

# Semantic Representation of Multimedia Content

Kalliopi Dalakleidi<sup>1</sup>, Stamatia Dasiopoulou<sup>2</sup>, Giorgos Stoilos<sup>1</sup>,  
Vassilis Tzouvaras<sup>1</sup>, Giorgos Stamou<sup>1</sup>, and Yiannis Kompatsiaris<sup>2</sup>

<sup>1</sup> School of Electrical and Computer Engineering,  
National Technical University of Athens,  
Zographou 15780, Athens, Greece  
[gstoil@image.ece.ntua.gr](mailto:gstoil@image.ece.ntua.gr)

<sup>2</sup> Informatics and Telematics Institute  
Centre for Research and Technology Hellas,  
Thermi-Thessaloniki, Greece  
[dasiop@iti.gr](mailto:dasiop@iti.gr)

**Abstract.** Multimedia documents constitute extremely rich information resources, whose efficacious management is intertwined with the effective capturing of the underlying semantics. The conveyed meaning may span along multiple levels and relates to search and retrieval tasks, much as to the very extraction and interpretation of content descriptions. In this chapter we consider the formal representation of multimedia semantics that pertain to media and domain specific descriptions, for the purpose of supporting both the extraction and subsequent semantic management of such descriptions. To this end, firstly, we present first an overview of existing approaches to the representation of multimedia content and discuss open issues. Subsequently, we present the ontology infrastructure developed in the context of the BOEMIE project tailored towards the formal representation of multimedia content. Concluding, we present what can non-standard formal representation technologies, such as fuzzy knowledge representation formalisms bring to multimedia document processing and management.

## 1 Introduction

Multimedia content made available nowadays on the Web and in digital archives amount to a striking volume, intensifying further the urge to process and manage the available content in a semantically rich way. As a multimedia document may convey a wealth of information ranging among others from thematic descriptions addressing scenes, objects and events (e.g. a landscape, a jet engine, scoring, running, etc.), to structural and signal level descriptions (e.g. blue/textured region, linear motion, etc.), the effective representation of such information becomes a critical requirement. This criticality relates not only to the consequent of enabling end-users to efficient query and retrieve multimedia content, but also to the intertwinement with the extraction of content semantics and the intricacies pertaining to automated multimedia interpretation.

During the last decade there have been intense research efforts aiming at developing a proper language by which one would be able to represent and query

(at semantic level) multimedia information. Such efforts gave rise to the MPEG-7 standard [32]. Through XML-Schema based definitions, MPEG-7 provides a rich set of tools for the description of multimedia content at different granularities and abstraction levels, including structural, low-level descriptors (e.g. colour, shape) and semantic descriptions, as well as aspects pertaining to authoring, user preferences, and so forth. Although the development of MPEG-7 has been a great advancement towards the systematic description of multimedia documents, significant deficiencies pertain to the means for and the axiomatisation of semantics representation [36, 57]. To a large extent, these deficiencies issue from the use of XML as the underpinning definition language, the flexibility allowed in the structuring of equivalent descriptions, as well as the restricted, and rather rigorous, model provided for the definition of domain specific semantic descriptions. As a result, the descriptions of multimedia information in a machine understandable way that would enable their sharing, reuse and interoperability has been hindered.

Towards this direction the approach of the Semantic Web [3] has proven to be the most promising way to achieve such goals, as ontologies can support a semantically rich, unambiguous and interoperable way of representing semantics, while additionally providing support for reasoning services that allow to extract further knowledge [48]. In addition to the well known paradigm of ontology based multimedia annotation, where domain specific ontologies are used to capture the semantics of subject matter descriptions associated to the multimedia content [27, 44], significant efforts have been undertaken in the last couples of years towards a more substantial deployment of ontologies in the management of multimedia semantics. Specifically, so called multimedia ontologies have been proposed to capture multimedia semantics through the formalisation and extension of the MPEG-7 modelling [1, 25, 41], while appropriately defined ontologies have been used to support tasks such as scene interpretation, object detection and retrieval [10, 33, 34, 45].

