QUEEN MARY
AND WESTFIELD COLLEGE
UNIVERSITY OF LONDON

Department of Computer Science

Research Report No. RR-03-02 ISSN 1470-5559 July 2003

Searching Multimedia Data using MPEG-7 Descriptions in a Broadcast Terminal

Mounia Lalmas, Benoit Mory, Katerina Moutogianni, Wolfgang Putz & Thomas Rölleke

SEARCHING MULTIMEDIA DATA USING MPEG-7 DESCRIPTIONS IN A BROADCAST TERMINAL

Mounia Lalmas[1], Benoit Mory[2], Katerina Moutogianni[1], Wolfgang Putz[3] and Thomas Rölleke[1]

[1]Queen Mary, University of London, E1 4NS London, England, United Kingdom
{mounia,km,thor}@dcs.qmul.ac.uk
[2]Laboratoires d'Électronique, Philips, 94453 Limeil Brévannes, France
benoit.mory@philips.com
[3]Fraunhofer-Institut Integrierte Publikations- und Informationssysteme (IPSI), D-64293 Darmstadt, Germany
wolfgang.putz@ipsi.fraunhofer.de

Abstract

MPEG-7 is an emerging standard for representing information carried by multimedia data. Such a standard is considered crucial for the oncoming integration of broadcast (TV) and Internet technologies and applications. This paper reports on the development of methods for searching multimedia data using MPEG-7 in the context of the SAMBITS application, and, in particular, in the development of a SAMBITS terminal. SAMBITS, a European broadcast application project, developed a multimedia studio and terminal technology for new interactive broadcast services based on the MPEG standards. This paper describes, after an introduction to MPEG-7 and a description of SAMBITS, a retrieval model for MPEG-7 in a broadcast terminal. The retrieval model was developed and implemented using the HySpirit software, which is a flexible framework for representing complex data and describing retrieval functions effectively. A user interface was also implemented to provide insights in integrating a search functionality in a broadcast terminal.

1. Introduction

The impact of multimedia data in our information-driven society is growing since tools for creating and manipulating multimedia data are becoming widely available. While the first generation of multimedia processing concentrated mainly on "re-playing" the data and users consumed the information directly, the second generation of multimedia tools supports the increasingly digital creation and manipulation of multimedia data. In the first generation, multimedia data was mainly gained from translating analogous data sources into digital data sources. In the second generation, we find real-time recording of digital data.

Such multimedia data sources need to be effectively and efficiently searched for information of interest to users or filtered to receive only information satisfying users' preferences. This may be the case for scenarios such as the recording and use of broadcast programmes, multimedia teaching material in educational and training institutes, or general multimedia data in security agencies, national archival centres and libraries, journalism, tourism and medical applications. However, the information extraction and semantic analysis is still a user-centred task, since automatic and semantic extraction of information is still considered as a task too complex to be exclusively carried out by computers [Sme00].

The increasingly diverse role that multimedia sources are destined to play in our society and the growing need to have these sources accessed made it necessary to develop forms of multimedia information representation that go beyond the simple waveform or sample-based, frame-based (e.g. MPEG-1 and MPEG-2) or object-based (e.g. MPEG-4) representations. **MPEG-7**, formally called "*Multimedia Content Description Interface*", is a new standard for describing the content of multimedia data [ISO00,MPE99,MPE00,MPE01]. MPEG-7 is a means of attaching metadata to multimedia content. MPEG-7 specifies a standard set of description tools, which can be used to describe various types of multimedia information. These tools shall be associated with the content itself to allow efficient and effective searching for multimedia material of users' interests.

MPEG-7 is a generic standard with broader application areas than broadcast, but the usage of MPEG-7 in any application is still an open issue and there are still little experiences in practical use of MPEG-7. This paper presents our investigation in using MPEG-7 descriptions for searching multimedia data in the context of a broadcast application, and more precisely in the development of the SAMBITS broadcast terminal. SAMBITS, a European broadcast application project, provides a studio technology for off-line and real-time production of integrated broadcast and Internet services, and a terminal technology for interactive *access* to these services, which involve various media content types (MPEG-2, MPEG-4, HTML). In particular, the terminal provides end-users (TV viewers) with access to the multimedia content using MPEG-7 description tools attached to the content. This is achieved through the development of *integrated search mechanisms* in the terminal that assist average TV viewers in identifying material that is specifically of interest to them.

This paper describes the development of access methods for searching multimedia data using MPEG-7 descriptions in the context of the SAMBITS terminal. The paper concentrates on the development and implementation of a retrieval model that exploits the nature of MPEG-7 descriptions for effectively accessing the corresponding multimedia data, in our case mostly audio-visual data sequence. One important characteristic of MPEG-7 descriptions is that they display a *structure*; that is they are composed of elements describing parts of the audio-visual sequence as well as the entire audio-visual sequence. By exploiting the structural characteristic of MPEG-7 descriptions, parts of as well as the entire audio-visual sequence can be searched, thus allowing users to precisely access the data of interest to them. The retrieval model was developed and implemented using the HySpirit software, which is a flexible framework for representing data that have an underlying structure and describing retrieval functions that adapt to both the complexity of the data and the requirements of the application.

The remainder of this paper is structured as follows. In Section 2, we provide the background of the research in which this work is being carried out. We introduce the SAMBITS project, MPEG-7 and its connection to SAMBITS, and general concepts of information retrieval, in particular, those related to the retrieval of structured data. In Section 3, we describe the part of MPEG-7 used in our broadcast application for searching audio-visual data sequence. Particular attention was paid to the MPEG-7 structural aspect. In Section 4, we describe the development and implementation of our model for accessing multimedia data based on their MPEG-7 descriptions using the HySpirit software. We first examine the requirements for such a model, describe briefly the HySpirit software, and then present in details the various modelling steps. We also show how the model captures the requirements that were identified. In Section 5, we present a prototypical user interface that integrates the search functionality in the SAMBITS broadcast terminal. Section 6 summarises the paper and draws conclusions, and in Section 7, we acknowledge contributions to the work.

2. Background

This section provides first some background information about the SAMBITS project [SAM00], in which this work has been carried out (Section 2.1). Then, the section discusses the background related to the development and standardisation of MPEG-7 in general, then with respect to broadcast applications, and finally with respect to SAMBITS (Section 2.2). The section continues with a general introduction to information retrieval techniques, and their application to searching structured documents such as MPEG-7 descriptions (Section 2.3). The section finishes with a brief description of related work (Section 2.4).

2.1 The SAMBITS Project

The advent of digital TV is producing a plethora of new innovative interactive multimedia services, where broadcast and interactive (e.g. Internet) applications are starting to converge. To be accepted by the user, these new services need to be designed so that they can be easily managed, readily and comfortably accessed. New tools are required to create these integrated broadcast and Internet services in a flexible way. This is the aim of the System for Advanced Multimedia Broadcast and IT Services (SAMBITS) project, a two-year European Commission funded, IST (IST-2000-12605) project, which aims to provide a platform for integrated broadcast and Internet services [SAM00].

In particular, SAMBITS aims to provide both a Multimedia Studio Technology for off-line and real-time production of new multimedia services, and a Multimedia Terminal Technology for interactive access to these new services. The new services offer access and use of a multitude of media content types (MPEG-2, MPEG-4 Audio-Video, MPEG-4 scenes, HTML) supplemented by MPEG-7 descriptions and employ new technologies and standards such as MPEG-2, MPEG-4, MPEG-7, DVB (Digital Video Broadcasting[1]), and MHP (Multimedia Home Platform[2]). The last two standards are broadcast standards developed in Europe. DVB, which is also used in South-America, Australia, Africa and Asia, is a set of standards covering all aspects of digital television, from transmission through interfacing, conditional access and interactivity for digital video and audio data. MHP provides a set of technologies that are necessary to implement cross-platform digital interactive multimedia services in the home based on DVB. MHP

[1] http://www.dvb.org

[2] http://www.mhp.org

defines a reference model and an open API that supports applications that are locally stored as well as those that are downloaded in either real time or non-real time, allows to preserve the "look and feel" of the application, and enables access to databases (e.g. DVB-SI).

The Studio System involves the development of various authoring and visualisation tools for creating the integrated broadcast and Internet content together with its MPEG-7 descriptions that will be sent to the Terminal System. The Terminal System provides users (TV viewers) with access to high quality digital video, as provided by DVB, and to the vast world-wide collection of interactive services and databases on the Internet at the same time.

A user requirements analysis was carried out to explore, from a user point of view, the perceived benefits and problems associated with broadcast and Internet integration [HML+01]. Among the findings was the need to develop *integrated search mechanisms* (at the terminal) that assist TV viewers in identifying additional material that is specifically of interest to them. For example, imagine a user watching the Ang Lee film "Crouching Tiger, Hidden Dragon". That user may decide to seek further information regarding that film (e.g. other films directed by Ang Lee), a particular event in the film (e.g. an additional clip provided by the broadcasters showing a longer version of the final sword fight in the trees), related web data provided by the broadcasters (e.g. an HTML document containing the biography of the actress "Michelle Yeoh") or other related data available on the Internet (e.g. price of the Motion Picture Soundtrack CD of the film).

The exploitation of MPEG-7 was proposed as a way to achieve the integrated search functionality in SAMBITS. In addition to using MPEG-7 data as a means to represent the material for performing searches, the MPEG-7 descriptions associated with the broadcast programmes can be exploited for building queries. Specifically, the description of the programme currently being broadcast on the terminal (derived from the associated MPEG-7 descriptions) can be used as the basis for a search (i.e. building a query) for additional material related to that programme. There are two challenging aspects in achieving such an integrated functionality. One is the use of the MPEG-7 descriptions to represent the query internally for the retrieval system. The other, tightly related with the first, is the design of a user interface that will allow this seamless, direct formulation of queries, taking into account that in the SAMBITS scenario only remote control will be used for interacting with the terminal. These aspects are addressed in Sections 3 and 5, respectively.

2.2 MPEG-7

MPEG-7 has been developed by the Moving Pictures Expert Group (MPEG), a working group of ISO/IEC [ISO00]. The goal of the MPEG-7 standard is to provide a rich set of standardised tools to describe multimedia (i.e. audio-visual) content. Unlike the preceding MPEG standards (MPEG-1, MPEG-2, MPEG-4), which have mainly addressed coded representation of audio-visual content, MPEG-7 focuses on representing information about the content at different levels. The structural level (e.g. "this video consists of a sequence of segments and each segment is composed of several shots") is supported in the same way as the (visual) feature level (e.g. "this object has the form of a fish") or the semantic level (e.g. "Mr. Blair meets Mr. Schroeder in Nizza"). The content itself is out of the scope of the standard and MPEG-7 states explicitly that the description tools are applicable for all kinds of multimedia content independent of its format and coding. The methods and technologies generating and using the descriptions are not part of the standard and the tools are not restricted to a specific set or class of applications. To reach this goal MPEG-7 restricts itself to few, but powerful concepts. These are:

- A set of *descriptors* (Ds), for representing features of audio-visual material, (e.g. colour histogram).
- A set of *description schemes* (DSs), which define the structure and the semantics of the relationships between its elements, which include Ds and DSs. An example is the hierarchical structure of a video.
- A *description definition language* (DDL) to specify Ds and DSs.
- System tools to support efficient binary encoding, multiplexing, synchronisation and transmission of the descriptions.

The DDL allows the representation of complex hierarchies as well as the definition of flexible relationships between elements [ISO00b]. The DSs and Ds are platform-independent and must be validated. An existing language that fulfils most of these requirements is XML Schema [XML00], which is used by MPEG-7 as the basis for its DDL.

The lower level of the DDL includes basic elements that deal with basic datatypes, mathematical structures, linking and media localisation tools as well as basic DSs, that are found as elementary components of more complex DSs. Based on this lower level, content description and management elements can be defined. These elements describe the content from five viewpoints [ISO00e]:

- *Creation & Production* (describing the creation and production of the content),
- *Media* (description of the storage media),
- *Usage* (meta information related to the usage of the content),
- *Structural aspects* (description of the audio-visual content from the viewpoint of its structure),
- *Conceptual aspects* (description of the audio-visual content from the viewpoint of its conceptual notions).

