Project no. FP6-507752

MUSCLE

Network of Excellence
Multimedia Understanding through Semantics, Computation and Learning

Guidelines on standards for representation and exchange of data, meta-data and semantics

Start date of project: 1 March 2004 Duration: 48 months

Deliverable Type: PU
Number: 2
Nature: Report
Task: WP 9

Name of responsible:Ovidio Salvetti
Editor & editor's address: Maria Grazia Di Bono

Revision 1.0

Project co-funded by the European Commission within the Sixth Framework Programme (2002-2006)		
Dissemination Level		
PU	Public	✓
PP	Restricted to other programme participants (including the Commission Services)	
RE	Restricted to a group specified by the consortium (including the Commission Services)	
CO	Confidential, only for members of the consortium (including the Commission Services)	

Keyword List: Multimedia Metadata,

Contents

Chapter 1

Introduction

MM metadata is the value-added information which documents the administrative, descriptive, preservation, technical and usage history characteristics associated with MM data. MM Metadata can be also classified according to some criteria considering the level of data description, the producibility and the domain dependence. They can be classified by [22]:

- *Level*: we can distinguish between a *technical level,* in which lower level aspects of the multimedia content are described, and a *semantic level,* in which aspects of higher level of abstraction on the multimedia content are taken into account.
- *Producibility*: the production of metadata can either be automatic which is a very desirable property from the economic point of view and regards more frequently the low level technical metadata; for semantic metadata describing the information covered by multimedia content it is typically required human knowledge. So in these cases the metadata production is manually performed.
- *Dependencies*: metadata can be domain-dependent, for instance the position of a tumor can be interesting for medical applications, while the colour distribution of an image can be useful for many application domains. Metadata can also be media type-dependent considering for instance the colour distribution as applicable only to visual media while the creation date applicable to any media.

From a state-of-the-art study performed outside the NoE, through the examination of several MM metadata standards in different scientific communities, and inside the NoE, through the formulation and distribution of a specific *Questionnaire*, we found that two different kind of applications on MM data emerged [15]. If from one hand people deal with MM data for scientific purposes using several consolidated MM data processing algorithms, from the other hand a content representation is required for applications following the *Content Based Query* (CBQ) paradigm. Respect to these needs, we found that many metadata standards have been developed by several communities with different requirements for a certain number of specific application contexts. The outlined scenario certainly appears as a possible limit for interoperability needs among different NoE communities. So, we pointed out two relevant aspects to take into consideration in order to achieve an efficient use of MM metadata that could allow to read, search and exchange MM data in the network, even understandable at a higher level. On one hand, we worked on the need of identifying a common MM standard format able to describe and represent the intrinsic heterogeneous nature of MM resources and their semantics, while on the other hand we studied the possibility to use more abstract models (*ontologies*) and related *mapping tools* to “represent” and “translate” different metadata sets whose elements are correlated on the basis of the same or similar meanings ("Semantic Mapping"), so that MM applications can use ontology knowledge in addition to the metadata.

Multimedia representation standards

A large number of metadata standardisation initiatives has been developed in recent years, in order to describe multimedia contents in so many different domains and to grant sharing, exchanging and interoperability across wide range networks. In particular we can discriminate among standardization proposals aiming at finding a standard and static representation for

specific domains and other initiatives that can provide models more easily applicable to more general domains.
From the study of the State of the Art on MM metadata standards, appears that one of the more recent approaches is to combine a specific MM metadata standard with other standards useful to describe other application domains, in order to have a more complete characterisation of the specific problem without creating a new standard. New metadata initiatives such as TV-Anytime [2], MPEG-21 [3], NewsML [4] and communities such as the museum, educational, medical and geo-spatial communities, want to combine MPEG-7 MM descriptions with new and existing metadata standards for simple resource discovery (Dublin Core [5]), rights management (INDECS [6]), geo-spatial (FGDC), educational (GEM [8], IEEE LOM) and museum (CIDOC CRM [9]) content, to satisfy their domain-specific requirements. In order to do this, it is necessary to have a common understanding of the semantic relationships between metadata terms from different domains. To this purpose, XML schema provides a first support for expressing semantic knowledge and RDF schema can provide a way to do this, but there are other several frameworks able to realise the same task.

Considering the NoE prospective, the questionnaire results have shown that the MM metadata standards are not commonly used within the NoE. This, together with the fact that the only standard used by some members is the MPEG-7, let us think and propose this one as a possible metadata standard to be used, eventually in conjunction with other standardised metadata tools (i.e. RDF, OWL or its extensions) that can allow the use of different metadata standards, through the representation of an *upper model* able to integrate information about heterogeneous resources of different and specific application domains.
So, two ways can be crossable at this point. One way could consider the possibility to identify and use only a specific standard able to represent all the useful information type and the second one, in line with other current initiatives, could think the definition of a larger and more comprehensive representation model able to grant interoperability across the network. This model can be obtained integrating different standards, maintaining MPEG-7 and selecting some other standard useful for the NoE purposes to represent MM data in the form of images, audios, videos and their decompositions and all the MM documents at low, high and semantic level.
Some of the studied metadata standards, useful to describe the NoE MM data, can be considered, taking advantage from the specific representation potentialities of each standard examined.
Among current integration initiatives, an interesting one is the Moving Image Collections (MIC) [10] [11], which is an integrated online catalogue of moving images held by a variety of organizations, including libraries, museums, archives and television broadcasting companies. Several standards have been examined by this initiative and a sub-set of them has been selected in order to achieve its purposes [12]. In particular, standards as MARC [13], DC [5], MPEG-7 [14] have been taken into account, exploiting the representative power of each of them. On one hand, the DC standard provides a specific information type and a great flexibility, is easy to learn, but even if it ensure interoperability with other schemes, it lacks of support for multiple formats and technical descriptions. On the other hand MPEG-7 [14] bridges this gap integrating technical, structural, administrative and descriptive metadata and allows integrating textual and non-textual indexing (e.g. melody and speech recognition with textual data elements).
At the moment, on the base of the actually acquired information from the NoE, we can sustain a strategy to consider the MPEG-7 as a base to start from for the definition of an eventually more complete description model. So, in this deliverable, the guidelines on MM representation standards mostly focused on MPEG-7 description [1][14].

Guidelines on some of the possible interoperability standard tools useful for the MM data, metadata and semantics integration have been provided, mostly focusing on ontology models and related languages for semantic web like RDF, OWL and its extensions.
In addition to MPEG-7, the Appendix A provides a list of further possible standards for MM data representation and exchange, considering the standards examined and documented in the previous deliverable [15].
In Appendix B a list of some of the more relevant reference links to MM descriptor extraction tools is shown.

Chapter 2

MPEG 7 overview

Context, objectives and applications

The use of digital cameras, personal computers and the Internet has made producing and distributing multimedia (MM) content easier today than ever before.
An incommensurable amount of MM information is becoming available in digital form, in digital archives, on the World Wide Web, in broadcast data streams and in personal and professional databases, and this amount is only growing [1].
A lot of different scenarios are taking place in several specific domains, as image understanding (surveillance, intelligent vision, smart cameras, etc.) and media conversion (speech to text, picture to speech, speech to picture, etc.). Other scenarios are information retrieval (quickly and efficiently searching for various types of MM documents of interest to the user) and filtering in a stream of audiovisual content description (to receive only those MM data items which satisfy the user's preferences).
MM information content often depends on how easy it can be found, retrieved, accessed, filtered and managed. Content that cannot be easily found is like content that does not exists.
The same digital technologies producing and publishing MM content can also help in analysing and classifying it, extracting and manipulating different type of features specific for applications and in searching and discovering content.
People looking for content have used text-based browser. These text-based engines rely on human operators for manually describing MM content using free text annotations and keyword lists, employing resources in an expensive process in which the cost increases with the growing amount of content. Furthermore, the descriptions are often subjective and related to the specific domain that them were created for. So, the necessity to automatically, objectively and domain independently describe, index and annotate MM content, using tools that automatically can extract possibly complex audiovisual features (e.g. colour, shape, texture, melody and sound envelope, etc.) from content to integrate them with text based descriptions.
In this context, the MPEG-7 project aimed at specifying a standard way to describe MM information to facilitate a fast and efficient identification and management of interesting and relevant MM descriptions. They are of different nature, both textual (e.g. annotations, names, titles, etc.) and non-textual (e.g. statistical features, technical parameters, etc.). So a standard way to produce MM content descriptions, allows content to be reached and exchanged across different systems on the network. Therefore, it defines an environment in which different providers can create infrastructures to transparently manage MM content, increasing interoperability.
The driven principles of MPEG-7, that seem to express the vision behind the standard, can be summarized as follows [1][14]:

- *Wide applications*: the standard can be applied to the content associated to any application domain, so the content can be stored and can be made available on-line, off-line or streamed.
- *Content related*: the standard allows the creation of descriptions that can be used stand-alone (e.g. a summary of a content), multiplexing them with the content itself (e.g. in broadcast applications) or linking them to one or more content versions.
- *Wide list of data types*: MPEG-7 deals with a wide range of data types (e.g. speech, audio, images, video, graphics, 3D-models, etc.) putting more emphasis on

audiovisual information and tools, while the existing tools could be considered (SGML, XML, RDF, etc.) for textual data.

- *Media independence*: the standard could be applicable independently of the media type (e.g. paper, film, tape, CD, hard disk, digital broadcast or internet streaming).
- *Format independence*: the standard could be applicable independently of the media formats. For example, in the case of audiovisual content, it can be represented using PAL, NTSC, MPEG-1/2/4.
- *Object based*: MPEG-7 allows constructing an object based MM content representation. In details, MM content can be described using a collection of MM objects independently accessible.
- *Abstraction level*: the standard allows defining two level of abstraction. The low-level (e.g. statistical features) and a high-level related to semantic meaning of content. Often the low-level features can be automatically extracted, while for the semantic features need to be manually or semi-automatically extracted.
- *Extensibility*: MPEG-7 allows the extension of the description tools in a standard way, to grant as much as possible interoperability.

