP bindings). In parallel the work is being discussed under the standardization process of an MPEG-7 Query Format, taking place within the MPEG context (ISO/IEC JTC1/SC29/WG11). The process started in July 2006 with the release of a "Call for Proposals on MPEG-7 Query Format" (ISO/IEC JTC1/SC29/WG11 N8220). Parts of this work were submitted as a response to the Call [10] and were evaluated during the 79th MPEG meeting in Marrakech, Morocco, January 2007. Parts of the submission were selected to be part of the standard, for which a first Committee Draft (CD) was approved during the 80th MPEG meeting in San Jose, USA, April 2007. It is expected that the MPEG Query Format (MPQF) will become an ISO standard after the 85th MPEG meeting in Hannover, Germany, July 2008.

6 Acknowledgments

This work has been partly supported by the Spanish government (DRM-MM project, TSI 2005-05277) and the European Network of Excellence VISNET-II (IST-2005-2.41.6), funded under the European Commission IST 6th Framework Program. It is worth mentioning that the referred submission to MPEG was a joint contribution from several VISNET-II partners.

References

[1] Common query language. See http://www.loc.gov/z3950/agency/zing/cql/cql-syntax.html.

[2] N. Fatemi, O. Khaled, and G. Coray. An xquery adaptation for mpeg-7 documents retrieval. See http://www.idealliance.org/papers/dx_xml03/papers/05-03-01/05-03-01.html.

[3] A. Graves and M. Lalmas. Video retrieval using an mpeg-7 based inference network. In *ACM SIGIR'02*, August 2002.

[4] J. Kang and al. An xquery engine for digital library systems. In *3rd ACM/IEEE-CS Joint Conference on Digital Libraries*, Houston, Texas, May 2003.

[5] P. Liu and L. Suhsu. Queries of digital content descriptions in mpeg-7 and mpeg 21 xml documents. In *XML Europe 2002*, Barcelona, Spain, May 2002.

[6] The open archives initiative. See http://www.openarchives.org/.

[7] Representational state transfer (rest), roy thomas fielding, author. 2000. See http://www.ics.uci.edu/ fielding/pubs/dissertation/top.htm.

[8] Soap version 1.2 part 1: Messaging framework. w3c recommendation 24 june 2003. See http://www.w3.org/TR/SOAP/.

[9] D. Tjondronegoro and Y. Chen. Content-based indexing and retrieval using mpeg-7 and xquery in video data management systems. *World Wide Web: Internet and Web Information Systems*, pages 207–227, 2002.

[10] R. Tous, J. Delgado, E. Peig, E. Rodrguez, A. Carreras, G. Cordara, F. Gianluca, F. Dufaux, and G. Galinski. Mpeg-7 search and retrieval web service: A response to the mpeg-7 query format call for proposals., January 2007. ISO/IEC JTC1/SC29/WG11 MPEG2007/M14117.

[11] Xquery 1.0: An xml query language. w3c proposed recommendation 21 november 2006. See http://www.w3.org/TR/xquery/.

[12] Z39.50. See http://www.ietf.org/rfc/rfc2616.txt.