The first three elements address primarily information related to the management of the content (*content management*) whereas the last two elements are mainly devoted to the description of perceivable information (*content description*). For instance, a segment can be decomposed into an arbitrary number of segments ("SegmentDecomposition"), which can be scenes or shots with an arbitrary number of temporal, spatial or content related relations to other segments. These segments can be described by additional elements. For instance, the "TextAnnotation" provides an unstructured or structured description of the content of the segment. An example of the structural content description scheme is given in Figure 1.

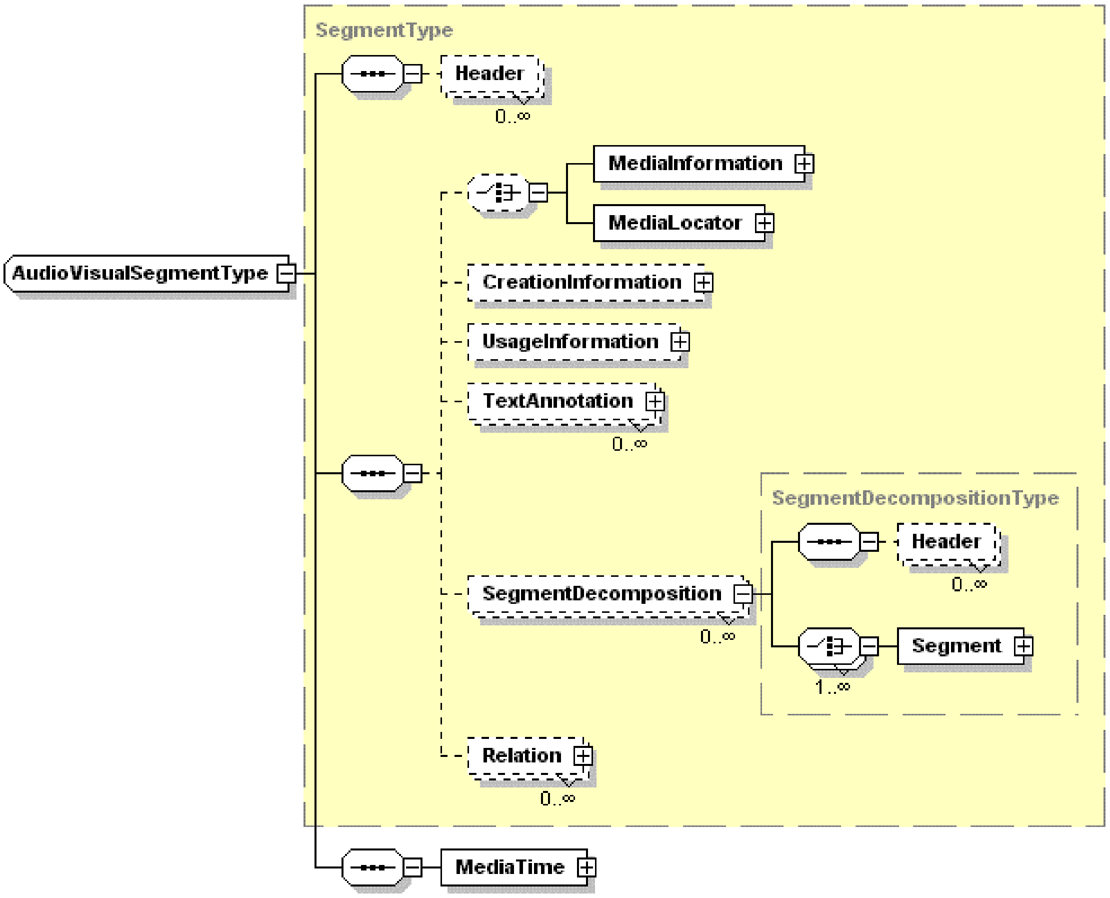


Figure 1: Audio-visual Segment Type of MPEG-7

Beside the management and description of the content, tools are also defined for *Navigation and Access*. For instance, browsing is supported by a "Summary" element and information about possible variations of the content is also given. Variations of the audio-visual content can replace the original, if necessary, to adapt different multimedia presentations to the capabilities of the client terminals, network conditions or user preferences [ISO00e].

As mentioned above, MPEG-7 is not tailored to a set or class of applications; it has to be aware of the increasing spectrum of audio-visual applications and to ensure that its tools are applicable to them. Many application domains will benefit from the MPEG-7 standard. A non-exhaustive list of application domains includes education, archives and directory services, information systems, entertainment, medicine and biomedicine, architecture and surveillance.

Of increasing importance are broadcast applications. Traditional broadcasting, including simple services like videotext, are slowly converging towards interactive television, for instance allowing the integration of the Internet. An intermediary step of this convergence is the use of Electronic Programme Guides (EPG). These services aim at providing more appropriate customisation and enhanced interactivity of a broadcast terminal, but they cannot be offered without descriptive information, i.e. metadata (e.g. [FA01, HI98,SC00]).

SAMBITS is one concrete example for this development in broadcast applications. Designing descriptive information for an application is always done with respect to the intended functionality. For SAMBITS, this can be best characterised at the user terminal by the following list [Put01]:

- Selection of parts of a broadcast programme for viewing additional content.
- Instant access to additional asynchronous material (both via DVB and the Internet).
- Easy and rapid access to programme-specific information.
- Instant access to additional synchronised material that is tightly linked to the programme such as additional audio or video signals.
- Navigation by user profiles or programme-related metadata allowing user-, object- or programme-specific searching.

To fulfil these requirements two different metadata sets were defined for the SAMBITS terminal: the content-related metadata for the different types of information supported by SAMBITS (MPEG-2 streams, MPEG-4 objects and scenes, and HTML information) and the programme-related metadata. The content-related metadata set allows a structural and an object-oriented view of the material to be broadcasted as TV programme elements. Each programme element is described by an element identifier, media characteristics and time information, low-level audio and/or visual features and high-level textual descriptions. The latter ones describe the programme element in terms of animate or inanimate objects, actions, places, time, and purpose.

As described in Section 3, MPEG-7 was used as a toolbox, from which elements were selected when they fitted the requirements. This selection of the appropriate subset of MPEG-7 is also suggested by MPEG-7 and, for SAMBITS, MPEG-7 turned out to provide the necessary tools.

2.3 Information Retrieval

Information retrieval (IR) is the science concerned with the efficient and effective storage of information for the later retrieval and use by interested parties. The general (and simplified) IR process is shown in Figure 2 [Rij79].

Information is stored in documents, the set of which constitutes a collection. For efficiency, an IR system generates a representation of this content (i.e. the indexing process). For a text collection, this corresponds to terms (or keywords, indexing terms) extracted from the text[3]. The indexing terms are usually weighted to reflect their importance in describing the document content and to capture the inherent uncertainty associated with the indexing process. More complex representations can be used, for example, noun-phrases and logical formulae (propositions and sentences of predicate logic or higher order logic).

A user with an information need submits a query to the IR system (i.e. the formulation process). The query is usually translated into an internal representation (often similar to that of the documents), and the IR system will compare the query representation to that of the documents forming its collection, and estimate the extent to which those documents satisfy or are relevant to the query. The comparison process is based on retrieval functions defined by so-called IR models [BR99]. For instance, in the Vector Space Model, the retrieval function corresponds to a similarity measure applied to the two representations, which consist of multi-dimensional vectors; in the Probabilistic Model, it corresponds to the probability that the document is relevant to the query based on distributions of terms in relevant and non-relevant documents. The comparison process returns a ranked list of documents, which is then displayed to the user, as an answer to his/her query.

In traditional IR systems, retrievable units are fixed. For example, the whole document, or, sometimes, pre-defined parts such as paragraphs constitute the retrievable units. The structure of documents (for instance, its logical structure, e.g. sections and sub-sections) is “flattened” and not exploited. Nowadays, with XML [XML00d] becoming widely used as a document language, the structure of documents is gaining increasing attention. As discussed in the previous

[3] Usually, these are the terms remaining after removal of insignificant (or non-content bearing) words (“the”, “over” in the English language) and stemming (e.g. “sailing” and “sailors” are stemmed into a root form such as “sail”).

section, MPEG-7 descriptions are XML documents, therefore they display a structure; they are composed of elements. With MPEG-7 descriptions, the retrievable units should be the elements as well as the whole audio-visual sequence. That is, the retrieval process should return elements of *various levels of granularity*, for example, an element when only that element is relevant, a group of elements, when all the elements in the group are relevant, or the full video itself, when the entire audio-visual sequence is relevant.

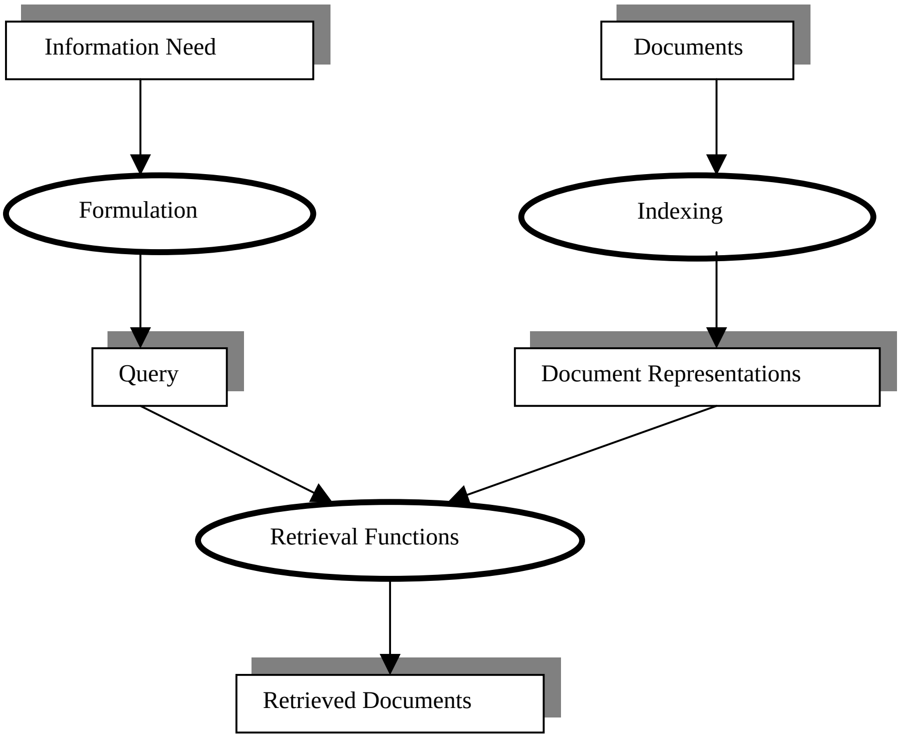


Figure 2: The (Simplified) Information Retrieval Process

For retrieving structured documents, the indexing process or the retrieval function has to pay attention to the structure of a document. Several approaches to structured document retrieval have been developed. The so-called passage retrieval (e.g. [Cal94,SAB93,Wil94]) determines retrieval weights for passages (paragraphs, sentences) to compute a final retrieval weight for a set of passages. Other approaches aim at computing an independent retrieval weight for each document component, so to find the *best entry points* into a structured document (e.g. [CMF96,Fri88,LM00,LR98,MJK98,Roe99]). As discussed in Section 4.1, an MPEG-7 retrieval model should support the notion of best entry points to return elements at various levels of granularity. The retrieval model developed in this paper therefore follows the second set of approaches.

2.4 Related Work

There are a number of MPEG-7-related projects that involve the adoption of MPEG-7 conformance [Hun99]. The HARMONY project[4] aims at exploiting (upcoming) standards such as RDF, XML, Dublin Core and MPEG-7 to develop a framework allowing diverse communities to define descriptive vocabularies for annotating multimedia content. A second project is DICEMAN[5] with a broad objective in developing an end-to-end chain for indexing, storage, search and trading of audio-visual content, where MPEG-7 is used to index this content. A third project is AVIR, which aims at developing an end-to-end system for delivering of personalised TV services over a DVB channel. AVIR has demonstrated (IBC 2000) how MPEG-7 metadata, delivered along with the content and used

[4] See http://www.ilrt.bris.ac.uk/discovery/harmony/.