To reach the objectives of the standard there is the need of a set of specific tools, in particular, Descriptors (to define the elements), Description schemes (to define the structures), Data Definition Language (DDL, to extend the predefined set of tools) and a certain number of System tools.
In more details:

1. Descriptors, that:
   i. Are a representation of a feature
   ii. Define the syntax and the semantic of a feature representation
   iii. Allows an estimate of the corresponding feature through the descriptor value. It is possible that the same feature is represented by several descriptors. An example can be a time code to represent the duration, a colour histogram for representing colour, and so on.
2. Description Schemes, that:
   i. Specify the structure and the semantics of the relations among the component objects, which can be Descriptors and also Description Schemes.
   ii. Have descriptive information and can participate into many-to-one relationships with other description elements. An example can be a movie that can be temporally structured into scenes, including some textual description.
3. The Description Definition Language (DDL), that:
   i. Is a language that allows the definition of new description schemes and descriptors
   ii. Also allows the extension of the existing description schemes
4. System tools, which related to the transformation in a binary format, transport and storage of descriptions and also the management of the intellectual property.

There are a lot of different application domains that can benefit of the MPEG-7 standard, from Education, Cultural services and Entertainment domains to the field of Geographic Information Systems and Remote Sensing, Surveillance and Biomedical applications.
The main application classes can be summarized as follows:

- *Pull applications*: based on storage and retrieval of audiovisual information from databases, like music applications, historical speech databases, movie scene retrieval, etc.
- *Push applications*: using agent-driven selection of MM information and filtering, intelligent multimedia presentations, etc.

- *Professional applications*: related to specialized application environments like biomedical, remote sensing, and educational applications.

In comparison with other emerging standards and solutions, MPEG-7 is characterized first of all by its *generality*, related to its ability to describe content from many different application domains, the *integration* of low-level and high-level descriptors into a singular architecture, its *extensibility* (through DDL) that allows users to evolve the standard adding new descriptors or modifying other to reach specific application needs.

Description Definition Language: a brief overview

The Description Definition Language (DDL) provides the foundations of the standard. It provides a language to define the structure and the content of an MPEG-7 document. It is a scheme-based language to represent the results of a modelling process of MM data according to a set of syntactic, structural and value restrictions to which MPEG-7 descriptors and description schemes have to conform. It also offers a set of syntactic rules that can be used by users to modify or to extend existing descriptors or description schemes by combining and refining them for specific applications.
The MPEG-7 instances are XML documents that conform to an MPEG-7 schema expressed in DDL describing MM content.
DDL consists in three logical components:

- XML Schema structural components
- XML Schema data types
- MPEG-7 specific extensions

The most significant structural components of an XML schema can be grouped as follows [1]:

- <u>Namespaces</u>: XML Namespaces provide a simple way to assign universally unique names to element types or attribute names, through a collection of names identified by an *Uniform Resource Identifier (URI)*, so that they can be reused in other XML documents. In the MPEG-7 context, the mechanism of the Namespace, allow the use of descriptors and description schemes belonging to different MPEG-7 schemas.
- <u>Element declarations</u>: specify a type definition for a schema element and can be explicit or by reference. It provides, for each element declared in the schema either occurrences of the elements (using information like minOccurr or maxOccurr) and default information (using default attributes). Sometimes it is more convenient to refer an existing element rather than to declare a new one. So, the declaration by reference consists in the use of an existing global element, declared elsewhere in the schema, moreover, its content must be consistent with the defined type and all the constraints about the element referred.
- <u>Attribute declarations</u>: allow specifying information about a declared element in the schema. In particular, an attribute specifies the name, the type, the use (which can be *required/optional/prohibited*) and the default value.
- <u>Type definitions</u>: define internal components of the schema that can be used in element or attribute declarations. XML Schema provides three classes of type definition, in particular *simple, complex* and *derived* type definitions:
    - *Simple types*: cannot have children element or support attributes. XML Schema provides a large set of simple types, using *"built-in primitive types"* and *"built-in derived types"* (*XML Schema data types*) and the application of constraints and restrictions (e.g. *enumeration* facet for string type or *minInclusive* and *maxInclusive* facet for integer type) on other simple types allows to derive new simple types. There are also two aggregate simple type:

    List and *Union*. The first type is composed by a sequence of atomic simple types, while the second one is the union of multiple atomic types and/or Lists.
  - *Complex types*: unlike simple types, they can have children elements carry attributes. Complex types provide mechanisms to constraint the appearance and nature of attributes and children elements and to derive other complex types from simple or complex types using extensions and restrictions.
  - *Derived types*: it is possible to create derived types by applying extensions or restrictions to simple or complex type definitions.
- Group definitions: the attribute Group and group elements provide a mechanism to create groups of attributes and groups of elements respectively. In particular, groups of elements can b created through a *sequence* or a *choice*.

MPEG-7 specific extensions allow supporting other important features, which are:
- *"Array"* and *"Matrix"* data types
- Derived data types: "basicTimePoint" and "basicDuration" data types, that specify respectively a time point according to the Gregorian dates and the duration of a time period days and time of day. In both cases the format is based on the ISO 8601 standard.

For the first case, using the *list* data type, there are two methods for specifying the size of array 1D or multidimensional matrices. The first method is the "*mpeg7:dimension"* facet, that is a list of positive integers and allows to specify the fix dimension of a matrix. This facet is an MPEG-7 extension and is not compliant to XML schema language and for this reason it is necessary to aid it into *<annotation><appinfo>* to permit the parser validation. The second method is the "*mpeg7:dim"* attribute, to support the parameterised array and matrix sizes that will be specified at the time of instantiation.

Overview of Multimedia Description Schemes and Schema Tools

MPEG-7 descriptors and description schemes can be organizing on the base of their functionalities in the follow classes [1]:
- *Basic elements* (Schema tools, Basic data types, links and media localization, basic tools)
- *Content Management* (media, creation and production, usage)
- *Content Description* (structural ad semantic aspects)
- *Content Organization* (collections and models)
- *Navigation and Access* (summaries, views and variations)
- *User Interaction* (user preferences and usage history)

MPEG-7 Multimedia Description Schemes (MDS) specification provide a certain number of *basic elements* that are used to provide specific data type and mathematical structures, as vectors and matrices that are very important for multimedia descriptions. The basic elements include mechanisms to link media files and to localize for example segments or regions. Moreover many of the basic elements are related to specific needs of MM content description, like time, places, persons, organizations and other additional information like textual annotations.

MPEG-7 provides also description schemes for *content management*; in details there are three principal class of information available:
- *Creation Information*, that specifies information like textual annotations or the creators, the place and date of the creation, and also provides information on how the

material has been classified into categories as subject, purpose, language and if there are other MM materials related to it.

- *Usage Information*, that specifies information related to rights even if this part is not explicitly expressed into the standard, but using links related to rights holders.
- *Media Information*, that specifies all the information related to the media compression, coding and storage format.

Regarding the *content description*, the standard also provides description schemes to describe MM content, like the structure (regions, video frames, audio segments, etc.) and the semantic (objects, events and abstract information).
It is possible to distinguish between structural and conceptual aspects. From the structural point of view, the MM data description turns around the concept of segment description scheme that represents spatial, temporal or space-temporal portions of a MM data. The segments can be organized into a hierarchical structure in order to produce an Index for searching by MM content. The segments are described by a set of descriptors representing features like colour, texture, shape, audio features and so on, also taking into account textual annotation representing the related semantic information. From the conceptual point of view, the semantic description schemes involve information related to entities like objects, events and relationships.

The standard provides also description schemes able to facilitate browsing and retrieval providing summaries, views and variations.
The *summary description schemes* can involve two possible modalities: hierarchical and sequential. In the first case, the information is organized in different levels of detail, with the course summaries near the root element and more detailed summaries further from the root. In the second case, the summaries can provide sequences of images or video frames eventually synchronized with audio.
The *view description schemes* use partitions and decompositions to represent different views of the MM content, in order to best match requirements for resolution accesses or progressive retrieval.
The *variation description schemes* allow the selection of different variation of MM document that can replace the original one, in order to satisfy different conditions or preferences.

MPEG-7 also provides description schemes for MM content organization, for organizing collections of MM contents, segments, events or objects, so that each collection can be seen as a total entity on the base of the common properties. Then, using the *Model description scheme*, it is possible to specify different models and statistics on the attribute values of the collections.

Finally, the *User Interaction description schemes* allow to describe user preferences, for example in order to match user preferences with MM contents and have a personalization of MM content accesses and presentations.

Overview of Visual Descriptors

One of the main objectives of MPEG-7 is to provide standard visual descriptions able to describe stored or streamed images or video, in order to make easy for users or applications to identify or categorize them.
The visual feature descriptors allow users or agents in applications to realize many tasks, as for example it will be possible to draw on the screen few lines and get images with similar graphics or logos, or given a set of video the standard is able to describe objects in the scene

and their relations, camera motions and get video with similar or dissimilar space-temporal relations. Moreover, given video content, the action descriptions make possible getting other video in which similar actions happen.
The MPEG-7 visual descriptors, as shown in Figure 1, can be organized into two class, a "*General class*", including colour, texture, shape and motion features, and a "*Domain specific class*", including application-dependent descriptors as the *face-recognition* descriptors.

General class

Colour	Texture	Shape	Motion
Colour Space *Colour Quantization* *Dominant Colour* *Colour Histogram* *Colour-structure histogram* *Compact Colour Descriptor*	*Homogeneous texture* *Texture Browsing* *Edge Histogram*	*Object Bounding box* *Contour-Based Descriptor* *Region-Based Shape descriptor*	*Camera motion* *Motion Trajectory* *Parametric Object motion* *Motion Activity*

Domain specific class
Face-recognition descriptors

Figure 1 Organization of visual descriptor in MPEG-7

In more details, from the general class the main descriptors are based on [1]:

- *Colour*: that is one of the widely used visual descriptor for image and video retrieval applications. There are six colour descriptors representing different colour features including colour distribution, spatial structural properties of colour. In details:
    - *Colour Space:* four colour spaces are defined (RGB, YcrCb, HSV, HMMD). Alternatively one can specify an arbitrary linear transformation matrix from RGB coordinate.
    - *Colour Quantization:* this descriptor is used to specify the quantization method which can be linear, non linear (in MPEG-7 uniform-quantization is referred as linear quantization and non-uniform quantiser as non-linear).
    - *Dominant Colour:* this feature describes the dominant colours in the underlying segment, including the number of dominant colours, a confidence measure on the calculated dominant colours, and for each dominant colour the value of each colour component and its percentage.
    - *Colour Histogram:* several types of histograms can be specified: *The common colour histogram,* which includes the percentage of each quantized colour among all pixels in a segment or in a region. *The GoF/GoP histogram,* which can be the average, median or intersection of conventional histograms over a group of frames or pictures.
    - *Colour-structure histogram,* which is indented to capture some spatial coherence of pixels with the same colour.
    - *Compact Colour Descriptor:* instead to specify the entire colour histogram, it will be possible to specify the first two coefficients of the Haar transform of the colour histogram.