However, the aforementioned ontologies are intended for specific applications and tasks, and as a result tend to address the issues involved with respect to the modelling and representation of multimedia semantics in a fragmented fashion. On the contrary, in the BOEMIE<sup>1</sup> project, the formal representation of multimedia semantics has been the subject of research within an integrated application scenario that includes knowledge acquisition and representation, reasoning, multimedia ontology evolution, retrieval and presentation. As such, the proposed representation of multimedia semantics addresses media (content structure and low-level descriptors) and domain specific aspects, and is tailored to the analysis, interpretation and retrieval tasks that constitute the aforementioned chain of semantic content management.

Aiming to provide a systematic view of the aspects involved in the representation of multimedia content semantics within the context of semantics modelling and extraction, we provide in this chapter, on one hand an overview of the relevant literature and its weaknesses, and on the other hand, the ontology-based

---

<sup>1</sup> <http://www.boemie.org/>

representation model that has been developed within the BOEMIE project, towards the integrated confrontation of the issues involved. Nevertheless, classical ontology languages are often not capable of handling the type of information that results from multimedia processing tasks, which in many cases is imperfect (vague and/or uncertain). For example, an image analysis algorithm is not always able to assess to 100% accuracy the existence of an object. To account for the handling of such imperfect knowledge in multimedia interpretation and management tasks, non-standard technologies, which extend the proposed ontology infrastructure are also presented.

The rest of the chapter is organised as follows. Sections 2 and 3 provide an overview of relevant approaches towards the representation of multimedia content semantics. Specifically, Section 2 presents the different multimedia ontologies that have been proposed to formally capture the semantics of content structure and of the applicable low-level descriptors, while Section 3 considers the representation of content from the perspective of knowledge-based extraction and interpretation of the underlying semantics. Section 4 describes the semantic model developed within the BOEMIE project for the representation of multimedia content, while Section 5 presents some novel and non-standard ontology languages, which can be used to extend the expressive power of the proposed semantic model in order to handle imperfect information. Finally, Section 6 concludes the chapter and discusses open problems.

## 2 Multimedia Semantics Representation in Content Management

Multimedia assets form extremely rich sources of information. The conveyed meaning is communicated not only through intertwined multimodal information channels, but also through implicit connotations, narrative and discourse relations that create new levels of meaning. To be able to develop applications and services that are aware of the semantics, both the content and the context of multimedia need to be made explicit. Aiming at interoperable multimedia content description, a variety of multimedia metadata standards have been proposed addressing different levels of the conveyed information. However, in the developed multimedia standards and vocabularies, the semantics are rendered mostly in the form of syntactic norms with respect to corresponding XML Schema definitions, rather than the attachment of formal meaning.

The Semantic Web initiative induced efforts further, pushing towards machine understandable rather than machine readable semantics through the use of ontologies, i.e. explicit specifications of conceptualizations [17]. Ontologies are used to make meaning explicit contributing to the communication, exchange, reuse and sharing of knowledge across heterogeneous agents and applications. A number of ontology languages with varying expressivity have been proposed but the currently most prevalent standard by the World Wide Web Consortium<sup>2</sup>(W3C) is the Web Ontology language (OWL) [6]. Building on the Semantic

---

<sup>2</sup> <http://www.w3.org/>

Web paradigm, a number of multimedia ontologies have been proposed aiming to attach formal semantics to multimedia content representation and allow for more intelligent content management.

In the following, we describe the proposed multimedia ontologies and discuss the encountered weaknesses. For reasons of completeness, a brief account of the most popular, yet lacking formal semantics, multimedia standards and vocabularies is also given.

## 2.1 Non-formal Representations

Within the Moving Pictures Expert Group, two relevant multimedia description standards have been developed, namely the Multimedia Content Description Interface (MPEG-7) and the MPEG-21 Multimedia framework.

The goal of MPEG-7 [32] is to provide a rich set of standardised tools for the description of multimedia content, and in addition to support some degree of interpretation of the meaning of information so as to enable the exchange of multimedia metadata across applications as well as their efficient management, e.g. in terms of search and retrieval. It offers a set of audiovisual description tools in the form of Descriptors (Ds) and Description Schemata (DSs), describing the structure of the metadata, their relationships and the constraints to which a valid MPEG-7 description should adhere. MPEG-7 is organised in 8 parts: *Systems*, the *Description Definition Language* (DDL), *Visual, Audio, Multimedia Description Schemes* (MDS), *Reference Software*, *Conformance*, and *Extraction and Use*. The DDL consists the standard's core part, specifying the language for the definition of the description tools. The Visual and Audio parts consist respectively of structures and low-level descriptors that cover basic visual and audio features, while the MDS part specifies generic description tools pertaining to multimedia.