[5] See http://www.teltec.dcu.ie/diceman/.

within a personal Electronic Programme Guide (EPG), could serve the non-IT expert user for automatic recording, later viewing, browsing and searching of broadcast video material. The CMIP[6] developed an “Article-Based News Browser”, a system that segments news, generates keyframes, and extracts additional information based on texts and icons. The generated top-level description structures are similar to those of SAMBITS, but the description of episodes in deeper hierarchies is not supported. The system supports a browsing functionality, but not retrieval. Another project is COALA[7], which aims to design and implement a digital audio-visual library system for TV broadcasters and video archive owners with facilities to provide effective content-oriented access to internal and external end-users. The application domain is news and one of its goals is the application of MPEG-7 to news content [FA01].

MPEG-7 is taking into account the activities of other international standardisation bodies that are also concerned with the management of content such as [Hun99]: MARC and Z39.50 for libraries [BR99]; SMTPE (Society of Motion Picture and Television Engineers)[8], Dublin Core[9], SMEF (the BBC Standard Exchange Framework) for electronic archives. MPEG-7 is combining effort with those standardisation bodies to maximise interoperability, prevent duplication of work and take advantage of work already carried out through the use of shared languages (e.g. XML Schema Language [XML00d]).

Another line of work related to ours is the development of interfaces for browsing audio-visual collections (e.g. [GBC+99,LST+00]). This is important for two reasons. First, there are technical constraints in transmitting and displaying high-quality video, which must be taken into account in building search tools for audio-visual data. Second, the effective interaction between end-users and video and speech retrieval systems is still an open research issue. It is important to devise retrieval systems with which user can effectively and efficiently carry their information-seeking activities. This is also a concern of SAMBITS. Although MPEG-7 provides descriptive tools (i.e. *Navigation and Access*) for customising interfaces (e.g. for browsing purpose), there is yet little experience in effectively applying these tools in real environment.

A break-through in metadata approaches such as MPEG-7 would be reached with the automatic generation of such metadata. This starts with the low-level feature extraction from audio-visual sequences, where the main concerns are the automatic segmentation of video sequences into shots using image processing algorithms, and information extraction using speech analysis and optical character recognition techniques, and mapping these low-level features to high-level concepts such as “sky”, “sea”, etc (e.g. [ABB01,Hau95,Sme00,ZTS+95]). User interventions have also been proposed as a means to map low-level to high-level features, for example in the two systems AMOS and IMKA developed at the University of Columbia [BZC+01].

Finally, more information about MPEG-7 applications, software, and approaches can be found in the MPEG-7 Awareness Event site, at http://www.mpeg-industry.com/.

3. SAMBITS MPEG-7 Description Schemes

In this section, we present an overview of the collection of MPEG-7 DSs that were experimented within SAMBITS and allowed the implementation of a typical enriched broadcast application.

As mentioned in Section 2.2, the MPEG-7 standard can be seen as a toolbox of generic description structures with associated semantics to be instantiated by any multimedia application making use of metadata. While having such schemes standardised is useful to maximise interoperability, the scopes and possible application domains are so broad that it is unrealistic to provide an exhaustive set. Consequently, MPEG-7 has been made extensible by means of its DDL so that application developers can extend the standard DSs to meet their specific requirements. Similarly, it is very unlikely that a single application will need the whole set of powerful and sometimes complex tools that MPEG-7 proposes. In that sense, a practical way to use the standard is to perform a selection of the description structures and validate those against specific targeted functional requirements. This application-driven process might lead in the future to the definition of “profiles”, a path for which technical means are currently being investigated in MPEG.

SAMBITS has hence worked along this line and has built a subset of relevant DSs from both the broadcast studio and terminal perspective. An overview of the terminal-related choices (the detailed list being out of the scope of this paper) is given hereunder. Particular attention is paid to the “Structural Aspects” of MPEG-7, which is of great

[6] See http://www.lg-elite.com/MIGR/cmip/
[7] See http://lithpc17.epfl.ch/
[8] See TV-AnyTime at http://www.davic.org.
[9] See Dublin Core Metadata Element Set at http://purl.org/DC/documents/rec-dces-19990702.htm.

importance when typical broadcast programmes are being enriched with metadata, as soon as a precise and granular description is targeted or when browsing and navigation within the programme is envisaged. Low-level features (DSs and Ds based on signals such as colour, shape, frequency) have not been considered by SAMBITS to be used in the terminal since they do not correspond to an abstraction level to which the non-expert average TV watcher is familiar with. It is unlikely that a non-expert average TV viewer will be seeking video shots that, for instance, contain a high distribution of the colour red and have a high noise level. The same does not hold at the studio side, where broadcast experts may have such queries, for instance, during a programme editing phase.

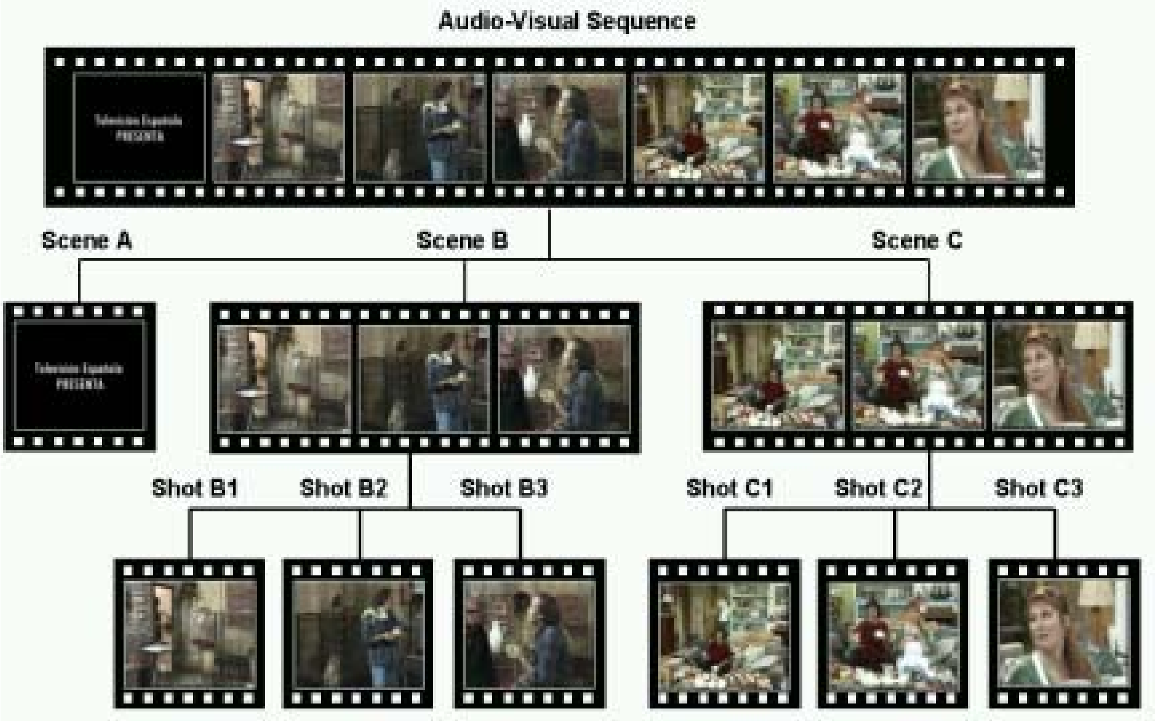


Figure 3: A Table of Content example

The description of the structural aspects of the content is extensively dealt with in MPEG-7. SAMBITS provides and uses in the terminal a hierarchically structured content description, referred to in the following as a *Table of Content*, a structure being very similar to its well-known equivalent in textual documents. Indeed, similarly to a book being decomposed into chapters, sections, and paragraphs, an audio-visual sequence can in most cases be structured into temporal nested elements such as scenes/shots in a movie or sets/games/points in a tennis match. In the general case, a Table of Content of an audio-visual sequence can be defined as a tree for which each node corresponds to a time interval that is decomposed into a partition of successive and temporally contiguous sub-intervals. An example of such a hierarchical structure is given Figure 3.

The so-defined audio-visual Table of Content can be automatically (or semi-automatically) generated at the production site by dedicated algorithms involving a temporal segmentation stage (usually into shots) and a tree creation process using low-level content features (sound, colour, motion) and possibly a priori knowledge and predefined models of the content structure [LS99,SLG00]. This structure provides a “skeleton” allowing the gathering of descriptive information in a hierarchical way, which is inherently scalable in the sense that any level of temporal granularity can be achieved for the description. Within SAMBITS, the Table of Content (together with the associated embedded DSs) is essential to:

- Provide more information / build queries at a given time with a controllable temporal precision.
- Retrieve not only entire programmes but also “pieces” of content that are most relevant given a user query.
- Browse through the query results or the recorded content in an efficient and intuitive way.

In the following two subsections, we will see how MPEG-7 description tools meet the requirements induced by the above approach through the definition of nested abstract “Segments” as dedicated placeholders for the description of a hierarchically structured content.

3.1 Segment Entities

Within MPEG-7, *Segment Entities* specify tools for describing spatial, temporal, or spatio-temporal segments of the audio-visual content and hierarchical structural decompositions of those segments. Figure 4 shows the tools that have been selected to be used in SAMBITS for representing segment entities and in particular the above mentioned Table of Content of a broadcast programme. The “Segment” DS provides the abstract base definition from which specialised segment entity classes are derived. Specialised tools include, among others, the “StillRegion” DS, used to describe regions of images and frames of video data, and the “AudioVisualSegment” DS used to describe spatio-temporal segments of the audio-visual content (in our case the nodes of the Table of Content). Finally, the “SegmentDecomposition” DS describes the hierarchical decomposition of a segment into sub-segments.

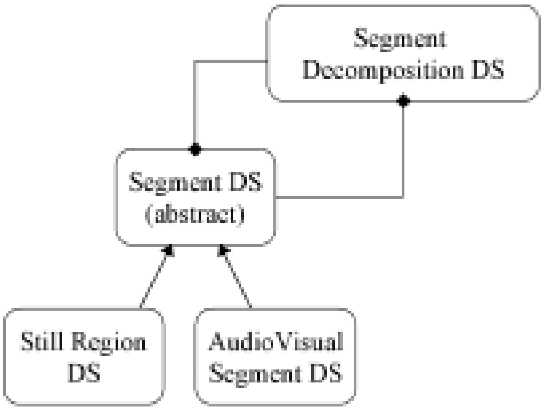


Figure 4: Overview of the MPEG-7 tools describing Segments in SAMBITS

Note that since the “Segment” DS is abstract, its actual type must be specified in the description instantiation, using the polymorphism feature of XML Scheme and the *xsi:type="ActualType"* construct [XML00a]. An example is shown in Figure 5, which applies to the Table of Content shown in Figure 3. Moreover, the “MediaLocator” descriptor is also needed to specify the location of a particular image, audio or video segment by referencing the media data using an URI (Universal Resource Identifier) [XML00e] and possibly a “MediaTime” reference. “(?)” refers to some instantiations, which are not shown in the example.

```
<AudioVisual id="Audio-Visual Sequence">
      <MediaLocator>
            <MediaURI> (?) </MediaURI>
            <MediaTime>(?) </MediaTime>
</MediaLocator>
      <SegmentDecomposition DecompositionType="temporal">
            <Segment xsi:type="AudioVisualSegmentType" id="Scene A"> (?) </Segment>
            <Segment xsi:type="AudioVisualSegmentType" id="Scene B"> (?) </Segment>
            <Segment xsi:type="AudioVisualSegmentType" id="Scene C"> (?) </Segment>
      </SegmentDecomposition>
</AudioVisual>
```

Figure 5: Example of “Segment” DSs

Apart from the description of temporal segments and still images, the “Segment” DS (“StillRegion” or “MovingRegion”) and associated tools can be used to implement object-based hypermedia applications (generally referred to as Video Hyperlinking) in an enhanced multimedia broadcast terminal. MPEG-7 has indeed been proven very efficient to achieve those functionalities, in particular compared to object-based video coding schemes such as MPEG-4 [HK99]. Such object-triggers are theoretically in the strict scope of SAMBITS applications in which the user should typically be able to ask for more information and build queries related to a specific object of the visual scene (e.g. “Who is this particular actor?”). However, since the technical breakthrough relies more on the spatial segmentation process to be performed at the studio side rather than in the description itself, this will not be discussed further in this paper.