  - *Colour Layout:* this is used to describe in a coarse level the colour pattern of the image. So an image can be reduced to an 8x8 blocks with each block represented by its dominant colour.
- *Texture*: that is related to the property of homogeneity on not of the colour in an image, resulting from the presence of multiple colours or intensities in the image. It is very useful for matching and retrieval applications. In details:
  - *Homogeneous texture:* this is used to specify the energy distribution in different orientations and frequency bands (scales). This can be obtained using Gabor transform with six orientation zones and five scale bands.
  - *Texture Browsing:* this descriptor specifies the texture appearance in terms of regularity, coarseness and directionality.
  - *Edge Histogram:* it is used to describe the edge orientation distribution in an image. Three types of edge histograms can be specified, each with five entries, describing the percentages of directional edges in four possible orientations and non-directional edges. The global edge histogram is accumulated over every pixel in an image; the local histogram consists of 16 sub-histograms, one for each block in an image; the semi-global histogram consists of eight sub-histograms, one for each group of rows and columns in an image.
- *Shape*: that offers an useful description for similarity matching applications, providing contour and region descriptors for 2D or 3D virtual objects. Moreover, the standard provides a 3D shape descriptor that is very useful for geometric transformation invariant 3D shape recognition. In details:
  - *Object Bounding box:* this descriptor specifies the rectangular box enclosing two- or three-dimensional object. In addition to the size, centre and orientation of the box, the occupancy of the object in the box is also specified by the ratio of the object area (volume) to the box area.
  - *Contour-Based Descriptor:* this descriptor is applicable to a 2-D region with a closed boundary. MPEG-7 has chosen the use of the peaks in the curvature scale space representation to describe a boundary, which has been found to reflect human perception of shape.
  - *Region-Based Shape descriptor*: it can be used to describe the shape of any 2-D region, which may consist of several disconnected sub-regions. MPEG-7 has chosen to use the Zernike moments to describe the geometry of a region. The number of moments and the value of each of them are specified.
- *Motion*: that assembles colour, texture, and shape of objects to index images into a video sequence. The main features captured by the standard descriptors are represented by the *camera motion* and *object motion* descriptors. In details:
  - *Camera motion:* seven possible camera motions are considered: panning, tracking (horizontal translation), tilting, booming (vertical translation), zooming, translation along the optical axis and rolling (rotation around the optical axis). For each motion two moving directions are possible. For each motion type and direction the presence (i.e., duration), speed and the amount of motion are specified.
  - *Motion Trajectory:* it is used to specify the trajectory of the non rigid moving object in terms of 2D or 3D coordinates of certain key point selected. For each key point the trajectory between adjacent sampling times is interpolated by a specific interpolation function (either linear or parabolic).
  - *Parametric Object motion:* this is used to specify the 2D motion of rigid objects. Five types of motion are included: translation, rotation/scaling, affine, planar prospective and parabolic. In addition, the coordinate origin and time duration need to be specified.

- *Motion Activity:* it is used to describe the intensity and the spread of activity on a video segment. Four attributes are considered: intensity activity, measured by the standard deviation of the motion vector magnitudes; direction of activity, determined from the average of the motion vector directions; spatial distribution of activity, derived from the run lengths of blocks with motion magnitudes; the temporal distribution of activity, described by the histogram of the quantized activity levels over individual frame in a shot.

Regarding the domain specific class, the main descriptor is the *"face-recognition descriptor"*. The human face perception and recognition is one of the most relevant topics in computer vision community. There are al lot of applications in which the face recognition is desirable, for example video database retrieval, security authentication, advanced surveillance applications, and so on.
A lot of techniques and methodologies dealing with the problem of face recognition have been proposed in recent years, among these the Principal Component Analysis (PCA) technique. The MPEG-7 face recognition descriptor is based on PCA. In details, this descriptor represents the projections of the face vector into a set of 48 basis vectors that cover the set of all the possible face vectors. These basis vectors are derived from the eigenvectors of a set of training faces considered robust enough to the view angle and light changes. Then, the similarity measure of face images is obtained through the Euclidean distance among the projections of the face vectors.

During the standardize process of MPEG-7, experiments were done in order to optimise the adopted methods and techniques. In particular, for visual descriptors, the *Query By Example* (QBE) paradigm was used in order to evaluate the goodness and the expressiveness of a specific descriptor in the retrieval process, first extracting the descriptor values from the query image and then comparing them with the values of the same reference descriptors in the database.

Overview of Audio Descriptors

MPEG-7 Audio provides structures for describing audio content, utilizing structures that are a set of low-level descriptors, for audio features that are across many applications (e.g., spectral, parametric, and temporal features of a signal), and high-level description tools that are more specific to a set of applications. Those high-level tools include the audio signature Description Scheme, musical instrument timbre Description Schemes, the melody Description Tools to aid query-by-humming, general sound recognition and indexing Description Tools, and spoken content Description Tools.
This part of the standard is oriented to a particular set of applications such as search and retrieval of audio and speech signals (Query by Humming, Query for Spoken Content, etc.).
Figure 2 shows a concise architecture of the audio descriptor organization in the standard.
Among the *low-level audio descriptors*, we can distinguish [1][14]:

- <u>*Basic descriptors*</u>: essentially two basic descriptors are available. The first, the *"AudioWaveformType"*, is usually used to describe the minimum and the maximum amplitude of the audio signal in the sampled period and gives to users an useful model of the envelope of the time domain signal. The second, the *"AudioPowerType"* that gives a temporally homogeneous measure of the signal content as a function of time. The AudioPowerType descriptor can be used in conjunction with other basic spectral descriptors in order to have an economic description of the signal behaviour taking into account both time and frequency.

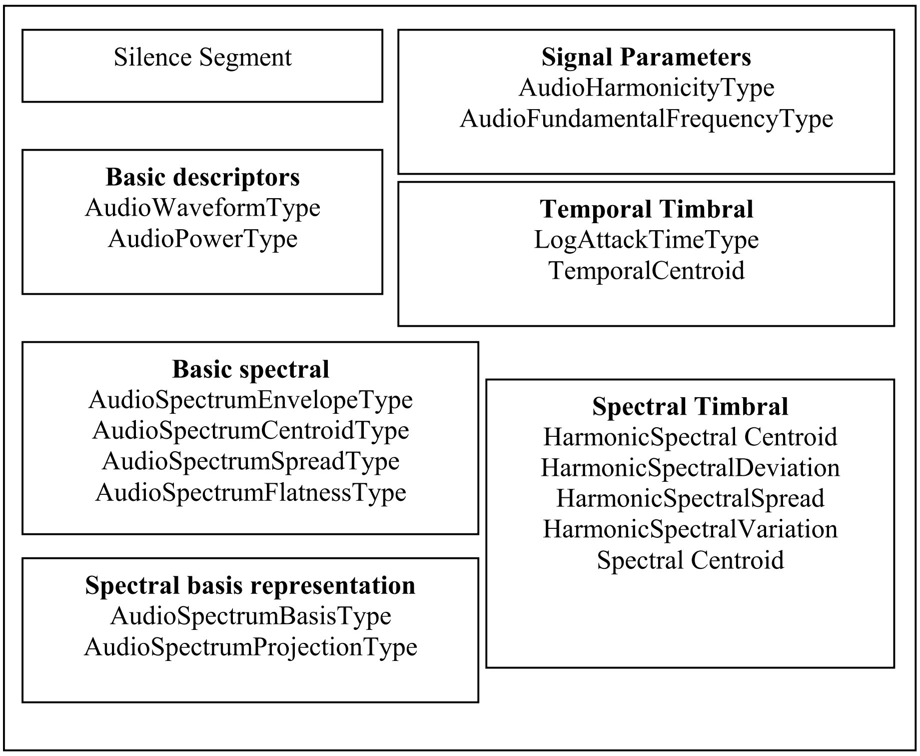

Figure 2 Organisation of the low-level audio descriptors in MPEG-7

- *Basic spectral descriptors*: the principal descriptor of this class is the *"AudioSpectrumEnvelopeType"*, tha6t can be computed from a simple power spectrum based on the Fast Fourier Transform (FFT) of the frame of audio signal samples. There are also a series of spectral feature descriptors that provide an estimation of the spectral content of the signal. These descriptors, listed in the Figure 2, are *"AudioSpectrumCentroidType"* that describes the centre of gravity of the AudioSpectrumEnvelope, the *"AudioSpectrumSpreadType"* that complements the previous Descriptor by describing the second moment of the log-frequency power spectrum, indicating whether the power spectrum is cantered near the spectral centroid, or spread out over the spectrum and the *"AudioSpectrumFlatness"* that describes the flatness properties of the spectrum of an audio signal for each of a number of frequency bands.
- *Basic signal parameters:* The two signal parameter descriptors are mainly applied to periodic or quasi-periodic signals. The "*AudioFundamentalFrequency"* descriptor describes the fundamental frequency of an audio signal. The representation of this descriptor allows for a confidence measure in recognition of the fact that the various extraction methods, commonly called "pitch-tracking," are not perfectly accurate, and in recognition of the fact that there may be sections of a signal (e.g., noise) for which no fundamental frequency may be extracted. The *"AudioHarmonicity"* descriptor represents the harmonicity of a signal, allowing distinction between sounds with a harmonic spectrum (e.g., musical tones or voiced speech), sounds with an inharmonic spectrum (e.g., metallic or bell-like sounds) and sounds with a non-harmonic spectrum (e.g., noise, unvoiced speech, or dense mixtures of instruments).
- *Temporal timbre descriptors:* The two timbral temporal Descriptors describe temporal characteristics of segments of sounds, and are especially useful for the description of musical timbre. Because a single scalar value is used to represent the evolution of a sound or an audio segment in time, these Descriptors are not applicable for use with

the Scalable Series. The *"LogAttackTime"* descriptor characterizes the "attack" of a sound, the time it takes for the signal to rise from silence to the maximum amplitude. This feature signifies the difference between a sudden and a smooth sound. The *"TemporalCentroid"* descriptor also characterizes the signal envelope, representing where in time the energy of a signal is focused.