The MPEG-21 [35] activities address the definition of an open framework that allows the integration of all components of a delivery chain necessary to generate, use, manipulate, and deliver multimedia content across heterogeneous networks and devices. The key elements of MPEG-21 are: *Digital Item Declaration*, *Digital Item Identification and Description*, *Content Handling and Usage*, *Intellectual Property Management and Protection*, *Terminals and Networks*, *Content Representation*, and *Event Reporting*. From the aforementioned, content handling and usage, addressing the provision of interfaces and protocols to enable creation, search, access, delivery and reuse of content across the content distribution and consumption value chain is specifically interesting for multimedia content description. The same holds for the aspects addressed in the content representation, digital item identification and description elements, etc.

In addition to the MPEG activities, a number of multimedia metadata vocabularies emerged as the outcome of efforts undertaken by individual communities towards shared multimedia content descriptions. We refer indicatively, the Visual Resource Association (VRA) Core that specifies a small and commonly used vocabulary targeted especially at visual resources, and the Exchangeable Image File Format (EXIF), which specifies the formats to be used for images, sound, and

tags, in digital still cameras. Finally, the Synchronised Multimedia Integration Language (SMIL), that is an XML-based two dimensional graphic language that enables simple authoring of interactive audiovisual presentations, while Scalable Vector Graphics (SVG) allows describing scenes with vector shapes, text, and multimedia.

For a more thorough presentation of multimedia related metadata specifications the reader is referred to [42]. As outlined previously, a common characteristic shared among these multimedia representation schemes is that the intended semantics remain implicit in the syntax and the accompanying normative specifications.

## 2.2 Formal Representations

To enable multimedia on the Semantic Web and alleviate interoperability issues, a number of initiatives engaged in building multimedia ontologies by attaching formal semantics to multimedia content representations. The relevant activities are distinguished in two categories: those building on the MPEG-7 specification, and those following ad hoc modelling choices that are customised to specific application contexts.

Chronologically, the first initiative to make MPEG-7 semantics explicit was taken by Hunter [25] in 2001. The RDF Schema (RDFS) language was proposed to formalise the decomposition patterns of the Multimedia Description Scheme (MDS), the descriptors included in the Visual part, and some additional descriptors representing information about production, creation, usage and media features. The developed ontology has been ported to DAML and eventually to OWL Full [26], while later, extensions that address image analysis terms of the MATLAB Image Processing Toolbox have been also included [19]. The translation approach taken follows rigorously the standard specifications, hence, preserving in this way the intended flexibility of usage. This flexibility however comes with the cost of the inherited ambiguities present in MPEG-7 [36, 57], resulting in descriptions with multiple possible interpretations and ambiguous meaning [11].

Two MPEG-7 based RDFS multimedia ontologies, namely the Multimedia Structure Ontology (MSO) and the Visual Descriptor Ontology (VDO), have been developed within the aceMedia<sup>3</sup> project. MSO covers the complete set of decomposition tools from the MDS, while VDO addresses the Visual Part. The use of RDFS restricts the captured semantics to subclass and domain/range relations [5]. Both these approaches still suffer from the ambiguities that are also observed in the case of the Hunter ontology.

Another effort towards an MPEG-7 based multimedia ontology has been reported within the context of the SMARTWeb<sup>4</sup> project [38]. The developed ontology focuses on the Content Description and Content Management DSs. The respective multimedia content and segment classes along with a set of properties

---

<sup>3</sup> <http://www.acemedia.org/aceMedia>

<sup>4</sup> <http://www.smartweb-projekt.de/>

representing the decomposition tools specified in MPEG-7 enable the implementation of the intrinsic recursive nature of multimedia content decomposition. Although in this approach, axioms have been used to make intended semantics of MPEG-7 explicit, ambiguities are still present due to the fact that the corresponding MPEG-7 normative descriptions have been directly translated into concepts and properties whose semantics lie again mostly in linguistic terms.