3.2 Descriptions Associated to Segments

In addition to the already mentioned and mandatory location of the media, the MPEG-7 abstract “Segment” DS specifies a set of optional DSs that can be filled to provide a high-level representation of the content at a given granularity. The set includes *CreationInformation*, *UsageInformation*, *TextAnnotation*, *MatchingHint*, *PointOfView*, and *Relation* ([ISO00e]). We intend here to highlight how SAMBITS uses some of them to fulfil the typical functional requirements such as providing more information at a specific request time, building queries in a seamless way and allowing navigational capabilities in the terminal. The overall induced representation of the programme structure is illustrated in Figure 7. Next, we describe briefly the components associated to segments that were instantiated and used in SAMBITS.

Key-Frames

To browse through the content structure represented by the Table of Content (e.g. for later non-linear viewing of a recorded programme or for presenting the results of a query), each segment needs to be represented by one (or possibly several) *key-frame*, which is one representative image of the given video segment. The association of a segment with its key-frame is done in SAMBITS using the *Relation* element of the “Segment” DS. Relations in MPEG-7, specifying how a *source* segment relates to a *target* segment, can be of many types including (together with their inverse equivalent):

- spatial relations (e.g. “left”, “inside”),
- temporal relations (e.g. “before”),
- neither spatial nor temporal relations (e.g. “annotatedBy” or “hasKeyOf”)

The latter one is used for the representation of key-frames. The source segment is therefore an “AudioVisualSegment” corresponding to a given node of the Table of Content, while the target segment is a “StillRegion”. A corresponding MPEG-7 description instance fragment is illustrated in Figure 6.

```
<Segment id="KFSceneA" xsi:type="StillRegionType">
    <MediaLocator> (?) </MediaLocator>
</Segment>
(?)
<Segment xsi:type="AudioVisualSegmentType" id="Scene A">
    <MediaLocator> (?) </MediaLocator>
    <Relation name="hasKeyOf"
          target="KFSceneA" xsi:type="BinaryOtherSegmentRelationType"/>
</Segment>
```

Figure 6: Example of an instance of a Key-Frame

Annotations

Besides the key-frame representation, the “TextAnnotation” DS is strongly required since it provides a description of the content directly usable at the application level for displaying information to the user for building (or answering to) queries. As far as annotations are concerned, MPEG-7 supports both free textual content and structured annotations in terms of action (“What”), people and animals (“Who”), objects (“WhatObject”), action (“WhatAction”), places (“Where”), time (“When”), purposes (“Why”) and manner (“How”).

Related Material

The last DS used in the SAMBITS terminal at the segment level (although the list of schemes given here is not intended to be neither exhaustive nor restrictive) is the “RelatedMaterial” DS, a description element within the “CreationInformation” DS. It allows to point from a given segment to any other audio-visual material (e.g. in another broadcast channel, locally archived or available on the internet) that could be of the viewer’s interest, providing typically additional information on a particular topic being dealt with in the current programme.

As a summary, Figure 7 depicts how the structural aspects, captured mainly by the Table of Content, are described in the broadcast terminal according to the implementation of the MPEG-7 standard within SAMBITS. One can see that the root segment corresponds to the whole audio-visual item, and can be treated as any other segment by gathering global information that is relevant for the entire duration of the content. The root segment is further decomposed into key-frames (instantiations of the “StillRegion”) and the Table of Content (consisting of nested instantiations of the “AudioVisualSegment”).

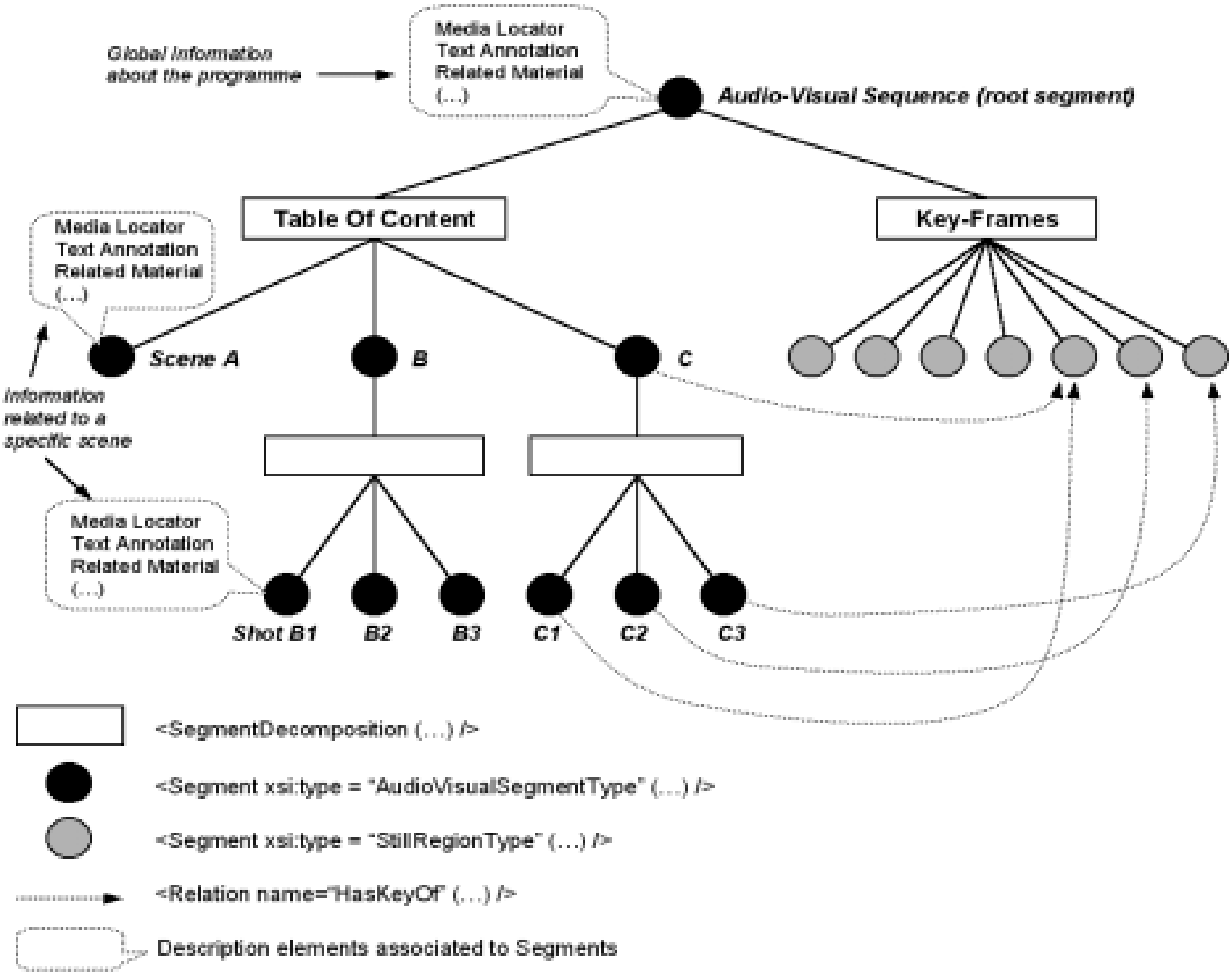

Global Information about the programme
Media Locator
Text Annotation
Related Material
(...)
Audio-Visual Sequence (root segment)
Table Of Content
Key-Frames
Media Locator
Text Annotation
Related Material
(...)
Scene A
B
C
Information related to a specific scene
Media Locator
Text Annotation
Related Material
(...)
Shot B1
B2
B3
C1
C2
C3
<SegmentDecomposition (...) />
<Segment xsi:type = "AudioVisualSegmentType" (...) />
<Segment xsi:type = "StillRegionType" (...) />
<Relation name="HasKeyOf" (...) />
Description elements associated to Segments

Figure 7: Structural aspects of MPEG-7 for the SAMBITS terminal

4. The Retrieval Model

The previous section described the MPEG-7 description tools necessary to implement a search functionality in a broadcast terminal. The present section describes the retrieval model for searching audio-visual material based on their associated MPEG-7 descriptions. First, we examine the requirements of a retrieval model for MPEG-7 (Section 4.1). Second, we present HySpirit, the software development kit that we used to develop and implement the retrieval model (Section 4.2). Third, we describe in details the actual design and implementation of the retrieval model using HySpirit (Section 4.3).

4.1 Requirements of a Retrieval Model for MPEG-7

MPEG-7 DSs define the schemes for representing structure, content and relationships of multimedia data; MPEG-7 DSs are specified as XML schemes. An MPEG-7 description is an instance of a DS, so we can consider an MPEG-7 description as an XML document. An XML document is a structured document, in the sense that the XML language is one way to capture the structure of a document. With this view in mind, the requirements for a model for structured document retrieval, and in particular, XML document retrieval, apply to MPEG-7 retrieval.

The first requirement applies to any retrieval model. We need a "*relevance-based" ranking function*, so that weights (e.g. probability values) are assigned to elements (e.g. segments) forming a retrieval result, reflecting the extent to which the information contained in an element is relevant to the query. This is particularly important when searching large repositories of multimedia data so that best matches (i.e. the most relevant elements) are displayed *first* to users[10].

A crucial requirement for structured document retrieval is that the *most specific element(s)* of a document should be retrieved (i.e. returned to the user). For MPEG-7 retrieval, this means that not only an entire video but also video parts can constitute a retrieval result, depending on how they match the query. Therefore, a retrieval model for MPEG-7 must determine *entry points* into the MPEG-7 structure. For example, suppose that a news broadcast (e.g. "AudioVisual" DSs) is structured into several news clips (e.g. "Segment" DSs). For a generic query, the entire news broadcast (i.e. the "AudioVisual" segment) would be an appropriate retrieval result, whereas for a specific query a particular news clip (one "Segment") would constitute a better retrieval result.

A third requirement relates to the *relationships between elements*, such as spatial and temporal relationships (see Section 3.2). In classical retrieval, we deal with independent documents and simple propositions in documents (e.g. terms occurring in a document). With structured documents, the relationships between the elements (here MPEG-7 DSs and Ds) must be considered, in particular for determining the most relevant document elements. Besides spatial and temporal relationships, relationships such as links (e.g. pointing to additional material such as an HTML page), order (sequence) and others should also be captured.

With respect to XML documents, the use of *attributes* leads to a further requirement. Standard IR, which has one of its aims the representation of the content of documents, will treat the attributes of an XML document as its content, and hence it will not explicitly model attributes. Attributes are used in databases to characterise entities (e.g. Name, Address, and Age of Person entity). In standard database approaches, content is often considered as an attribute, and again, there is no conceptual support that distinguishes between content and attributes (e.g. [ACC+97]). For accessing XML documents, and hence MPEG-7 descriptions, more refined retrieval methods are necessary to distinguish between attributes and content of XML documents.

The next requirement arises from one of the goals of MPEG-7, which is to describe multimedia data in terms of the *objects (*persons, places, actions, etc.) that occur in them. As pointed out in Section 3.2, the representation of objects in a scene is part of an MPEG-7 instance (see descriptors "Who", "Where", "WhatAction", etc.). Those objects add a new dimension to a retrieval model, when we can distinguish between content-bearing entities (retrievable document units) such as videos and video segments, and "semantic" entities such as persons, actions, etc.

The next requirement refers to the *data propagation* of MPEG-7 descriptions. That is, some attributes and elements (e.g. descriptors) defined at upper levels may be valid for elements at lower levels. For example, a "FreeTextAnnotation" specified for a video description root is the description of all contained video elements, if not specified at the video segment level. A retrieval model for MPEG-7 should be able to specify which elements, if any, are propagated up or down an MPEG-7 structure.