- Spectral timbre descriptors: the five timbral spectral Descriptors, shown in Figure 2, are spectral features in a linear-frequency space especially applicable to the perception of musical timbre. The *"SpectralCentroid"* descriptor is the power-weighted average of the frequency of the bins in the linear power spectrum. The remaining four descriptors operate on the harmonic regularly-spaced components of signals. The *"HarmonicSpectralCentroid"* is the amplitude-weighted mean of the harmonic peaks of the spectrum. The *"HarmonicSpectralDeviation"* Descriptor indicates the spectral deviation of log-amplitude components from a global spectral envelope. The *"HarmonicSpectralSpread"* describes the amplitude-weighted standard deviation of the harmonic peaks of the spectrum and the *"HarmonicSpectralVariation"* descriptor is the normalized correlation between the amplitude of the harmonic peaks between two subsequent time-slices of the signal.
- Spectral basis representations: the two spectral basis descriptors represent low-dimensional projections of a high-dimensional spectral space to aid compactness and recognition. The *"AudioSpectrumBasis"* descriptor is a series of basis functions that are derived from the singular value decomposition of a normalized power spectrum. The *"AudioSpectrumProjection"* descriptor represents low-dimensional features of a spectrum after projection upon a reduced rank basis.
- Silence segment: this descriptor simply attaches the simple semantic of "silence" (i.e. no significant sound) to an Audio Segment. Although it is extremely simple, it is a very effective descriptor. It may be used to aid further segmentation of the audio stream, or as a hint not to process a segment.

Among the *high-level description tools*, the standard considers [1][14]:

- *General sound recognition and indexing tools*: the general sound recognition and indexing Description Tools are a collection of tools for indexing and categorization of general sounds, with immediate application to sound effects.
- *Spoken content description tools*: the spoken content Description Tools allow detailed description of words spoken within an audio stream. The Spoken Content Description Tools are divided into two broad functional units: the *"SpokenContentLattice"* description scheme, which represents the actual decoding produced by an Automatic Speech Recognition (ASR) engine, and the *"SpokenContentHeader"*, which contains information about the speakers being recognized and the recogniser itself.
- *Musical instrument timbre description tools*: the aim of the these description tools is to describe the perceptual features related to notions such as "attack", "brightness" or "richness" of a sound. Among four possible classes of musical instrument sounds, two classes are well detailed, and had been the subject of extensive development within MPEG-7. Harmonic, coherent, sustained sounds, and non-sustained, percussive sounds are represented in the standard.
- *Melody description tools*: these description tools include a rich representation for monophonic melodic information to facilitate efficient, robust, and expressive melodic similarity matching. The Melody Description Scheme includes a *"MelodyContour"* description scheme for efficient melody contour representation, and a *"MelodySequence"* description scheme for a more complete and expressive melody representation.

Content Organization

The evolution of MM systems and technologies can be seen as a process that leave the perspective of dealing with MM data as signals to meet a new point of view, in which the building block is the MM data feature extraction and processing in order to facilitate the MM data understanding at a semantic and knowledge representation level [1][14].
MPEG-7 allows content-based retrieval of multimedia extracting and searching several features through a similarity measures. An important aspect is the labelling of MM content at a semantic level, allowing the content to be easier to search, filter, summarize and so on. Then, semantic labels allow the extraction of knowledge through MM metadata-mining applications.
The gap between MM data and knowledge extracted from them is covered by *Models*, which provides parametric and statistical descriptions of MM information. For example, given a set of images representing the same content, the colour features of the images can be used to form a *Model* of the image set, by extracting the colour histograms from the images and computing the centroid colour feature vector. Then, the colour histogram model can be used to semantically classify new images comparing the histogram colour vector of the new images with the centroid of the colour histogram vector of the semantic classes.
Given the importance of semantic labels for MM content descriptions, MPEG-7 provides a set of standardized tools able to represent the key concepts of *Collections* and *Models*, in the follow way [1]:

- *Collection*: provides sets of MM content, including descriptor instances, segments, concepts or mixed sets of them. Collection can be used to describe for example albums of songs, or more in general sets of objects, or at a lower level colour feature clusters. In particular, the *content collection* describes a collection of MM documents, while *segment collection* describes collections of MM data segments like image regions or video segments. The *descriptor collection* describes collections of MM content descriptions like Dominant Colour Descriptor and other MPEG-7 descriptors. The *concept collection* describes semantic concepts like events, objects and relationships among them. Finally, the *mixed collection* describes collection of content, descriptions and concepts, while the *structured collection* describes relationships among collections.
- *Model*: provides representations of instances or classes of MM content, descriptors or collections. A Model is organized in the following way:
    - *Probability Model*: that describes statistics or probabilities associated with attributes of MM content, descriptors or collections.
    - *Analytic Model*: that associates labels and semantics to MM content, descriptors or collections.
    - *Cluster Model*: that describes the association between the probability and the analytic models, associating statistics, probabilities, labels and semantics with MM content collections.
    - *Classification Model*: that provides information on known collections of MM content (labels, semantics and models), in order to classify new MM content.

So, MPEG-7 provides a way to describe MM content organization providing a description of both low-level and semantic type from an upper level perspective, allowing as example applications to classify positive and negative samples of images or to model video sequences taking into account temporal relations among them.

Description of a MM document

MM document supported by MPEG-7 is not only audiovisual documents such as images, videos, audio sequences, but also documents of a different type like web pages containing audio, images, text and other different MM data at the same time, putted together using specific tools for MM synchronized presentations (e.g. SMIL). So, the MPEG-7 tools are also applicable to this type of MM documents. In particular, the standard provides tools for management, organization and semantics of a single MM document. The *management tools* describe the creation, usage properties of content, while *structure tools* describe spatial or temporal segments of MM contents, their attributes and relations among them. Finally *semantic tools* describe semantic entities extractable from MM content and their attributes and relations.

In more details, the *Management tools* include three basic classes of description schemes [1]:

- Media Information: that is composed by the following descriptors
    - The *MediaFormat* Descriptor, that describes the coding parameters of the media, for example the file format, size, coding scheme and other parameters describing frames, pixels and so on.
    - The *MediaInstance* description scheme
    - The *MediaTranscodingHints* Descriptor
    - The *MediaQuality* Descriptor
- Content Creation, that is composed by the follow description schemes:
    - The *CreationDescription* Scheme,
    - The *Classification* Description scheme,
    - The *RelatedMaterial* Description Scheme
- Content Usage, which is composed by the following description schemes:
    - The *Rights Data Type*,
    - The *Financial data Type*,
    - The *Availability* Description Scheme,
    - The *Usage Record* Description scheme

About the *Content Structure tools*, they describe the spatial and temporal relations among the interrelated segments in the structure of multimedia content. In particular they describe the segments, their attributes, the structural relations among them as their hierarchical decomposition. In more details, the content structure tools provide instruments to express:

- *Segment Entities*: that describes space-temporal segments of generic types of MM content, such as images, videos, audio and audio-video sequences. Segment entities could be still-images describing spatial region of 2D images or video-frame or temporal interval of an audio sequence. MPEG-7 also provides a set of specific tools for describing 3D spatial region of 3D images, or space-temporal segment of ink content, created by a pen based system or an electronic white board.
- *Segment Attributes*: that can be used to describe segment properties such as media, creation and usage information, visual and audio features, but also considering other specific features like characterization of the connected components of the segment or the segment importance, while other segment attributes specify *HandWritingRecognition* information related to ink segments. Audio and visual descriptors can be also used to describe specific characteristics of segments.
- *Segment Decomposition*: that is related to the possibility to decompose segments in order to form a hierarchy to be able to successively define efficient global or local searching strategies. The main decomposition tools allow to decompose segments following space-temporal criteria, for example an image can be spatially divided into still regions, corresponding to the objects in the scene, that can be at the same time decomposed into other segments corresponding to different region of the object itself.

It is also possible to perform a decomposition of segments into a set of segments of different nature (e.g. a video segment can be divided into other video segments or still or moving regions).

- *Structural Relation*: that supports other types of decomposition. In fact, the hierarchical structure provided by segment decomposition tools can be not appropriate in certain context, so it can be possible to define different type of structural relations. MPEG-7 provides different Classification Schemes (CS) such as *SpatialRelation CS* or *TemporalRelatio CS* and *GraphRelation CS* that allow respectively to describe how the entities are placed and related to each other in the 2D space, how entities are related in different temporal intervals and other semantic relations among segments.

About the *Semantic Tools*, they are able to express semantics of MM contents allowing the description of *semantic entities*, *attributes* and *relations*. The standard allows the description of abstractions, which refer to the process of extracting semantic description from an instance of MM content and generalising it to multiple instances of MM content (*Media Abstraction*), for example "Ms A and Mr B are talking" for any picture or video sequences in which can be decomposed the same MM data, or to a set of concrete semantic descriptions (*Formal Abstraction*), through the description of a pattern that is common to a set of examples, e.g. "Ms A and Mr B are talking" or "Any women is talking with any man".
In more details, semantic tools are composed of the following specific elements:

- *Semantic Entities*: describing semantic entities such as objects, events, concepts state, places, times and *narrative world*. A narrative world is represented by the *Semantic Description Scheme*, which is described by a certain number of semantic entities and graphs of their relations. Other description schemes derived from the Semantic Description Scheme are *Object*, *AgentObject*, *Event*, *SremanticPlace*, *SemanticTime*, *SemanticState* and *Concept*, which can be used to describe entities that populate the narrative world.
- *Semantic Attributes*: that allows to complementary describe semantic entities through the definition of *Labels*, *Textual Definitions*, *Properties* and *Features* of the segments in which they occur. In particular, a semantic entity can have multiple associated labels that allow a fast retrieval of all the semantic entities belonging to the same abstract class defined by a certain label. The Textual Definition can be defined by both free and structured text.
- *Semantic Relations*: that allows describing the MM data semantics providing instruments to express both *Normative* and *Non Normative* semantic relations. The first class includes relation types such as *agent*, *agentOf*, *patient*, *patientOf*, *cause*, *causeOf*, *result*, *resultOf*, describing how several semantic entities are related in a narrative or story, or such as *combination*, *specialise*, *generalise*, *component*, *componentOf*, describing haw the definitions of several entities are related to each other, or finally *location*, *locationOf*, *source*, *sourceOf*, *destination*, *destinationOf*, and so on, describing spatial, temporal localisation of semantic entities into segments, models, collections or other semantic entities. Finally, structural relation tools like *GraphRelation*, *SpatialRelation* and *TemporalRelation CS* can be used to describe semantic relations among semantic entities.