Based on the work within the ReDeFer<sup>5</sup> project, the Rhizomik approach proposes a fully automatic translation of the complete MPEG-7 Schema to OWL [41], by mapping the XML schema of MPEG-7 to OWL. Human intervention is required only to resolve name conflicts stemming from the independent name domains for complex types and elements in XML. The resulting MPEG-7 ontology is in OWL DL and has been validated through its comparison against the manual translation of [26], which showed their semantic equivalence. The obvious advantage of the Rhizomik is the automatic translation of the complete MPEG-7 Schemata. However, when it comes to integration with domain specific ontologies, the Rhizomik approach is applicable only under the presumption that these domain ontologies have been re-engineered beforehand so that they extend the classes resulting from the corresponding Semantic DS structures.

An alternative approach has been adopted by the DS-MIRF framework [56]. Exploiting the MPEG-7 semantic description capabilities provided by the SemanticBaseType DS, the resulting ontology intends to serve as an upper multimedia ontology. A systematic methodology has been presented for the integration of domain specific semantics with the general-purpose semantic entities of MPEG-7 [55]. The developed ontology has been conceptualised manually and is in OWL DL. Transformation from XML to OWL, and conversely, is supported through a separate OWL DL ontology that holds the mappings between the original XML Schema and the corresponding OWL entities. Although sharing the same goal with Rhizomik, in terms of using MPEG-7 as a core multimedia content representation ontology, DS-MIRF does not require for the MPEG-7 Schema to be extended, allowing for efficient translation of MPEG-7 metadata to OWL assertions, and inversely.

The most recent approach to the formalisation of MPEG-7 semantics is the Core Ontology for MultiMedia (COMM) initiative [1] developed within the K-Space<sup>6</sup> and X-media<sup>7</sup> projects. Aiming to serve as a core ontology for multimedia, COMM utilises DOLCE [14] to provide a common foundational framework for the description of multimedia documents. COMM is in OWL DL and covers selected descriptors from the media, location and decomposition patterns of MDS, as well as the visual part. COMM extends the design patterns of *Descriptions & Situations (D&S)* [15] and *Ontology of Information Objects (OIO)* [13] in order to axiomatise the description at structural (content decomposition), algorithmic (functionality and parameters), and conceptual (semantics annotation) level. Thereby, COMM underpins at semiotic level the process of integrating

---

<sup>5</sup> <http://rhizomik.net/redefer>

<sup>6</sup> <http://kspace.qmul.net>

<sup>7</sup> <http://www.x-media-project.org/>

multimedia and domain ontologies for the description of various aspects of content, reinforcing conceptual clarity in the descriptions per se.

As aforementioned, besides the multimedia ontologies that have been developed based on MPEG-7, a number of customised multimedia ontologies have been proposed within specific applications. Thonnat et. al [24, 31], proposed a visual ontology that provides qualitative descriptions with respect to color, texture, and spatial aspects of the characterised content. Analogous qualitative visual descriptors have been also employed in the Breast Cancer Imaging Ontology (BCIO) [23]. In SCULPTEUR [30], an ontology for the museum domain has been combined with a graphical concept browser interface that allows navigation through the domain ontology semantic layer, as well as display of the different content types in appropriate viewers. In [4], a so called pictorially enriched ontology is proposed that uses visual prototypes to represent semantic concepts instead of linguistic concepts. In [20], a visual ontology (VO) is described, which combining MPEG-7 and WordNet descriptions, allows the representation of visual attributes, such as shape, colour, visibility, etc.

Despite sharing a common vision, the aforementioned approaches present substantial conceptual differences, reflected both in the modelling of content semantics as well as in the linking with domain ontologies. The various customised multimedia ontologies, adhering to application specific requirements, are hardly concerned with interoperability issues, while the MPEG-7 based multimedia ontologies, although aiming to alleviate interoperability issues, have introduced new ones, this time at a semantic level [11, 54].

The COMM ontology addresses the axiomatisation of multimedia description patterns, but does not confront the semantic ambiguities that relate to the extensions of the provided definitions through more specialised descriptions as those provided by the rest MPEG-7 based multimedia ontologies. The latter demonstrate a tendency for continually higher utilisation of the expressiveness provided by the ontology languages, yet they all suffer, to a lesser or greater degree, from ambiguous semantics. As a consequence, one ends up with descriptions that have multiple interpretations, even when construed with respect to the reference ontology, thus hindering not only their management but as well their linking with descriptions pertaining to different multimedia ontologies.