Specific to the SAMBITS terminal is the requirement for the *integrated search* (see Section 2.1), allowing TV viewers to submit a search request based on the description of what they are watching at a given time. This requires that the retrieval system uses parts or all of the MPEG-7 descriptions attached to the scene on display to determine the internal representation of the query.

Various users with varying background, experience and interest will use the SAMBITS terminal. With user profiles, we can adapt the terminal search functionality to *user preferences*. A last requirement of a model for MPEG-7 retrieval is the conceptual integration of user profiles into the retrieval model, so that an element is not only retrieved with respect to its information content, but also according to user preferences.

In summary, the requirements for a retrieval model for accessing multimedia data based on MPEG-7 descriptions are:

- Provide a "relevance-based" ranking function.
- Retrieve the most relevant document elements (an entire audio-visual sequence vs. a segment).

[10] Statistics show that users look on average at the first two screens of a web search result; this is equivalent to looking at the twenty top-ranked hits [BR99].

- Consider relationships (spatial, temporal, etc) for representing (indexing) and retrieving MPEG-7 elements.
- Support query formulation with respect to content and attributes.
- Model content-bearing entities (segments) as well as “semantic” entities (persons, places, etc).
- Reflect the data propagation of MPEG-7 attributes and elements.
- Support an integrated search, i.e. an MPEG-7 description can constitute a query.
- Consider the user preferences in the retrieval model.

Requirement 1 holds for all IR models, requirements 2 and 3 hold for structured document retrieval, requirement 4 is particular to XML retrieval, requirements 5 and 6 are particular to MPEG-7, and requirement 7 is particular to the SAMBITS terminal. Requirement 8 is common to IR models where user preferences are encapsulated in the query. Requirement 4 is provided by IR approaches that are based on the integration of IR and database models and technologies [FR98]. Requirement 8 is well known to content-oriented push scenarios where queries are static entities (and correspond to user profiles) and documents are incoming entities that are matched to the established queries.

Requirements 1 to 6 are addressed in the retrieval model, in particular in Section 4.3.1. Requirement 7 is addressed in Section 5, which describes a prototypical design of the user interface for the SAMBITS terminal. For effectively taking into account user preferences (requirement 8), longitudinal user studies are first necessary. We will however discuss briefly in Section 4.3.2 an example that incorporates content-based user preferences into the retrieval model.

4.2. HySpirit

HySpirit is a software development kit[11] [Roe99], which provides support for representing complex documents and describing retrieval functions. HySpirit is based on generic and well-established approaches to data and information management such as relational database model, logic, and object-orientation. HySpirit therefore provides the necessary expressiveness and flexibility for capturing content, facts, and structure in retrieving information from large data sources. Users (administrators, developers) have extensive control on many aspects of the retrieval process. This allows the development of novel, experimental and highly flexible retrieval systems, which can be dedicated to specified requirements and tasks.

HySpirit represents knowledge (content, fact, and structure) by a *probabilistic object-oriented four-valued predicate logic* (POOL). The object-oriented nature of POOL was motivated by F-Logic [KLW95], which combines object-oriented principles (e.g. classification and attribute values) with logical rules (e.g. Datalog rules). The semantics of POOL is based on the semantic structure of modal logics [HM92]. This allows for a context-dependent interpretation of knowledge representation, which is necessary for modelling the nested structure of MPEG7 descriptions. The uncertainty of knowledge is modelled with a probabilistic-extended semantics [FH94]. The retrieval functions are implemented as *inference* process based on the logical approach to IR (see [Rij86]), which computes the probability $P(d \rightarrow q)$ that a document d implies a query q. For implementation issues and the integration of relational database management system, POOL is mapped onto a *probabilistic relational algebra* (PRA) [FR97].

As it will be demonstrated in the next section, HySpirit provides all concepts and paradigms necessary to design and implement a retrieval model for MPEG-7 that satisfies the requirements listed in Section 4.1. In particular, HySpirit supports the design and implementation of retrieval models for data with an underlying structure, a key characteristic of MPEG-7 documents, thus allowing the retrieval of elements at various level of granularity. Finally, unlike XQL [RLS98] and OQL [ACC+97], HySpirit provides means to represent the uncertainty inherent to the IR process, which leads to a relevance-based ranking of the retrieval results.

4.3. Design and Implementation of the Retrieval Model with HySpirit

This section describes the procedure followed to develop and implement the retrieval model using HySpirit. POOL is first used to provide a probabilistic representation of MPEG-7 data and MPEG-7 based queries (Section 4.3.1). The POOL representation is then mapped to a PRA representation (Section 4.3.2). The PRA representation is then interpreted by an inference engine (retrieval functions) that produces a ranked list of results (Section 4.3.3).

[11] www.hyspirit.de.

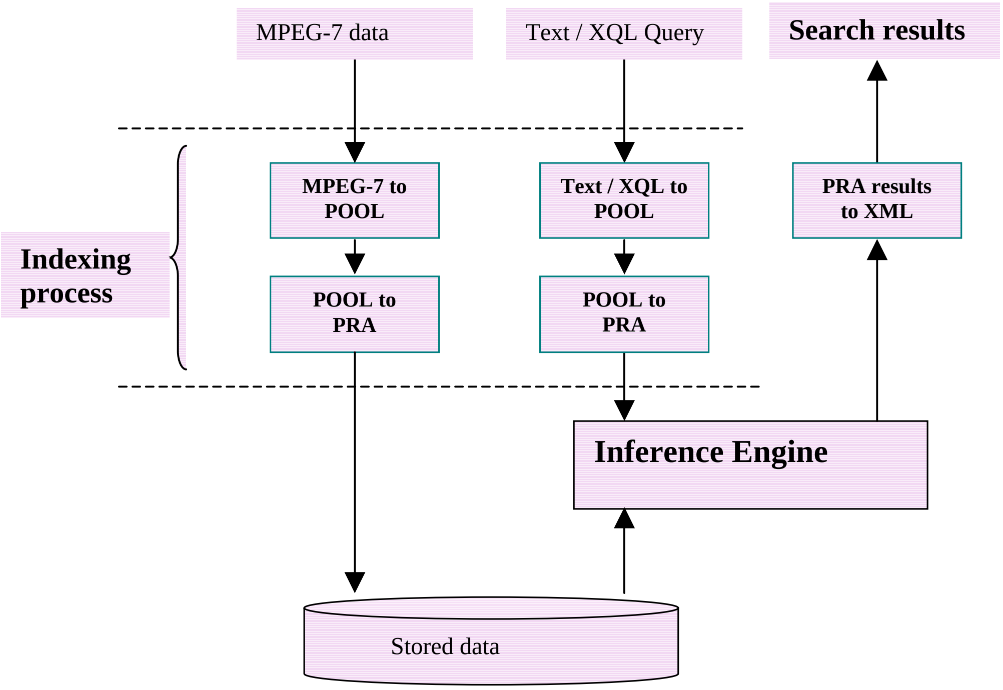


Figure 8: The architecture of the system

The overall architecture of the system that implements the model is shown in Figure 8. MPEG-7 data (DS instances) are indexed (MPEG-7 to POOL and POOL to PRA as explained above). The indexed data is then stored (e.g. locally on the terminal set-top box, or remotely on a broadcast server). Queries can be formulated as stand-alone keywords (text queries) or as structured queries formulated for example with XQL [RLS98]. The formulated queries are mapped into POOL and then PRA representations. The inference engine compares the indexed queries to the indexed MPEG-7 data, which results in a ranked list of entry points (i.e. segments at various granularity levels of the hierarchical structure) in the MPEG-7 data. The entry points, expressed within PRA, are then mapped to the original MPEG-7 data (XML elements) that will then constitute the search results.

Throughout this section, we illustrate the procedure using the extract of a sample MPEG-7 description of a soccer game[12] (shown in Figure 9). The extract consists of an audio-visual segment ("AudioVisualSegmentType"), composed of two sub-segments ("SegmentDecomposition"). Creation information is provided for the audio-visual segment, such as a "Title", an "Abstract", the "Creator", the "Genre" and "Language" (the content management part of MPEG-7). The segment has also a free text annotation. The sub-segments ("VideoSegmentType") correspond to video shots. Each sub-segment has a free text annotation component.

[12] This extract is based on the MPEG-7 structure experimented with in SAMBITS, using content from the MPEG-7 Monster Description of the MPEG soccer game video.

```
<AudioVisual xsi:type="AudioVisualSegmentType">
<CreationInformation>
  <Creation>
     <Title>Spain vs Sweden (July 1998)</Title>
     <Abstract><FreeTextAnnotation>Spain scores a goal quickly in this World Cup soccer
               game against Sweden. The scoring player is Morientes.
     </FreeTextAnnotation></Abstract>
     <Creator>BBC</Creator>
        </Creation>
        <Classification>
     <Genre type="main">Sports</Genre>
     <Language type="original">English</Language>
   </Classification>
</CreationInformation>
<TextAnnotation>
     <FreeTextAnnotation>Soccer game between Spain and Sweden.</FreeTextAnnotation>
</TextAnnotation>
<SegmentDecomposition decompositionType="temporal" id="shots" >
     <Segment xsi:type="VideoSegmentType" id="ID84">
              <MediaLocator> (?) </MediaLocator>
              <TextAnnotation><FreeTextAnnotation>Introduction.</FreeTextAnnotation>
              </TextAnnotation>
     </Segment>
     <Segment xsi:type="VideoSegmentType" id="ID88">
              <MediaLocator> (?) </MediaLocator>
              <TextAnnotation><FreeTextAnnotation>Game.</FreeTextAnnotation>
              </TextAnnotation>
     </Segment>
</SegmentDecomposition>
</AudioVisualContent>
```

Figure 9: Extract of a MPEG7 Description

4.3.1. From MPEG-7 Description to POOL Representation

POOL is a probabilistic object-oriented logic, which enables the integration of content-based and fact-based querying, as well as the structure of documents. The *knowledge of* (content) or *knowledge about* (fact) multimedia data is expressed in terms of POOL programs. These combine:

- object-oriented modelling concepts like aggregation, classification, and attributes,
- classical IR concepts such as weighted terms,
- a probabilistic aggregation of knowledge necessary for structured document retrieval.

The retrieval units are modelled as *contexts*, where a retrievable unit can be any “Segment” DS of the audio-visual sequence at any level of granularity. This includes the root segment, corresponding to the complete video sequence, or a segment corresponding to a particular scene or shot in the video (see Figure 7).

Since a video sequence is decomposed into segments (e.g. scenes), which can themselves be decomposed into (sub-)segments (e.g. shots), etc. (as represented by the “SegmentDecomposition” element modelling the hierarchical structure of the video), contexts are nested into each other. The retrieval process should therefore return the most relevant level(s) of granularity. For example, for a long video sequence with only one relevant scene, the scene (i.e. sub-segment level) should be retrieved instead of the whole video (i.e. segment root level).

Representation of MPEG-7 Descriptions

An example of a POOL program that illustrates the MPEG-7 description of Figure 9 is given in Figure 10. The POOL program consists of a set of *clauses*, where each clause is either a context or a *proposition*. The *nested* contexts represent the structure, whereas the propositions represent the content (terms) and the attributes at the respective context level.

```
audiovisualsegment(audiovisualsegment_1) % classification of object audiovisualsegment_1
audiovisualsegment_1.title("Spain vs. Sweden")
% optionally created in addition to the title_1 context
audiovisualsegment_1.creator("BBC")
audiovisualsegment_1[
  title_1["Spain vs. Sweden" 0.8 spain 0.8 sweden]
  % probabilities express uncertain contribution of terms to title content
  % title is modelled both as content and as attribute
  abstract_1[spain scores a goal  …]
  soccer game between spain and sweden

  segment_1.medialocator("(?)")
  segment_1[introduction]

  segment_2.medialocator("(?)")
  segment_2[game]
]
audiovisualsegment_1.genre("Sports")
audiovisualsegment_1.language("English")
```

Figure 10: Example of the POOL representation of the MPEG-7 Description of Figure 9

The first clause classifies the context “audiovisualsegment_1” to be an instance of the class “audiovisualsegment”. The second clause states that “Spain vs. Sweden” is the title of the context “audiovisualsegment_1”. The third clause means that “BBC” is the creator of the context “audiovisualsegment_1”. These three clauses express facts, i.e. knowledge about the context “audiovisualsegment_1” (its type, its title and its creator). The fourth clause reflects the structure of “audiovisualsegment_1”, which is composed of four *sub-contexts*: “title_1”, “abstract_1”, “segment_1” and “segment_2”. The context “audiovisualsegment_1” is called the *super-context* of these four sub-contexts. The content of “audiovisualsegment_1” is given by the terms “soccer game between spain and sweden”. The content of “segment_1” is the term “introduction”; that of “segment_2” is the term “game”.