The binary description stream format (BiM)

The XML language has not been designed to deal ideally in a real-time constrained and streamed environment like in the multimedia or mobile industry.

MPEG-7 uses XML to model this rich and structured data and to overcome the lack of efficiency of textual XML, has defined a generic framework to facilitate the processing of MPEG-7 descriptions: the BiM (Binary description stream Format for MPEG-7) format. Actually there are two possible ways to transmit MPEG-7 descriptions, in textual (TeM, staying for Textual Format for MPEG-7) or binary (BiM) fomat, but the last one provides an additional feature consisting in the compression of the verbose XML text.

BiM coders and decoders can deal with any XML language. Technically, the schema definition (DTD or XML Schema) of the XML document is processed and used to generate a binary format that has three main properties [1][14]:

- Is connected to Schema knowledge
- Structural redundancy (element name, attribute names, and so on) is removed from the document, so the document structure is highly compressed (98% in average).
- Unlike a zipped XML document, a BiM file can be processed directly at binary level, allowing fast parser and filtering.

Using the BiM format, each document can be transmitted in one or more pieces. At the lowest level of granularity, each attribute value or document leaf can be modified to allow a minimal transmission in case of a minimal change in the sent document.
For instance, the streaming capability of BiM enables to cut a large XML document in many pieces. These pieces can be delivered separately to the client. It is not required for the decoder to download (and keep in memory) the entire XML file before being capable of processing it. It can reduce both memory required at terminal side and consumed bandwidth. It improves overall quality of service and response time of XML based services.
In a streamed environment, a single XML document can be maintained at terminal side by constantly refreshing its sub parts. Different refresh rate can be imposed to different sub part of the same XML document. This updating capability allows to finely manage consumed bandwidth and to provide the best quality of service for the minimal bandwidth consumption.

Unlike a zipped XML document, a BiM file can be processed directly at binary level. Moreover, document sub-trees can be skipped improving the overall performance of an application dealing with large documents. This optional “skipping” process can be triggered on the basis of element names, types or attribute values. This feature considerably improves browsing and searching through large XML files.

So, in conclusion we can say that the binary description stream format BiM provides a flexible and efficient instrument for compression and streaming of MPEG-7 documents.
From one hand it allows to have compact representation of the descriptions compared to the XML format, while on the other hand it allows the incremental transmission and the dynamic update of the descriptions.
The format has been defined in order to allow fast search and filtering operations on binary level, without the need to decompress the incoming description stream. It can be also applied to general XML files, if the XML schema and the respective XML schema definition are available.

Chapter 3

The Semantic Web and interoperability tools: a brief overview

According to the World Wide Web Consortium (W3C) [33], *"The Semantic Web is an extension of the current web in which information is given well-defined meaning, better enabling computers and people to work in cooperation"*. It provides a common framework that allows data to be shared and reused across different applications, enterprises, and communities.

The Semantic Web has two main motivations [16]:

- The first is *data integration*, which is a significant bottleneck in many IT applications. Current solutions to this problem are mostly ad hoc. In fact, each time a specific mapping is made between the data models (schemas) of the data sources involved. If the semantics of data sources were described in a machine-interpretable way, the mappings could be constructed at least in semi-automatic way.
- The second motivation is *more intelligent support for end users*. If computer programs can infer consequences of information on the web, they can give better support in finding information, selecting information sources, personalizing information, combining information from different sources, and so on.

In order to reach the interoperability needs, the main problem related to the use of the *standardised metadata description schemes* is the fixed number of describing attributes and their specialisation on fixed domains with the consequent lesser flexibility in representing needs of particular application domains.
In the light of these difficulties another standard class has been evolved, considering the possibility of integrating more metadata standards mapped on different application domains, providing rich metadata models for media descriptions together with languages allowing one to define other description schemes for arbitrary domains.
These last standards can be referred as *standardised metadata frameworks*. They promised simpler interoperability among different application domains and also the possibility to simpler create generic tools able to process multimedia descriptions.

We can summarise the main characteristics of the standardised metadata frameworks in the following points:

- No domain-dependence description scheme.
- Rich generic data-model for multimedia content description.
- Schema definition languages that allow defining description schemes for arbitrary application domains.

A list of other relevant interoperability tools and their specific references have been provided in the first deliverable of this WP [15], some of which are listed in the follow:

- Platform for Internet Content Selection (PICS) [23]
- Meta Content Framework (MCF) [24]
- XML Topic Maps (XTM) [25]

During the research phase particular attention has been given to one of the latest initiative following this tendency that is represented by the MPEG-21 [26], a new MPEG metadata standard framework.

Moreover, the W3C has defined such open standards for metadata syntax as the Resource Description Framework (RDF) [20] and the Web Ontology Language (OWL) [21], and support for these standards from both industry and academia is rapidly increasing.

Professional groups increasingly are building metadata vocabularies (or *ontologies)*. Large ontologies just exist for medical terminology, genomic, geographic information systems, and law, just to mention a few. These terminologies are typically hand built, but systems are rapidly getting better at learning them in a semi-automatic way from large volumes of text.

An important related problem is that of automatically finding translations between different terminologies that have been designed for the same domain, the so-called *"ontology mapping problem"*.
A brief overview on XML language and RDF framework is presented in the follow.

An eXtensible Mark-up Language: XML

Metadata must be expressed in a standardized way in order to be read, searched, exchanged by computer systems, and understood by humans. It can be recorded in a table or a database, or expressed in an HTML (Hypertext Mark-up Language) or XML (eXtensible Mark-up Language) document [19].
XML is playing an increasingly important role in the exchange of a wide variety of data on the Web and elsewhere. It and its associated technologies, XML Namespaces, XML Query languages, XML Databases are enabling implementers to develop metadata schemas, application profiles, large repositories of XML metadata, and search interfaces using XML Query Language. These technologies are the key to enabling the automated computer processing, integration and exchange of information over Internet [18].
An important feature of this language is the separation of content from presentation, which makes it easier to select and/or reformat the data. However, due to the probability of numerous industry and domain specific DTDs, those who wish to integrate information will still be faced with the problem of semantic interoperability [17]. This problem is not solved by XML, but there are other frameworks, among which RDF, able to support such solution to this difficulty. In particular, while XML is able to embody several types of documents taking into account specifically syntactic aspects, RDF also allows to express semantic aspects, however it represents only a partial solution to the interoperability problem. A way to achieve better results is provided by specific semantic-rich languages appositely constructed for expressing a more complete representation of semantic entities and their relations. An example of that is the Web Ontology Language (OWL) is in [21], which represents an extension of RDF and will be examined in more details in the next chapter.

Resource Description Framework: RDF

The Semantic Web provides a common framework that allows data to be shared and reused across application, enterprise, and community boundaries. It is a collaborative effort led by W3C with participation from a large number of researchers and industrial partners. It is based on the Resource Description Framework (RDF), which integrates a variety of applications using XML for syntax and URIs for naming.

RDF is *"to specify semantics for data based on XML in a standardized interoperable manner."* [17].
Technically, RDF is not a language, but a data model of metadata instances. The basic data model is very simple; it consists of nodes connected by labelled arcs, where the nodes represent web resources and the arcs represent properties of these resources.

The development of RDF has been motivated by the following uses [20]:

- Web metadata: providing information about Web resources and the systems that use them (e.g. content rating, capability descriptions, privacy preferences, etc.)

- Applications that require open rather than constrained information models (e.g. scheduling activities, describing organizational processes, annotation of Web resources, etc.)

- To do for machine processable information (application data) what the World Wide Web has done for hypertext: to allow data to be processed outside the particular environment in which it was created, in a fashion that can work at Internet scale.

- Inter-working among applications: combining data from several applications to arrive at new information.

- Automated processing of Web information by software agents: the Web is moving from having just human-readable information to being a worldwide network of cooperating processes.

RDF is designed to represent information in a minimally constraining, flexible way. It can be used in isolated applications, where individually designed formats might be more direct and easily understood, but RDF's generality offers greater value from sharing. The value of information thus increases as it becomes accessible to more applications across the entire Internet.
RDF uses the key concepts exposed briefly in the follow [20][15]:

- *Graph data model*: the structure of any expression in RDF is a collection of triples, each consisting of a subject, a predicate and an object. A set of such triples is called an RDF graph. This can be illustrated by a node and directed-arc diagram, in which each triple is represented as a node-arc-node link, hence the term "graph". Each triple represents a statement of a relationship between the things denoted by the nodes that it links.

- *URI-based vocabulary*: A node may be a URI (Uniform Resource Identifiers) with optional fragment identifier (URI reference, or URIref), a literal, or blank. A URI reference or literal used as a node identifies what that node represents. A URI reference used as a predicate identifies a relationship between the things represented by the nodes it connects. A predicate URI reference may also be a node in the graph. In particular, a blank node is just a unique node that can be used in one or more RDF statements, but has no intrinsic name.

- *Data types*: RDF uses Data types in the representation of values such as integers, floating point numbers and dates. A data type is defined by a lexical space, a value space and a lexical-to-value mapping. RDF predefines just one data type “rdf:XMLLiteral”, used for embedding XML in RDF.

- *Literals*: literals are used to identify values such as numbers and dates by means of a lexical representation. Anything represented by a literal could also be represented by a URI, but it is often more convenient or intuitive to use literals. They can be “plain literals” meaning that the string can be combined with other language tags, or “typed literals” meaning that the string can be combined with data type URI.