We note that in the case of MPEG-7 based multimedia ontologies, the observed semantic ambiguities refer in principle to the representation of the content structure information and of the applicable decomposition schemes, and not to the modelling of the MPEG-7 low-level description tools, since the latter comprise rigid numerical structures rather than conceptual notions. This is no longer the case for the customised multimedia ontologies though, where the different application contexts induce additional discrepancies. Moreover, since correspondence to the MPEG-7 structural and low-level descriptors cannot be always guaranteed, further questions are raised regarding the reuse and linking with existing MPEG-7 based descriptions. Consequently, a critical requirement for enabling the effective extraction and subsequent handling of multimedia semantics is the construction of multimedia ontologies with well-defined semantics.

### 3 Knowledge-Based Interpretation of Multimedia Content

Multimedia interpretation constitutes a particularly challenging problem that has engaged strong and continuous research interest. It refers to the lack of coincidence between the descriptions that can be extracted automatically from multimedia content at signal level, and the corresponding interpretations as acquired by a human [47]. In this endeavour, the use of background knowledge holds a central role, as the complexity of the problem renders purely data-driven approaches, severely inadequate to approximate what would consist a human like perception of the conveyed meaning.

This background knowledge is usually structured at levels of increasing abstraction, ranging from perceptual representations to logical interrelations that define the entities and notions of interest. Different perspectives on what constitutes multimedia semantics have resulted in the development of knowledge models that address different levels and types of knowledge, and define different interrelations between the employed abstraction levels. These differences affect in turn the espoused knowledge representation formalism as well as the configuration of the multimedia interpretation process as an inferencing task. In each case, the adopted representation formalism determines the degree at which explicit and formal semantics are supported.

In the following, we outline the effects pertaining to the representation of multimedia semantics from the perspective of content interpretation. First, general considerations that apply in the use of knowledge and reasoning in the extraction of multimedia semantics are discussed, and in the sequel characteristic examples of existing works are discussed.

#### 3.1 Knowledge-Based Multimedia Semantics Extraction

The development of knowledge-based approaches to multimedia semantics extraction confronts two crucial questions: i) which representation formalism is suitable for capturing the semantics at hand, and ii) what pieces of information constitute the knowledge that is required for solving the addressed problem.

Regarding the first, and bearing in mind that the focus of this chapter is on formal semantics, the various alternatives, as suggested by the existing literature, have been largely influenced by the Semantic Web initiative. Ontology languages such as OWL [6] and their logical underpinnings, Description Logics [21] have become prevailing choices. The popularity of DLs issues not only from the direct relation with OWL, but also from the fact that they constitute expressive fragments of first order logic, for which decidable reasoning algorithms exist [2].

The expressivity provided by the different representation formalisms, determines the appropriate choice in accordance with the types of knowledge and reasoning tasks that comprise the extraction of content semantics. As will be described in the subsection 3.2, there are approaches that employ hybrid schemes, combining more than one representation formalisms. Given the differences in the



provided expressivity such observations constitute important issues with respect to the kind of expressivity required for supporting multimedia interpretation. It should be noted though that in some cases, the more intuitive formalisms are the ones that finally prevail.

The second question is intertwined to the espoused perspective on what multimedia semantics consists in. An aspect shared among the different approaches, is that the employed knowledge, in addition to providing support for the analysis and extraction processes, it also provides the vocabulary and the semantics of the produced content annotations. This enables content management services, such as search, retrieval, filtering, etc., at a semantic level. This vocabulary is not necessarily restricted to domain specific descriptions, but may include other aspects as well, such as content structure. The latter is a prerequisite in order to provide finer indexing and retrieval services, and support transcoding applications.