Fact vs. Content

Contexts have a unique identifier, a content and a set of attributes. This is driven by a conceptual distinction between content-oriented and factual querying. In the context of MPEG-7, this would mean that, from the descriptors of a segment that can be exploited for searching, some are considered to represent the content (knowledge) of a segment, and others are considered to represent facts (knowledge) about these segments. For example, a query seeking videos about dinosaurs is considered a content-oriented query, whereas a query seeking videos in English produced by BBC is a factual query. In this sense, seeking a BBC documentary about dinosaurs corresponds to both a content-oriented and factual query. In our representation, we consider the “Title” descriptor to contribute to the content of a segment and to be a fact about the segment. Therefore, the “Title” descriptor is translated to both knowledge of (content) and knowledge about (fact) the context “audiovisualsegment_1” in Figure 10.

Note that not all MPEG-7 descriptors available at the terminal contribute to deriving the content of a video (for searching purposes). For example, “Media Time” or “Media URI” provides technical information such as links etc. Therefore, the retrieval is not based upon such descriptors. Table 1 summarises the MPEG-7 descriptors of a segment upon which retrieval is based and the distinction we have made in terms of fact or content contribution.

The “Text Annotation” DS and its elements are considered to describe the content of a segment, whereas elements of the “Creation” DS and Classification DS are considered to describe facts about the content, with the exception of the “Title” and the “Abstract” descriptors. “Abstract” was thought to express the content, whereas “Title” was considered to express both some content as well as the fact of a segment being entitled as such. Therefore, when modelling the MPEG-7 data in POOL, we classify the above descriptors as content or as attributes of the segment context, respectively. We also include the “Media Time” and the “Media Locator” elements to be attributes of the segment context, since they provide information necessary for locating and presenting the retrieved segments.

Segment elements expressing Content	**Segment elements expressing Facts**
TextAnnotation/FreeTextAnnotation	CreationInformation/Classification/Genre
/StructuredAnnotation/Who	/Classification/Language
/StructuredAnnotation/Where	/Classification/Country
/StructuredAnnotation/What	
	CreationInformation/Creation/Creator
CreationInformation/Creation/Title	/Creation/Title
/Creation/Abstract	

Table 1: MPEG-7 DSs and Ds used for retrieval

Querying MPEG-7 Descriptions

An information need such as “I am looking for every instance of a goal” can be described as querying for all contexts (segments) where the logical formula (the proposition) “goal” is true. For example, the query

```
?- D[goal]
```

searches for all contexts D where the formula “goal” is true (the result here consists of the two contexts “audiovisualsegment_1” and “segment_1”; this is explained later in this section). An example of a combined content and factual query is

```
?- D[goal] & D.title(“Spain vs. Sweden”)
```

This query combines a typical IR criterion referring to the content (all contexts about “goal”) with a typical database selection referring to the attribute values (all contexts with title ‘Spain vs. Sweden”). The result consists of one context, “audiovisualsegment_1”. The query corresponds to a conjunction (AND combination). A disjunction (OR combination) query is expressed via *rules*. For instance, the following query (which is also a combined content and factual query)

```
retrieve(D) :- D[goal]
retrieve(D) :- D.title(“Spain vs. Sweden”)
```

searches for all contexts (segments) about (showing a) “goal” *or* with title “Spain vs. Sweden”. A disjunctive query will retrieve a higher number of contexts; however, the ranking function will assign higher retrieval weights to the contexts fulfilling both rules (showing a “goal” *and* with title “Spain vs. Sweden”).

Structure

A major concern in an MPEG-7 retrieval model is to capture the hierarchical structure of the MPEG-7 descriptions for determining the entry points. Consider the following modified extract from Figure 10:

```
audiovisualsegment_1[ segment_1 [introduction game]
                      segment_2 [goal game]
    ]
```

This program expresses that ““audiovisualsegment_1” is composed of two segments, “segment_1” and “segment_2”, and the content of these segments is given by the terms “introduction game” and the terms “goal game”, respectively. The query

```
?- D[introduction]
```

retrieves both “segment_1” and “audiovisualsegment_1”; they both constitute entry points. The context “segment_1” is retrieved, since the term “introduction” occurs in “segment_1”, whereas the context “audiovisualsegment_1” is retrieved, since “segment_1” contains the term “introduction” and it is part of “audiovisualsegment_1.” Consider the following query:

```
?- D[introduction & goal]
```

The conjunction “introduction & goal” is true in the context “audiovisualsegment_1(segment_1, segment_2)”, i.e., the context that consists of both sub-contexts “segment_1” and “segment_2”. The term “goal” is true in "audiovisualsegment_1(segment_1, segment_2)" since it is true in "segment_2", and the term "introduction" is true in "audiovisualsegment_1(segment_1, segment_2)" since it is true in "segment_1". Neither sub-context on its own ("segment_1" or "segment_2") satisfies the query; only the context "audiovisualsegment_1(segment_1, segment_2)", i.e. "audiovisualsegment_1", satisfies the query. In other words, only "audiovisualsegment_1" is an entry point for that query. We call "audiovisualsegment_1(segment_1, segment_2)" an *augmented* context since its knowledge is augmented by the knowledge of the sub-contexts. An augmented context *accesses* its sub-contexts.

Uncertainty

A major task of the IR process (e.g. in the indexing phase) is the incorporation of the intrinsic uncertainty in representing documents and queries. Unlike XML, POOL provides probabilities that can be used to reflect this intrinsic uncertainty. POOL programs address two dimensions of uncertainty:

- the uncertainty of the content representation,
- the uncertainty that a super-context accesses its sub-contexts.

For the uncertain content representation, probabilities can be defined. For instance, in Figure 10, a probability value of 0.8 is assigned to the terms "spain" and "sweden", which means that the probability that the terms "spain" and "sweden" is true (false) is 0.8 (1.0 – 0.8 = 0.2) in the context "title_1", respectively. This could also be read as follows: 0.8 (0.2) is the probability that the term "spain" is (is not) a good indicator of the content of the context "title_1".

MPEG-7 provides tools that address the organisation of content (*content organisation*) [MPE00] that may be used to estimate these probabilities. For instance, the "Probability Model" DS provides a way to describe statistical functions for representing samples and classes of audio-visual data and descriptors using statistical approximation. A second DS, the "Analytic Model" DS, provides a way to describe properties of groups of objects, groups of descriptors and classifiers that assign semantic concepts based on descriptor characteristics, training examples and probabilities models. In our current implementation of the search component, such data was not available. Therefore, we use standard statistic-based techniques from IR. These are based on term frequency information (i.e. how often a term occurs in an element) and inverse document frequency (how many elements contain a particular term) [BR99].

The uncertain access reflects the effect of a sub-context on the knowledge of an augmented context. A weight can precede the opening of a sub-context. Consider the following modified extract of Figure 10:

```
audiovisualsegment_1[ 0.5 segment_1 [0.8 goal]
                      0.5 segment_2 [0.6 goal]
                    ]
```

In context "segment_1", "goal" is true with a probability of 0.8. In "segment_2", "goal" is true with a probability of 0.6. These probability values reflect the uncertain indexing of the two sub-contexts as described above. The two sub-contexts "segment_1" and "segment_2" are accessed by "audiovisualsegment_1" with a probability of 0.5. This probability reflects the effect of the knowledge of "segment_1" and "segment_2" on the augmented knowledge of "audiovisualsegment_1". The query

```
?- D[goal]
```

retrieves three contexts (i.e. identifies three entry points) with the following probabilities:

```
0.8 segment_1
0.6 segment_2
0.58 audiovisualsegment_1(segment_1, segment_2)
```

The sub-contexts are retrieved with the probabilities of "goal" being true in them. The augmented context "audiovisualsegment_1(segment_1, segment_2)" is retrieved with a probability of 0.58 which is the summation of three probabilities[13]: the probability that goal is true if both sub-contexts are accessed (0.5×0.5×(0.8×0.6+0.8×0.4+0.2×0.4)=0.23) plus the probability that goal is true if only "segment_1" is accessed (0.5×0.8×0.5=0.2) plus the probability that goal is true if only "segment_2" is accessed (0.5×0.6×0.5=0.15). The use

[13] The semantics of the probability computation is beyond the scope of this paper, and readers should refer to [Roe99].

of probabilities provide the “relevance-based” ranking of the segments forming the video sequence, which corresponds to determining entry points to the MPEG-7 structure.

Assigning access probabilities to sub-segments makes it possible to differentiate between the sub-segments of a segment. For instance, in a news program, the first shot of a news item often contains a summary of that item. The content of that sub-segment could then be given higher priority (a higher probability) than other segments in contributing to the content of the augmented segment, and the news program itself. The probabilities need however to be estimated either automatically (e.g. see [RLK+02] for a general methodology for deriving the estimates) or manually (e.g. the broadcast content producer) via, for instance, the instantiations of the “Probability Model” and “Analytic Model” DSs discussed above.

Data Propagation

One requirement for a model for MPEG-7 retrieval is to capture the propagation of MPEG-7 descriptions. Propagation is expressed in POOL via rules. For example, the rule

```
S.title(T) :- segment(X) & X[segment(S)] & X.title(T)
```

expresses that the title T is assigned to each segment S if S is a segment within the context of segment X and T is the title of X. In this way, we can model the decomposition of segments and the propagation of attributes downwards in the hierarchy [CMF96]. The above rule defined the title relationship in the root context, whereas the rule

```
X[S.title(T)] :- segment(X) & X[segment(S)] & X.title(T)
```

assigns the title to segments in the context of a decomposed segment X only.

4.3.2. From POOL to PRA Representation

For execution, POOL programs are translated into PRA (probabilistic relational algebra) programs. The translation of POOL into PRA follows the so-called object-relational approach; PRA programs consist of probabilistic relations that model aggregation, classification, and attributes as well as the terms and structure. The relations necessary for modelling MPEG-7 descriptions include:

- *term*: represents the terms occurring in the MPEG-7 descriptions.
- *attribute*: represents the relationships between MPEG-7 elements.
- *acc*: represents the structure of MPEG-7 descriptions.

As an example, Figure 11 shows an extract of the PRA representation of the MPEG-7 data of Figure 9, based on the POOL representation of Figure 10. The *term* relation models the occurrence of a term in a document; a high weight (probability) corresponds to a high term frequency. The *acc* relation reflects the hierarchical structure of the documents. Here, “segment_1” is a sub-context of “audiovisualsegment_1”. The higher the probability, the more impact has the content of the sub-context in describing the content of the super-context. In our example, no differentiation is made between the sub-segments, so the impact is full (the probability is 1.0). Relations reflecting spatial and temporal relationships between segments could also be used. The same criterion can be used in quantifying the impact of segments with respect to the content of spatially and temporally related segments. The *attribute* relation models relationships between elements. The last parameter of the attribute relation gives the context in which the relationship holds; for instance “audiovisualsegment_1” and “db” (the up-most context, i.e. the database of audio-visual material).

```
1.0 attribute(title,audiovisualsegment_1,"Spain vs. Sweden", db).
% The attribute "title" of audiovisualsegment_1 having value
% “Spain vs Sweden” in the context of the database db
1.0 attribute(genre,audiovisualsegment_1,"Sports”, db).
1.0 attribute(language,audiovisualsegment_1,"English”, db).

1.0 term(soccer, audiovisualsegment_1).
1.0 term(game, audiovisualsegment_1).
0.8 term(spain, audiovisualsegment_1).
0.8 term(sweden, audiovisualsegment_1).
% The term tuples represent the "word" content of audiovisualsegment_1.

1.0 acc(audiovisualsegment_1, segment_1). % representation of the logical structure

1.0 attribute(medialocator,segment_1,”(?)”,audiovisualsegment_1).
1.0 term(introduction,segment_1)
```

Figure 11: Example of the (simplified) PRA representation of the MPEG-7 Description of Figure 9

One can see that PRA programs are “assembler-like”. The assembler nature of PRA programs was the motivation for defining POOL, thus having a more abstract and object-oriented view of the data than that provided by PRA.