- *Expression of simple facts*: some simple facts indicate a relationship between two things. Such a fact may be represented as an RDF triple in which the predicate names the relationship, and the subject and object denote the two things. Thus, a more complex fact is expressed in RDF using a conjunction (logical-AND) of simple binary relationships. RDF does not provide means to express negation (NOT) or disjunction (OR).

- *Entailment*: in brief, an RDF expression A is said to entail another RDF expression B if every possible arrangement of things in the world that makes A true also makes B true. On this basis, if the truth of A is presumed or demonstrated then the truth of B can be inferred.

An RDF document is written in XML. The XML language used by RDF is called RDF/XML. By using XML, RDF information can easily be exchanged between different types of computers using different types of operating system and application languages.

The RDF data model defines a simple model for describing interrelationships among resources in terms of named properties and values. RDF properties may be thought of as attributes of resources and they also can represent relationships between resources. As such, the RDF data model can therefore be similar to an *entity-relationship* diagram.
The RDF data model, however, provides no mechanisms for declaring these properties, it does not provide any mechanisms for defining the relationships between these properties and other resources. That is the role of *RDF Schema*.
To allow the creation of controlled, sharable, and extensible vocabularies, the RDF working group has developed the RDF Schema Specification. This specification defines a number of properties that have specific semantics.
A schema defines not only the properties of the resource (e.g. Title, Author, Subject, Size, Colour, etc.) but may also define the kinds of resources being described (e.g. books, Web pages, people, companies, etc.). It specifies the mechanisms needed to define such elements, to define the classes of resources they may be used with, to restrict possible combinations of classes and relationships, and to detect violations of those restrictions.
Since it is possible that different schemas may use the same strings to represent different conceptual meanings, RDF uses *XML namespaces* to assign a separate namespace to each schema. In RDF, schemas are extended by simply referring to objects from that schema as resources in a new schema. Since schemas are assigned to unique URIs (Uniform Resource Identifiers), the use of XML namespaces guarantees that exactly one object is being referenced.
However, RDF does not have any mechanism to define *general axioms* [17], which can allow further interpretation of the described data providing a stronger semantic representation, for example it could be interesting to specify that such property is the inverse of another. For example, general axioms can be used for the mapping process between different representations of the same concept.
In order to bridge this gap, other interoperability tools, providing mechanisms able to enrich the semantic description of data, have been developed and will be discussed in the next sections.

Chapter 4

Knowledge representation and interpretation through ontologies

An ontology consists of a set of concepts, axioms, and relationships that describes a domain of interest. An ontology is similar to a dictionary or glossary, but with greater detail and structure and expressed in a formal language (e.g., OWL) that enables computers to process its content.
Ontologies can enhance the functioning of the web to improve the accuracy of Web searches and to relate the information in a resource to the associated knowledge structures and inference rules defined in the ontology.
Ontologies can differ in respect to the scope and purpose of their content. The most prominent distinction is between the [15]:

- *domain ontologies* describing specific fields, like medicine.
- and *upper level ontologies* describing the basic concepts and relationships considered when information about any domain is expressed in natural language.

Upper ontologies provide a structure and a set of general concepts upon which domain-specific ontologies (e.g. medical, financial, engineering, sports etc.) could be constructed. An upper ontology is limited to concepts that are abstract and generic enough to address a broad range of domain areas at a high level.
Computers utilize upper ontologies for applications such as data interoperability, information search and retrieval, automated inferring, and natural language processing.
Ontologies should be publicly available and different data sources should be able to commit to the same ontology for shared meaning. Also, they should be able to extend other ontologies in order to provide additional definitions. In fact, it can be possible that an organization find that an existing ontology provides only a part of what it needs, in which case, it should not have to create a new ontology, but an ontology which extends an existing one and adds any desired identifiers and definitions.
Different ontologies may model the same concepts in different ways. The language should provide primitives for relating different representations, thus allowing data to be converted to different ontologies and enabling a "web of ontologies".

About the history of the languages realised and adopted for the expression of the web ontologies, the semantic web has followed the way that starting from RDF and crossing on other intermediate ways (DAML, OIL) has reached the place of OWL and its extensions, providing several new initiatives still evolving.
While RDF was the first language specified by the W3C for representing semantic information about arbitrary resources, RDF Schema (RDFS) was a W3C recommendation for an extension to RDF to describe RDF vocabularies. RDFS can be used to create ontologies, but it is trivial for the purpose, with less expressive power than OWL [21].
Like OWL, RDFS includes classes and properties, as well as range and domain constraints on properties. It provides inheritance hierarchies for both classes and properties. Upon its release users began requesting additional features, including data types, enumerations and the ability to define properties in a more rigorous way.
Instead of continuing with separate ontology languages for the Semantic Web, a group of researchers, got together in the Joint US/EU ad hoc Agent Markup Language Committee to create a new Web ontology language. This language DAML+OIL built on both OIL and

DAML-ONT, was submitted to the W3C as a proposed basis for OWL, and was subsequently selected as the starting point for OWL.
In addition to ontology languages, various taxonomies and existing ontologies are already commercially in use. In e-Commerce sites they facilitate machine-based communication between buyer and seller, enable vertical integration of markets and allow descriptions to be reused in different marketplaces. Examples of sites that are actually ok allmaking commercial use ontologies include VerticalNet [32] , that currently hosts 59 industry-specific e-marketplaces that span diverse industries such as manufacturing, communications, energy, and healthcare.

Various medical or drug-related ontologies have been developed to help manage the overwhelming mass of current medical and biochemical research data that can be difficult to tie together into a cohesive whole. One major resource is the Gene Ontology Consortium which is defining ontologies for Molecular Function,Biological Process, and Cellular Components.

There exist large taxonomies in use today that would be ripe for extension into the OWL space. For example, the North American Industry Classification System (NAICS) [31] defines a hierarchy of over 1900 items that identify industry types. NAICS is also tied to the International Standard Industrial Classification System (ISIC, Revision 3), developed and maintained by the United Nations.

The Web Ontology Language: OWL

The term "Ontology" is rented from philosophy and refers to the science of describing entities in the world and how they are related.
The Web Ontology Language OWL is a semantic markup language for publishing and sharing ontologies on the World Wide Web. OWL is developed as a vocabulary extension of RDF (the Resource Description Framework) and is derived from the DAML+OIL Web Ontology Language. This part of the document contains a structured informal description of the set of OWL language constructs, the full guide, that serve as a reference for OWL users who want to construct OWL ontologies is available in [21].
The OWL Web Ontology Language is intended to provide a language that can be used to describe the classes and relations between them that are inherent in Web documents and applications.
The use of the OWL language is intended to :

- formalize a domain by defining classes and properties of those classes
- define individuals and assert properties about them
- reason about these classes and individuals to the degree permitted by the formal semantics of the OWL language

The W3C's Web Ontology Working Group defines OWL as three different sub-languages [28]:

- *OWL Full*: the entire language is called OWL Full, and uses all the OWL language primitives. It also allows to combine these primitives in arbitrary ways with RDF and RDF Schema. This includes the possibility, also present in RDF, to change the meaning of the predefined (RDF or OWL) primitives, by applying the language primitives to each other. For example, in OWL Full we could impose a cardinality constraint on the class of all classes, essentially limiting the number of classes that can be described in any ontology. The advantage of OWL Full is that it is fully upward compatible with RDF, both syntactically and semantically: any legal RDF document is also a legal OWL Full document, and any valid RDF/RDF Schema conclusion is also a valid OWL Full conclusion.

- *OWL DL*: in order to reach computational efficiency, OWL DL (OWL Description Logic) is a sub-language of OWL Full which restricts the way in which the constructors from OWL and RDF can be used, ensuring that the language corresponds to a well studied description logic. The advantage of this is that it permits efficient reasoning support. The disadvantage is that we loose full compatibility with RDF: an RDF document will in general have to be extended in some ways and restricted in others before it is a legal OWL DL document, while every legal OWL DL document is still a legal RDF document.
- *OWL Lite*: it consists in an ever further restriction limits OWL DL to a subset of the language constructors. For example, OWL Lite excludes enumerated classes, disjunction statements and arbitrary cardinality (among others). The advantage of this is a language that is both easier to control (for users) and easier to implement (for tool builders). The disadvantage is of course a restricted expressiveness.

Most of the elements of an OWL ontology concern classes, properties, instances of classes and relationships between these instances.
The essential language components can be briefly summarized in the following classes [21] [28]:

- *Simple Classes and Individuals*: there is an important distinction between a class and an individual in OWL. A class is simply a name and collection of properties that describe a set of individuals. Individuals are the members of those sets. So, classes correspond to the sets of things in a domain of discourse, while individuals correspond to concrete entities that can be grouped into these classes. In particular:
    - *Simple Named Classes*: every entity (individual) in the OWL world is a member of the class owl:Thing. Thus each user-defined class is implicitly a subclass of owl:Thing.
    - *Individuals*: in addition to the classes definition, it is also possible to describe individuals as instances of a certain class in the domain of interest .
- *Simple Properties*: properties allow to assert general facts about the members of classes and specific facts about individuals. A property is a binary relation and can be distinguished into two categories:
    - *Data type properties*: relations between instances of classes and RDF literals and XML Schema datatypes
    - *Object properties*: relations between instances of two classes.