Regarding the extraction per se, the tasks for which knowledge and reasoning have been utilised fall roughly into three categories: i) the translation of automatically extracted features to semantic entities, ii) the extraction of descriptions of higher abstraction based on the logical associations that underly the semantic entities that are directly detectable by means of analysis, and iii) the specification of the control strategy, i.e. of the steps and parameters comprising the analysis process itself. Plausibly, the tasks at hand have a strong interrelation with the types of knowledge captured. For example, in approaches tackling the first task, there exists representations of features pertaining to audiovisual manifestations as well as corresponding domain concept definitions with respect to the constraints and range values that apply with respect to the modelled audiovisual features (e.g. colour, texture, motion, etc.). Approaches addressing the second task on the other hand, focus more on the capturing of semantic interrelations and attributes between domain entities. Hence, the background knowledge is populated mostly with concept definitions that reflect complex notions whose meaning lies in logical aggregations, rather than audiovisual manifestations.

It is interesting to note that although multimedia semantics extraction aims at educing descriptions close to what a human interpretation would be, the overview of the state of the art reveals that the majority of the approaches considers mostly the first task. This means that the employed knowledge, even when adequately capturing the specific domain semantics, is mostly utilised for the purposes of annotation, while in the extraction only semantics relevant to audiovisual features are used. Adding to this the fact that axioms defining concept with respect to audiovisual features entail numerical computations rather than logical inference, shows that despite using very expressive knowledge representation languages, with powerful inference services, their potential is poorly exploited.

Another issue relevant to multimedia semantics extraction is the handling of uncertainty, a feature inherent in multimedia analysis and understanding. The numerical nature of segmentation and the incompleteness, to a large extend due to the inability to capture semantics only by means of audiovisual manifestations, of the perceptual models describing semantic entities, allow only for partial

matching against these models. As a result the extracted analysis representations cannot be interpreted as indisputable evidences. From the aforementioned representations, none provides directly the means to handle this uncertainty. As will be described in the next subsection, most approaches handle the uncertainty indirectly, by defining thresholds with respect to the degree of similarity against the defined audiovisual features' models that is acceptable. However, once the similarity is evaluated and the decision is taken, the uncertainty information is usually dismissed, i.e. in the resulting assertions (facts) that comprise the content annotation, there are no degrees. This also means that whichever reasoning is applied afterwards, is performed over crisp terms.

The aforementioned aspects lie in the core of the development of knowledge-based approaches for the extraction of semantic descriptions from multimedia; however, these are not the only dimensions involved. Knowledge acquisition, supported media type, and sequential vs interactive extraction, are indicative examples of relevant issues.

### 3.2 Related Work

In the following, we briefly summarise different approaches of knowledge-based multimedia systems.

In the series of works presented in [24, 31], an ontology-based approach is followed for the representation of knowledge. The employed knowledge builds upon the premise of addressing separately the three abstraction levels as defined by Marr. A domain ontology provides the corresponding conceptualization for the various domains of images considered, i.e., pollen grain, galaxies, rose diseases, transport vehicles, etc., while a visual concept ontology is employed to provide symbolic, intermediate level definitions related to color, texture and spatial information, that allow linking the domain concepts with the raw image data. The extraction of semantic description for images is realised in the form of rule-based reasoning, performed in a linear fashion in order to derive descriptions of successively higher-abstraction in a stepwise fashion, starting from the available at visual level information.

A similar approach is taken in [26], where rule-based reasoning is employed in an non-iterative manner to derive semantic annotations based on the manually defined mappings between domain concepts and visual characteristics. Three OWL ontologies capture the different knowledge components involved, i.e., low-level visual features, microscope information, and domain specific knowledge (fuel and pancreatic cells). Contrary to the customised visual descriptions of the adopted in [24, 31], the low-level visual features ontology builds on the corresponding MPEG-7 visual descriptors [25].

The ontology-based framework proposed in [5] adopts a similar perspective. A domain ontology captures the logical associations that define the relevant concepts and relations, while two MPEG-7 based ontologies model low-level visual descriptors and content structure, as described in Section 2. The linking of domain concepts with prototypical low-level descriptors' values is realised through

M-Ontomat-Annotizer [39], which formalises the interconnection between the two ontologies.

In [10], semantic concepts in the context of the examined domain are defined in an ontology, enriched with qualitative attributes (e.g., color homogeneity), low-level features (e.g., color model components distribution), object spatial relations, and multimedia processing methods (e.g., color clustering). The RDF(S) language has been used for the representation of the developed domain and analysis ontologies, while for the rules that determine how tools for multimedia analysis should be applied depending on concept attributes and low-level features, are expressed in F-Logic. Compared to the previous approaches, [10] brings in the modelling of analysis new dimensions as well, while for the linking of visual descriptions with domain concepts a similar rationale is followed.