4.3.3 The Retrieval Function

The retrieval function is performed through a probabilistic interpretation of standard relational algebra operations (e.g., UNION, JOIN, PROJECT, etc), where the relations are those obtained from the translation from POOL to PRA (e.g. see Figure 11). The retrieval process is implemented through PRA programs.

As described in the previous section, probability values are attached to tuples (e.g. *term* tuples, *acc* tuples, *attribute* tuples) and capture the uncertainty of the representation. The retrieval process accesses these tuples, and the probabilities are combined to infer the ranking. For instance, in the PROJECT operation, when independence is assumed, the probabilities of duplicate tuples are added. The complete details of the calculation of the probabilities are beyond the scope of this paper, and interested readers should refer to [FR97]. In this section, we describe the retrieval process through examples of PRA retrieval functions.

A simple retrieval function that considers only the terms present in segments and queries would be implemented as follows ($n refers to the columns of the relations). For instance, a query about “goal” leads to the execution of the following PRA program[14]:

```
qterm(goal)
segment_index = term                    % renaming of relation for later use
retrieved_segments = PROJECT[$3](JOIN[$1=$1](qterm,segment_index))
```

The *qterm* relation represents the actual query (the terms composing the query, here the term “goal”). The first equation (*segment_index*) computes the segment index. The second equation matches (JOIN) the query terms with the document terms, and then returns (PROJECT) the matched segments. Applying this PRA program to our example, the two contexts “audiovisualsegment_1” and “segment_1” are retrieved. The retrieval status value (RSV) of “segment_1”, which is used to rank segments in the retrieval result is computed as follows[15]:

$$\text{RSV(segment\_1)} = \text{qterm(goal)} \times \text{segment\_index(goal, segment\_1)}$$

More sophisticated indexing schemes can be used. For instance, inverse document frequency can be modelled by a relation *termspace*:

```
0.6 termspace(goal)
0.2 termspace(game)
```

[14] In POOL, such query was formulated as “?-D[goal]”, see Section 4.3.1.

[15] The computation requires the specification of the so-called disjointness key of the termspace relation. The clause _disjointness_key(termspace, “”) tells HySpirit about the disjointness of the termspace tuples. Details can be found in [FR97] and [RF98].

…

This relation corresponds to the probabilistic interpretation of the inverse document (here segment) frequency of terms. The retrieval function (the segment indexing equation) is then expressed as follows:

```
segment_index = JOIN[$1,$1](term,termspace)
```

User preferences with respect to topicality can be incorporated in a similar way. For instance, suppose that a user has a preference for goals, and in particular those being scored by the French player Zidane. These preferences can be modelled by the relation *userspace* as follows:

```
0.7 userspace(goal)
1.0 userspace(zidane)
```

An additional equation would be inserted before the *retrieve_segments* equation joining the *segment_index* relation with the *userspace* relation. This would retrieve at higher rank any segments showing goals scored by Zidane, then segments showing goals scored by other players.

With the *acc* relation, the representation of super-contexts (e.g. "audiovisualsegment_1") is specified in terms of the representation of its sub-contexts (e.g. "segment_1"). The content of the super-context is augmented by that of its sub-contexts. This is modelled by the following PRA program:

```
term_augmented = PROJECT[$1,$3](JOIN[$2=$2](term,acc))
```

Applied to our example, augmentation produces "term_augmented(goal, audiovisualsegment_1)", i.e. we have augmented the description of "audiovisualsegment_1" by that of "segment_1". The probability of the augmented term in the context "audiovisualsegment_1" is:

P(term_augmented(goal, audiovisualsegment_1))
= P(term(goal,segment_1)) × P(acc(audiovisualsegment_1,segment_1))

The *term* probability is multiplied with the *acc* probability, resulting in a smaller or at most equal probability in the super-contexts. The probability of an *augmented term* should be smaller since the super-contexts are larger contexts than the sub-contexts, and the probability should be normalised with respect to the document size and the *acc* value.

Retrieval is now performed with the *term_augmented* relation, yielding the super-contexts. In our example, "audiovisualsegment_1" and "segment_1" are retrieved. The *acc* relation allows the retrieval of entry points instead of components of fixed granularity levels.

We finish with an example of a combined content-based and factual query. The query on videos about "goal" or produced by the "BBC" would be expressed as the following PRA program (a simple indexing schema is used):

```
qterm(goal)
segment_index = term
retrieved_segments_1 = PROJECT[$3](JOIN[$1=$1](qterm,segment_index))
retrieved_segments_2 = PROJECT[$2](SELECT[creator =~ /BBC/](attribute))
retrieved_segments = UNION(retrieved_segments_1,retrieved_segments_2)
```

The first three expressions of the retrieval function have already explained. The fourth expression (the third equation) retrieves all segments with creator similar to "BBC"[16]. The last expression (fourth equation) performs a probabilistic union over the two query criteria.

5. A User Interface for Searching

This section describes the user interaction involved in searching (5.1), and then proposes a design for integrating the search functionality on the SAMBITS terminal (Section 5.2). The design supports TV viewers in formulating searches for further material based on the description of what they are viewing, and enables the use of remote control as the only controlling input device. This is the proposed front-end to using the underlying MPEG-7 data for

[16] PRA provides regular expression matching.

searching, as described in the previous sections. A brief description of the techniques used in its implementation is given in Section 5.3.

5.1 Model of Interaction

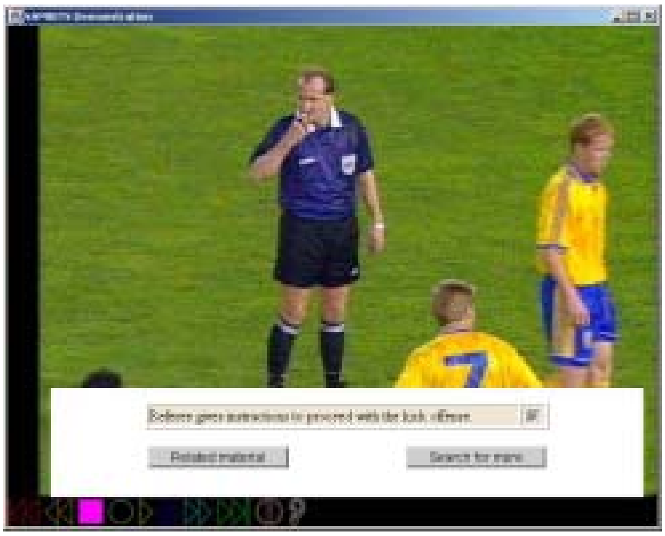
Our model of the search interaction consists of the following three stages:

- Viewing description of what is currently being presented.
- Optionally selecting specific terms of interest in the description.
- Searching for more related material or information.

After viewing the description of a programme element, users are given the option to directly perform a search for additional related material or restrict their query to only those terms of the description that best represent their information need. For example, in [HML+01], a description of a segment of a football match consists of the name of the player, the name of the team, the name of the stadium and so on. Users may use the whole description to search for additional material, which is then used as the query. However, if they want additional information about the stadium only, they can select this item (i.e. the name of the stadium) to be used as the query.

5.2 Design of the User Interface

A design of the user interface that supports the above search interaction model is proposed in Figures 13 and 14. As shown in Figure 13, the description of what is currently being viewed is presented overlaid on the video screen.



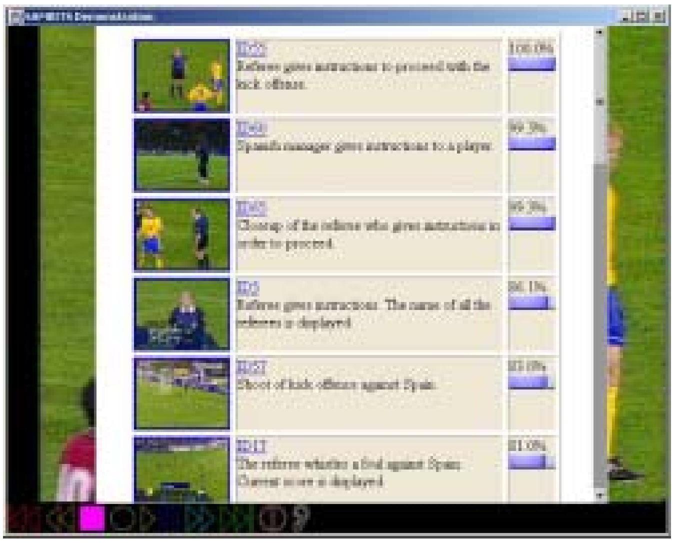
Referee gives instructions to proceed with the kick offsure ☑

Figure 13: Sample screen shot showing the building of a query

A button provides viewers with the option to submit a search for additional material. Check boxes enable selection of items, in order to restrict the search according to the specific interest of users. All check boxes are initially ticked, indicating that by default the complete description is taken into account for searching.

In the context of SAMBITS, a remote control button is mapped to the search button of the screen, the backward and the forward buttons are used to navigate between items in the description, and the ok button is used to enable selection of a desired item.

The list of search results is also presented overlaid on the video screen, as shown in Figure 14. For each result item, a thumbnail image is displayed, which provides a visual indication of the material, a brief description of it, and its percentage of relevance and a graphical representation of the relevance value. A link to the video segment provides access to the retrieved material, which is then presented in a full screen.



[illegible]
Referee gives instructions to proceed with the kick offence.
100.0%

[illegible]
Spanish manager gives instructions to a player.
99.3%

[illegible]
Close-up of the referee who gives instructions in order to proceed.
99.3%

[illegible]
Referee gives instructions. The names of all the referees is displayed.
86.1%

[illegible]
Shoot of kick offence against Spain.
85.0%

[illegible]
The referee whistles a foul against Spain. Current score is displayed.
81.0%

Figure 14: Sample screenshot showing the list of search results

In the context of SAMBITS, the backward and the forward buttons of the remote control are used for navigating between the items in the list of search results, and the ok button for following the link to view the selected item.

Note that viewers have the option of editing a set of parameters in their personal profile concerning the display of the list of search results, exploiting the *Navigation and Access* component of MPEG-7. Such parameters are, for example, the number of results per page, whether they see a thumbnail or not, or the level of detail of the description of the results [ILR01].

5.3 Implementation

In SAMBITS, the user interface that supports the search functionality is implemented as a web application. The integrated web browser of the terminal is used for rendering and displaying the overlays, and a web server has been set up locally at the terminal to handle the interactivity and pass the search requests to the retrieval system. The description of what is being presented at a given time is extracted from the MPEG-7 structure, and transformed from XML to HTML in order to be displayed using an XSLT processor and an appropriate style sheet. Another XSLT style sheet is used for transforming the search results from the XML format in which they are produced by the retrieval engine to HTML suitable for rendering by the web browser.

6. Conclusions and Future work

This paper describes the design and implementation of a model for searching multimedia data (audio-visual data) based on their associated MPEG-7 descriptions, in the context of a broadcast application. First, MPEG-7 description tools were examined to determine the parts of MPEG-7 that are relevant to our application. It was found that the MPEG-7 standard provides the necessary tools for searching audio-visual data based on their associated MPEG-7 descriptions. Second, we identified a number of requirements for a retrieval model for audio-visual data annotated with MPEG-7 DSs and Ds. Third, we designed our model using the HySpirit software kit, and took take into account these requirements. In particular, our model:

- explicitly incorporates the structural aspect of MPEG-7 descriptions so as to retrieve elements at various levels of granularity (e.g. segments vs. sub-segments),
- permits content-based and factual queries since both should be allowed from the nature of MPEG-7 descriptions.