  Defying properties, there is a certain number of restrictions applicable on them. For example the domain and the range can be defined, or a property can be specified as a specialization of another property.
- *Property Characteristics*: consist in mechanisms used to further specify properties. 'Property Characteristics' provides a powerful mechanism for enhanced reasoning about a property. A list of property characteristics provided by the language is presented in the follow:
    - *Transitive Property*: if a property, P, is specified as transitive then for any x, y, and z, P(x,y) and P(y,z) implies P(x,z)
    - *Symmetric Property*: if a property, P, is tagged as symmetric then for any x and y, P(x,y) iff P(y,x)
    - *Functional Property*: if a property, P, is tagged as functional then for all x, y, and z, P(x,y) and P(x,z) implies y = z

    - *InverseOf Pproperty*: if a property, P1, is tagged as the "owl:inverseOf P2", then for all x and y, P1(x,y) iff P2(y,x)
    - *Inverse Functional Property*: if a property, P, is tagged as InverseFunctional then for all x, y and z, P(y,x) and P(z,x) implies y = z
- *Property Restrictions*: in addition to the property characteristics, it is also possible to restrict the range of a property into some specific context.
    - *allValuesFrom:* The "owl:allValuesFrom" restriction requires that for every instance of the class that has instances of the specified property, the values of the property are related to all members of the class indicated by the "owl:allValuesFrom" section.
    - *somValuesFrom:* the meaning of this restriction is similar to the previous one with the difference that the previous formulation consider an universal quantification, while "owl: *somValuesFrom*" is based on an existential quantification.
    - *Cardinality:* the restriction "owl:cardinality" allows to specify exactly the number of elements in a relation. Cardinality expressions with values limited to 0 or 1 are part of OWL Lite. This permits the user to indicate 'at least one', 'no more than one', and 'exactly one'. Positive integer values are also permitted in OWL DL. The constraint "owl:maxCardinality" can be used to specify an upper bound, while "owl:minCardinality" can be used to specify a lower bound. These restrictions can be used together to express a numeric interval.
    - *hasValue:* this restriction allows to specify classes based on the existence of particular property values. For example, an individual will be a member of such a class with at least one of its property values equal to the "hasValue" resource.

OWL also provides additional constructors in order to form more complex classes. These constructors can be used to create so-called *class expressions*. OWL supports the basic set operations, namely *union*, *intersection* and *complement* that selects all individuals from the domain of discourse that do not belong to a certain class. These are named "owl:unionOf", "owl:intersectionOf", and "owl:complementOf", respectively.
Additionally, classes can be enumerated. In particular, OWL provides the means to specify a class via a direct enumeration of its members. This is done using the "owl:oneof" construct. Notably, this definition completely specifies the class extension, so that no other individuals can be declared to belong to the class.
The disjunction of a set of classes can be expressed using the "owl:disjointWith" constructor. It guarantees that an individual that is a member of one class cannot simultaneously be an instance of a specified other class.

Ontologies will change over time, so they can be versioned. Through an "owl:Ontology" element, it is possible to link to a previous version of the ontology being defined. The "owl:priorVersion" property is intended to provide this link, and can be used to track the version history of an ontology. It could be possible that different versions of an ontology could not be compatible, so the tags "owl:backwardCompatibleWith" and "owl:incompatibleWith" cab be used to indicate compatibility or the lack with the previous ontology versions.

The Rule Languages

Many of the limitations of OWL come from the fact that, while the language includes a relatively rich set of class constructors, as we have seen in the previous section, the language provided for talking about *properties* is weaker. In particular, there is no composition constructor, so it is impossible to capture relationships between a composite property and another.
One way to do this can be to extend OWL DL in order to increase the language power to describe properties, but in this way for *decidability* reasons the property composition should be limited, that is not all the relationships between composed properties could be captured. Another way to overcome this OWL lack of expressiveness can be obtained extending OWL with a set of “rule languages” .

Interesting initiatives and their references are listed below [27]:

- *RuleML (Rule Markup Language)* [29] : The Rule Markup Initiative is working towards an XML-based markup language that permits web-based rule storage, interchange and retrieval applications. The RuleML Initiative has taken steps towards defining a shared Rule Markup Language (RuleML), permitting both forward (bottom-up) and backward (top-down) rules in XML for deduction, rewriting, and further inferential-transformational tasks. RuleML covers a hierarchy of rules, including *reaction rules* (event-condition-action rules), *transformation rules* (functional-equational rules), *derivation rules* (implicational-inference rules), also specialized to facts ('premiseless' derivation rules) and *queries* ('conclusionless' derivation rules), as well as *integrity-constraints* (consistency-maintenance rules).
- *SWRL (Semantic Web Rule Language)* [30]: The proposal extends the set of OWL axioms to include Horn-like rules. It thus enables Horn-like rules to be combined with an OWL knowledge base. The proposed rules are of the form of an implication between an antecedent (body) and consequent (head). The intended meaning can be read as: whenever the conditions specified in the antecedent hold, then the conditions specified in the consequent must also hold. Both the antecedent (body) and consequent (head) consist of zero or more atoms. An empty antecedent is treated as trivially true ad an empty consequent is treated as trivially false. Moreover rules with conjunctive consequents could easily be transformed into multiple rules each with an atomic consequent. Atoms in these rules can be of the form *C(x)*, *P(x,y)*, *sameAs(x,y)* or *differentFrom(x,y)*, where *C* is an OWL description, *P* is an OWL property, and *x,y* are either variables, OWL individuals or OWL data values. An XML syntax is also given for these rules based on RuleML and the OWL XML presentation syntax. Moreover, an RDF concrete syntax based on the OWL RDF/XML exchange syntax has been presented
- *ORL (OWL Rules Language)* [27]: the proposal adds a simple form of Horn-style rules to OWL. In ORL, rules are syntactically and semantically coherent with the ontology language, being the basic idea to add Horn rules as a new type of axiom in OWL DL with similar semantics to OWL *subClassOf* axioms. In particular:
  - The basic syntax is an extension of the abstract syntax for OWL DL and OWL Lite.
  - ORL rules XML syntax is based on the OWL XML presentation syntax
  - ORL rules mapping to RDF graphs is based on the OWL RDF/XML exchange syntax
  - ORL semantics extend the OWL DL model-theoretic semantics.

  It extends OWL axioms by adding the production:

  axiom ::= rule

rule ::= 'Implies(' {annotation} antecedent consequent ')'
antecedent ::= 'Antecedent('{atom}')'
consequent ::= 'Consequent('{atom}')'

So, ORL extends OWL with the basic Horn rules, in which predicates can be formed by OWL classes or properties, but the expressiveness power is also limited, for example it is not possible to express arithmetic predicates, taking into account arithmetic relationships between data values, that is a very desirable feature in a lot of applications.

Ontology mapping

In order to have the maximum impact for ontologies, they need to be widely shared. In order to minimize the intellectual effort involved in developing an ontology they need to be re-used. In the best of all possible worlds they need to be composed. For example, you might adopt a date ontology from one source and a physical location ontology from another and then extend the notion of location to include the time period during which it holds.

It is possible to distinguish different interesting aspects that have to be taken into account in order to reach these goals. In particular [21]:

- *Equivalence between classes and properties*: it is frequently useful to be able to indicate that a particular class or property in one ontology is equivalent to a class or property in a second ontology. This capability must be used with care. If the combined ontologies are contradictory (all A's are B's vs. all A's are not B's) there will be no extension (no individuals and relations) that satisfies the resulting combination. The property "owl:equivalentClass" is used to indicate that two classes have precisely the same instances. Note that in OWL DL, classes simply denote sets of individuals, and are not individuals themselves. In OWL Full, however, we can use "owl:sameAs" between two classes to indicate that they are identical in every way.
- *Identity between Individuals:* there is also a mechanism similar to that for classes, but declares two individuals to be identical using the OWL constructor "owl:sameAs". In fact, OWL does not have a unique name assumption. Just because two names are different, it does not mean that they refer to different individuals.
- *Different Individuals:* otherwise, the mechanisms "differentFrom" and "AllDifferent" provide the opposite effect from "sameAs". In fact, there will be cases where it is important to ensure such distinct identities.

In the context of MM data representation and exchange, the importance and the relevance of the possibilities provided by these constructors represents a focal point along the potential way, adopted by several current initiatives, that consider the construction of an upper ontology able to represent different domain specific ontologies build upon the promising selected standards.

Chapter 5

Conclusions

A review of metadata standards and tools useful to represent MM content is presented in this document in order to point out their context and motivation and their specific characteristics fitting the NoE's MM data representation needs and provide guidelines and a rich reference list to use them.
In the light of the existing state of the art, inside and outside the NoE, it has been pointed out how the interoperability needs of the NoE can be covered by some of these standards.
In particular, two possible crossable ways have been identify and highlighted. From one hand, it could be useful to consider the possibility to identify and adopt only a specific standard able to represent all the opportune information type. From the other hand, in line with other current initiatives, we could think to the definition of a larger and more comprehensive representation model, which can be obtained through the integration of different metadata standards forming an upper-level ontology and, at the same time, able to grant interoperability across the networks through the definition of specific semantic mapping procedures.

Following the aforementioned strategy lines, the report followed a structure that starts from the detailed description of the MPEG-7 standard, which can be considered as a bridge between the field of the "static" *MM standardized description schemes* (e.g. DC, MARC, LOM, SMPTE, etc.) defined for specific domains, and the field of the *standardized description frameworks* (standard tools, e.g. RDF, OWL, etc) able to integrate different standards and also define graphs allowing the representation of semantic relations. The report also tracks the flow that comes from the overview of XML and RDF interoperability tools and goes towards OWL and the definition of an upper ontology model.

On the base of the currently acquired information from the NoE, we can better sustain a strategy that considers the MPEG-7 as a good base to start from for the definition of an eventually more complete description model, taking into account other metadata standard covering MM data aspects that are not tractable using only MPEG-7. This is the reason why, in this deliverable, the guidelines mostly focused on MPEG-7, also providing references for available tools able to help in MPEG-7 descriptor extraction procedures.

Appendix A

List of possible additive standards for MM data representation and exchange

1. MARC Standard (http://www.loc.gov/marc/)
2. Dublin Core Metadata Element Set, Version 1.1, 2 July, 1999. http://www.purl.org/dc/documents/rec-dces-19990702.htm
3. CDWA Standard (http://www.getty.edu/research/conducting_research/standards/cdwa/)
4. VRA Core (http://www.vraweb.org/vracore3.htm)
5. Content Standard for Digital Geospatial Metadata (CSDGM), http://www.fgdc.gov/metadata/contstan.html
6. NISO Z39.87 Standard (http://www.niso.org/standards/)
7. IEEE Learning Technology Standards Committee's Learning Object Meta-data Working Group. Version 3.5 Learning Object Meta-data Scheme (http://ltsc.ieee.org/wg12/20020612-Final-LOM-Draft.html)
8. DIG35 Standard (http://www.i3a.org/i_dig35.html)
9. METS (Metadata Encoding and Transmission Standard) http://www.jisc.ac.uk/index.cfm?name=techwatch_report_0205
10. SMPTE Metadata Dictionary, (http://www.smpte-ra.org/mdd/)

Appendix B

A list of some of reference links to MM descriptor extraction tools is shown below.