More advanced search methods can be implemented. Some of them have been highlighted in the paper (e.g. data propagation, user preferences). Using a software development kit such as HySpirit makes it possible to design and implement many requirements for the effective retrieval of broadcast data.

A complete prototype of the search component was developed and integrated with the SAMBITS broadcast terminal. The integrated prototype involved many issues not discussed in this paper: the decoding of incoming MPEG-7 data from the broadcast (MPEG-2) stream; timing information to specify which MPEG-7 description corresponds to which part of the video sequence; the automatic building of queries via the user interfaces; the local storage of the video stream and MPEG-7 description on the terminal; discarding stored videos from the local storage that are too old.

Finally, exactly what types of searches will be performed by real TV users need to be investigated through user studies, since it can be that end-users may only be willing to perform simple searches. The SAMBITS consortium is currently carrying out user studies, that will not only provide answers to this question, but general input regarding the full SAMBITS terminal.

7. Acknowledgements

This work was funded by SAMBITS (System for Advanced Multimedia Broadcast and IT Services, IST-1999-12605). We would like to thank all the members of the SAMBITS consortium for their involvement in this work. In particular, we would like to thank Alan Pearmain and Yakup Parker for their insightful comments and feedback, Pat Healey for his input for the design of the user interface, and Damien Papworth for his technical input.

8. References

[ABB01] J Assfalg, M Bertini and A Del Bimbo. Information Extraction from Video Streams for Retrieval by Content. *23rd European Colloquium on Information Retrieval Research*, Darmstadt, Germany, April 2001.

[ACC+97] S Abiteboul, S Cluet, V Christophides, T Milo, G Moerkotte and J Simeon. Querying documents in object databases. *International Journal on Digital Libraries*, 1:1-9, 1997.

[BR99] R Baeza-Yates and B Ribeiro-Neto. *Modern Information Retrieval*. Addison Wesley, 1999.

[BZC+01] AB Benitez, D Zhong, S-F Chang, and JR Smith. MPEG-7 MDS Content Description Tools and Applications. *International Conference on Computer Analysis of Images and Patterns (CAIP-2001)*, Warsaw, Poland, 2001.

[Cal94] J Callan. Passage-Level Evidence in Document Retrieval. *ACM SIGIR*, pp 302-310, Dublin, 1994.

[Chi97] Y Chiaramella. Browsing and querying: two complementary approaches for multimedia information retrieval. *Hypermedia - Information Retrieval – Multimedia*, Dortmund, Germany, 1997.

[CMF96] Y Chiaramella, P Mulhem and F Fourel. A model for multimedia information retrieval. Technical Report Fermi ESPRIT BRA 8134, University of Glasgow, 1996.

[FA01] N Fatemi and O Abou Khaled. Indexing and Retrieval of TV News Programs based on MPEG-7. *IEEE International Conference on Consumer Electronics (ICCE2001)*, Los Angeles, CA, June 2001.

[FH94] R Fagin and J Halpern. *Reasoning about Knowledge*. MIT Press, Cambridge, Massachusetts, 1995.

[Fri88] M Frisse. Searching for information in a hypertext medical handbook. *Communications of the ACM*, 31(7):880-886, 1988.

[FR97] N Fuhr and T Roelleke. A probabilistic relational algebra for the integration of information retrieval and database systems. *ACM Transactions on Information Systems*, 14(1), 1997.

[FR98] N Fuhr and T Roelleke. HySpirit - A probabilistic inference engine for hypermedia retrieval in large databases. *International Conference on Extending Database Technology (EDBT)*, Valencia, Spain, 1998.

[GBC+99] D Gibbon, A Basso, R Civanlar, Q Huang, E Levin and R Pieraccini. Browsing and Retrieval of Full Broadcast-Quality Video.
http://www.research.att.com/~mrc/pv99/CONTENTS/PAPERS/GIBBON/sbtv_htm, 1999.

[Hau95] AG Hautmann. Speech Recognition in the Informedia Digital Video Library: Uses and Limitations, *ICTAI95*, 1995.

[HI98] J Hunter and R Iannella. The application of Metadata Standards to Video Indexing. *Second European Conference on Research and Advanced Technology for Digital Libraries*, Crete, Greece, September 1998.

[HK99] O Hori and T Kaneko. Results of Spatio-Temporal Region DS Core/Validation Experiment. *ISO/IEC JTC1/SC29/WG11/MPEG99/M5414*, Maui, December 1999.

[HLM+00] PGT Healey, M Lalmas, E Moutogianni, Y Paker and A Pearmain. Integrating internet and digital video broadcast data. *4th world Multiconference on Systemics, Cybernetics and Informatics (SCI 2000), Information Systems*, Vol I, pp 624-627, Orlando, Florida, U.S.A, July 2000.

[HM92] J Halpern and Y Moses. A Guide to Completeness and Complexity for Modal Logics of Knowledge and Belief. *Artificial Intelligence*, 54:319-379, 1992.

[HML+01] PGT Healey, E Moutogianni, M Lalmas, Y Paker, D Papworth and A Pearmain. Requirements for Broadcast and Internet Integration. *Media Future*, Florence, Italy, May 2001.

[Hun99] J Hunter. MPEG-7 Behind the Scenes. *D-Lib Magazine*, 5(9), September 1999.

[ILR01] D Ileperuma, M Lalmas and T Roelleke. MPEG-7 for an integrated access to broadcast and Internet data. *Media Future*, Florence, Italy, May 2001.

[ISO00a] ISO MPEG-7. Text of ISO/IEC FDIS 15938-1 Information Technology - Multimedia Content Description Interface - Part 1 Systems, ISO/IEC JTC 1/SC 29/WG 11 N4285, October 2001

[ISO00b] ISO MPEG-7. Text of ISO/IEC FDIS 15938-2 Information Technology - Multimedia Content Description Interface - Part 2 Description Definition Language, ISO/IEC JTC 1/SC 29/WG 11 N4288, October 2001

[ISO00c] ISO MPEG-7. Text of ISO/IEC FDIS 15938-3 Information Technology - Multimedia Content Description Interface - Part 3 Visual, ISO/IEC JTC 1/SC 29/WG 11 N4358, October 2001

[ISO00d] ISO MPEG-7. Text of ISO/IEC FDIS 15938-4 Information Technology - Multimedia Content Description Interface - Part 4 Audio, ISO/IEC JTC 1/SC 29/WG 11 N4224, October 2001

[ISO00e] ISO MPEG-7. Text of ISO/IEC FDIS15938-5 Information Technology - Multimedia Content Description Interface - Part 5 Multimedia Description Schemes, ISO/IEC JTC 1/SC 29/WG 11 N4242, October 2001

[KLW95] M Kifer, G Lausen and J Wu. Logical Foundations of Object-Oriented and Frame-Based Languages. *Journal of the Associations for Computing Machinery*, 42(4):741-843, 1995.

[LM00] M Lalmas and E Moutogianni. A Dempster-Shafer indexing for the focussed retrieval of hierarchically structured documents: Implementation and experiments on a web museum collection. *RIAO*, Paris, France, 2000.

[LR98] M Lalmas and I Ruthven. Representing and retrieving structured documents with Dempster-Shafer's theory of evidence: Modelling and evaluation. *Journal of Documentation*, 54, 5: 529-565. 1998.

[LS99] J Llach and P Salembier. Analysis of video sequences: Table of contents and index creation. *International Workshop on Very Low Bit-rate Video Coding, VLBV'99,* pp 52-56, Kyoto, Japan, October 1999.

[LST+00] H Lee, AF Smeaton, C O'Toole, N Murphy, S Marlow and NE O'Connor. Recording, Analysing and Browsing System, *RIAO*, Paris, France, 2000.

[MJK+98] S Myaeng, DH Jang, MS Kim and ZC Zhoo. A flexible model for retrieval of SGML documents. *ACM-SIGIR Conference on Research and Development in Information Retrieval*, Melbourne, Australia, pp 138-145, 1998.

[MPE99] MPEG Requirements Group. *MPEG-7 Requirements Document v.16*, ISO/IEC JTC 1/SC 29/WG 11 N4510 , MPEG Pattaya Meeting, December 2001.

[MPE00] MPEG Requirements Group. *Overview of the MPEG-7 Standard (v.5.0),* ISO/IEC JTC 1/SC 29/WG 11 N4031, MPEG Singapore Meeting, March 2001.

[MPE01] MPEG Requirements Group. *MPEG-7 Applications Document v.10*, ISO/IEC JTC 1/SC 29/WG 11 N3934, MPEG Pisa Meeting, January 2001.

[PLM+01] A Pearmain, M Lalmas, E Moutogianni, D Papworth, PGT Healy and T Roelleke. Using MPEG-7 at the Consumer Terminal in Broadcasting, EURASIP (European Association for Signal, Speech and Image Processing) Journal on Applied Signal Processing, April 2002

[Put01] W Putz. The usage of MPEG-7 Metadata in a Broadcast Application. *Media Future*, Florence, Italy, May 2001.

[Rij79] CJ van Rijsbergen. *Information Retrieval*. Butterworths, London, 2 edition, 1979.

[Rij86] CJ van Risjbergen. A Non-Classical Logic for Information Retrieval. *The Computer Journal*, 29(6):481-485, 1986.

[RLK+02] T Roelleke, M Lalmas Gabriella Kazai, Ian Ruthven and S Quicker. The Accessibility Dimension for Structured Document Retrieval, *24th ECIR'02*, Glasgow, March 2002.

[RLS98] J Robie, Lapp and D Schach. XML Query Language (XQL). *Query Language Workshop, W3C*, December 1998, http://www.w3.org/TandS/QL/QL98/.

[Roe99] T Roelleke. POOL: Probabilistic Object-Oriented Logical Representation and Retrieval of Complex Objects - A Model for Hypermedia Retrieval. PhD thesis, University of Dortmund, Germany, 1999.

[SAB93] G Salton, J Allan and C. Buckley. Approaches to passage retrieval in full text information systems. *ACM SGIR*, Pittsburgh, pp 49-58, 1993.

[SAM00] *System for Advanced Multimedia Broadcast and IT Services*, IST-2000-12605, http://www.irt.de/sambits/, 2000.

[SC00] B Smyth and P Cotter. A personalised Television Listings Service. *Communication of the ACM*, 43(8):107-111, 2000.

[SLG00] P Salembier, J Llach and L Garrido. Visual segment tree creation for MPEG-7 description schemes. *International Conference on Multimedia and Expo, ICME'2000*, Vol 2, pp 907-910, New York City, NY, USA, July 2000.

[Sme00] A Smeaton. *Indexing, Browsing and Searching of Digital Video and Digital Audio Information*. Tutorial Notes, European Summer School in Information Retrieval, Varenna, Lago di Como, Italy, 2000

[Wil94] R Wilkinson. Effective Retrieval of Structured Documents. *ACM-SIGIR*, Dublin, pp 311-317, 1994.

[XML00a] XML Schema Part 0. *Primer, W3C Candidate Recommendation,* 24 October 2000, http://www.w3.org/TR/xmlschema-0/

[XML00b] XML Schema Part 1. *Structures, W3C Candidate Recommendation,* 24 October 2000, http://www.w3.org/TR/xmlschema-1/

[XML00c] XML Schema Part 2. *Datatypes, W3C Candidate Recommendation*, 24 October 2000, http://www.w3.org/TR/xmlschema-2/

[XML00d] Extensible Markup Language (XML) 1.0 (Second Edition), W3C Recommendation, 6 October 2000, http://www.w3.org/TR/REC-xml

[XML00e] *Namespaces in XML, W3C Recommendation*, 14 January 2999, http://www.w3.org/TR/REC-xml-names/

[ZTS+95] H Zhang, SY Tan, SW Smoliar and G Yihong. Automatic parsing and indexing of news video. *Multimedia Systems*, 2:256-266, 1995