1. Sources for MPEG-7 eXprerimentation Model XM Software (http://www.lis.e-technik.tu-muenchen.de/research/bv/topics/mmdb/e_mpeg7.html)
2. IBM MPEG-7 Annotation Tool (http://www.alphaworks.ibm.com/tech/videoannex)
3. IBM Multimodal Annotation Tool (http://www.alphaworks.ibm.com/tech/multimodalannotation)
4. Ricoh MovieTool Home (http://www.ricoh.co.jp/src/multimedia/MovieTool/)
5. MPEG-7 Audio Low Level Descriptors (http://www.whisper.elec.uow.edu.au/mpeg7)
6. Link to search semantic annotation tools eventually under construction (http://mpeg-industry.com/domain.cool)
7. BilVideo: A Video Database Management System (http://www.cs.bilkent.edu.tr/~bilmdg/bilvideo/)

References

[1] B. S. Manjunath, P. Salembier, T. Sikora, "Introduction to MPEG-7: Multimedia Content Description Interface", April 2002.
[2] TV-Anytime Forum, http://www.tv-anytime.org/
[3] MPEG-21 http://www.chiariglione.org/mpeg/standards/mpeg-21/mpeg-21.htm
[4] NewsML http://www.newsml.org/
[5] Dublin Core Metadata Element Set, Version 1.1, 2 July, 1999. http://www.purl.org/dc/documents/rec-dces-19990702.htm
[6] Indecs Metadata Model, November, 1999 (http://www.indecs.org/)
[7] Content Standard for Digital Geospatial Metadata (CSDGM), http://www.fgdc.gov/metadata/contstan.html
[8] GEM, The Gateway to Educational Materials http://www.geminfo.org/
[9] CIDOC Documentation Standards Group, Revised Definition of the CIDOC Conceptual Reference Model, September 1999 (http://cidoc.ics.forth.gr/index.html)
[10] Moving Image Collections (http://gondolin.rutgers.edu/MIC/)
[11] Moving Image Collections (http://mic.imtc.gatech.edu/index.php)
[12] MIC standards overview http://mic.imtc.gatech.edu/catalogers_portal/cat_standrs.htm#metastandards)
[13] MARC Standard (http://www.loc.gov/marc/)
[14] MPEG7 http://www.chiariglione.org/mpeg/standards/mpeg-7/mpeg-7.htm
[15] WP9, deliverable D9.1 "A Review of Data and Metadata Standards and Techniques for Representation of Multimedia Content".
[16] Frank van Harmelen, "The Semantic Web: What, Why, How, and When", IEEE Distributed Systems Online, Vol. 5, No. 3, March 2004
[17] Heflin, J. and Hendler, J. *"Semantic Interoperability on the Web"*. In Proceedings of Extreme Markup Languages 2000. Graphic Communications Association, 2000. pp. 111-120.
[18] J. Hunter, *"Working Towards MetaUtopia - A Survey of Current Metadata Research"*, Library Trends, Organizing the Internet, Edited by Andrew Torok, 52(2), Fall 2003.
[19] W3C eXtensible Murkup Language (XML) (http://www.w3.org/XML)
[20] Resource Description Framework (RDF) http://www.w3.org/RDF/
[21] W3C Web Ontology Language (OWL) (2003, March 31). Guide, Version 1.0, W3C Working Draft (http://www.w3.org/TR/owl-guide)
[22] Uts Westermann, Wolfgann Klas, *"An Analisys of XML Database Solutions fot the Management of MPEG-7 Media Descriptions"*, ACM Computing Surveys, Vol. 35, No. 4, December 2003, pp. 331-373.
[23] Platform for Internet Content selection (PICS) http://www.w3.org/PICS/
[24] Meta Content Framework (MCF) using XML http://www.w3.org/TR/NOTE-MCF-XML-970624/
[25] XML Topic Maps (XTM) http://www.topicmaps.org/xtm/index.html
[26] MPEG-21 http://www.chiariglione.org/mpeg/standards/mpeg-21/mpeg-21.htm
[27] Horrocks, I. and P. F. Patel-Schneider, *"A Proposal for an OWL Rules Language"*, WWW2004, *May, 17-22, 2004, New York, NY USA.*

[28] Grigoris Antoniou, Frank van Harmelen, "Web Ontology Language: OWL", Handbook on Ontologies in Information Systems, S. Staab and R. Studer (eds), Springer-Verlag, 2003. (http://www.cs.vu.nl/~frankh/postscript/OntoHandbook03OWL.pdf).
[29] The Rule Markup Initiative, (http://www.ruleml.org/)
[30] Horrocks, I., *"SWRL: A Semantic Web Rule Language Combining OWL and RuleML", Version 0.6 of 30, April, 2004, (http://www.daml.org/2004/04/swrl/rules-all.html)*
[31] North American Industry Classification System (NAICS) (http://www.census.gov/epcd/www/naics.html)
[32] Aseem Das, Wei Wu, and Deborah L. McGuinness, "Industrial Strength Ontology Management", Stanford Knowledge Systems Laboratory Technical Report KSL-01-09 2001. In the Proceedings of the International Semantic Web Working Symposium, Stanford, CA, July 2001 (http://www.ksl.stanford.edu/people/dlm/papers/ontologyBuilderVerticalNet-abstract.html).
[33] W3C Semantic Web (http://www.w3.org/2001/sw/)
[34] Nack, F.; Lindsay, A.T., "Everything you wanted to know about MPEG-7. 1", Multimedia, IEEE ,Volume: 6 , Issue: 3 , July-Sept. 1999, Pages:65 – 77.
[35] Nack, F.; Lindsay, A.T.; Multimedia," Everything you wanted to know about MPEG-7. 2", IEEE ,Volume: 6 , Issue: 4 , Oct.-Dec. 1999, Pages:64 – 73.
[36] Martinez, J.M.; Koenen, R.; Pereira, F., "MPEG-7: the generic multimedia content description standard, part 1", Multimedia, IEEE ,Volume: 9 , Issue: 2 , April-une 2002, Pages:78 – 87.
[37] Martinez, .M., "Standards - MPEG-7 overview of MPEG-7 description tools, part 2", Multimedia, IEEE ,Volume: 9 , Issue: 3 , July-Sept. 2002 Pages:83 – 93.
[38] van Ossenbruggen, J.; Nack, F.; Hardman, L., "That obscure object of desire: multimedia metadata on the Web, Part-1", Multimedia, IEEE ,Volume: 11 , Issue: 4 , Oct.-Dec. 2004, Pages:38 – 48.

Bibliography

Ching-Yung Lin, Belle L. Tseng, Milind Naphade, Apostol Natsev and John R. Smith, “VideoAL: A Novel End-to-End MPEG-7 Automatic Labeling System” , *IEEE Intl. Conf. on Image processing, Barcelona, Sep. 2003.*

Ching-Yung Lin, Belle L. Tseng and John R. Smith, “VideoAnnEx: IBM MPEG-7 Annotation Tool for Multimedia Indexing and Concept Learning,”, *IEEE Intl. Conf. on Multimedia & Expo, Baltimore, July 2003.*

Carlo Meghini, Fabrizio Sebastiani and Umberto Straccia. “A model of multimedia information retrieval”. *Journal of the ACM, 48(5):909-970, 2001.*

SMITH, J. R., AND CHANG, S.-F. 1997. Visually searching the Web for content. *IEEE Multimedia 4*, 3, 12–20.

J. Magalhães, F. Pereira, "Using MPEG standards for multimedia customization", *Signal Processing: Image Communication, 2004.*

Eyvind Fossbakk, Pilar Manzanares, Jose L. Yago, Andrew Perkis, “An MPEG-21 framework for streaming media ”, *in Proc. 2001 IEEE Workshop on Multimedia Signal Processing, Cannes 3-5 October, 2001.*

Y. Wang, Z. Liu and J. Huang, "Multimedia content analysis using audio and visual information," *IEEE Signal Processing Magazine, Vol. 17 n. 6,* November 2000.

van Beek P., Smith J.R., Ebrahimi T., Suzuki T., Askelof J. ,“Metadata-driven multimedia access”, *IEEE Signal Processing Magazine, Vol. 20, n. 2, p. 40-52*, March 2003.

Bormans J., Gelissen J., Perkis A., “MPEG-21: The 21st century multimedia framework”, *IEEE Signal Processing Magazine, Vol. 20, n. 2, p. 53-62*, March 2003.

Gal Ashour, Arnon Amir, Dulce Ponceleon, Savitha Srinivasan, “Architecture for Varying Multimedia Formats”, *ACM Multimedia Workshop*, Marina Del Rey CA USA, 2000.

Zheng Yin, Zhengfang Xu, and Abdulmotaleb El Saddik, “Study of Metadata for Advanced Multimedia Learning Objects”, CCECE 2003 – CCGEI 2003, Montreal, May 2003.

Martin Doerr, Jane Hunter, Carl Lagoze, “Towards a Core Ontology for Information Integration”, *Journal of Digital Information, Vol. 4 Issue 1*, April 2003.

C. Lagoze, J. Hunter, "The ABC Ontology and Model" (v3.0), *Journal of Digital Information, Vol 2 , Issue 2*, November 2001.

J. Hunter, "MetaNet - A Metadata Term Thesaurus to Enable Semantic Interoperability Between Metadata Domains", *Journal of Digital Information, Special Issue on Networked Knowledge Organisation Systems, Volume 1, Issue 8*, April 2001.

J. Hunter, F. Nack, "An Overview of the MPEG-7 DDL Proposals", *Signal Processing: Image Communication Journal, Special Issue on MPEG-7, Vol 16 pp 271-293*, 2000.

van Harmelen, F., “ The semantic web: what, why, how, and when“, *IEEE Distributed Systems Online, Vol. 5 n. 3 p. 1-4* , March 2004.

Greg Goth, “Multimedia Search: Ready or Not?”, *IEEE Distributed Systems Online, Vol. 5, n.7*, July 2004.

Catarci, T.; Donderler, M.E.; Saykol, E.; Ulusoy, O.; Gudukbay, U., “BilVideo: a video database management system”, *IEEE Multimedia Vol. 10, n. 1, p. 66-70*, March 2003.