OntoPic [43], is a supervised learning system that utilises DL-based reasoning, treating concept recognition as a classification problem. An appropriately constructed TBox provides the hierarchy of the domain concepts and their spatial topology. The initial definitions are extended during the learning phase with feature roles that associate domain concepts to the features and feature value constraints that resulted from the training. A pseudo-extension to fuzzy DL is introduced to avoid overspecification. During a postprocessing step, the resulted membership values can be re-adjusted according to feature weights reflecting their discriminative power. Finally, the classified regions are checked in terms of spatial consistency, utilising once again the DL inference services. To avoid ending up with inconsistent ABoxes, the violations of spatial constraints are treated as *non-concept* definitions which OntoPic removes successively, starting from the one with the lowest degree of membership, until a consistent ABox, i.e., image description, is reached.

Hence, as in the previously described approaches, two abstraction levels are employed for the representation of content semantics, i.e. domain specific descriptions and low-level visual descriptions. However, contrary to the previous approaches, OntoPic utilises the axiomatic definitions that link the descriptions of the two levels in a more semantically rich way. Specifically, the linking axioms are not used simply as the means to realise the transition from visual descriptions to semantic domain specific notions in the form of “IF” “THEN” production rules, but support the construction of semantically constituent, logical models.

In [37], Description Logics are used for acquiring scene interpretations. The notion of *aggregate concept* is introduced for realising scene interpretation as a stepwise process utilising taxonomical and compositional relations. The interpretation process works on top of primitive descriptions derived directly from visual evidence, and further contextual information is introduced in the form of spatial and temporal constraints. Four kinds of steps, namely aggregate instantiation, instance specialization, instance expansion and instance merging, are used to realise scene interpretation as model construction. In addition, coupling with a probabilistic framework is proposed in order to provide guidance among the different plausible interpretations.

Rule-based reasoning is employed in the approach to video understanding presented in [28]. Visual, auditory and textual aspects of the video are taken into consideration to semi-automatically construct multimedia ontologies that will provide the definitions required in the sequel for the extraction of video semantics. After automatic speech recognition (ASR) and alignment to video shots, the produced textual data along with the available text annotations are processed using *KAON*<sup>8</sup> and exploiting Wordnet<sup>9</sup>, in order to select relevant concepts included in the employed TGM I vocabulary. Similarly, visual detectors based on low-level content features (color, texture, etc.) are used and associated with corresponding terms, while reasoning concerns the application of context rules to adjust the confidence values of the visual detectors.

In [8], an approach to fuzzy reasoning is proposed in order to integrate image annotations at scene and region level, into a semantically consistent final description, further enhanced by means of inference. An ontology is used to capture the underlying domain semantics and allow the detection of incoherences, while rules are used to allow the effective representation of spatial related axioms. The assimilation of fuzzy semantics allows to handle the uncertainty that characterises multimedia analysis and understanding, while the use of DLs allows to benefit from the high expressivity and the efficient reasoning algorithms in the management of the domain specific semantics. The initial annotations forming the input may come from different modalities and analysis implementations, and their degrees can be re-adjusted using weights to specify the reliability of the corresponding analysis technique or modality.

The aforementioned approaches constitute characteristic examples, where the representation of content semantics not only serves in the semantic structure and management of multimedia descriptions, but in addition underpins the extraction of such descriptions. In their most straightforward form, the proposed approaches involve the representation of some types of perceptual features (often in the form of MPEG-7 descriptors) and the definition of axioms that link domain specific concepts with combinations of valid feature values. In this manner though, reasoning is employed in a rather trivial fashion as it assimilates more the functionality of production rules rather than the construction of logical models. Reasoning as logical entailment is investigated more thoroughly in [8, 37, 43], where the captured semantics are used in order to ensure the construction of consistent content interpretations.

Furthermore, the proposed approaches are tailored to the adopted interpretation perspective, and as such they address only selected content representation aspects. As a result, there exists a lack of an integrated representation framework that would enable to address the formal modelling of the different types and abstraction levels of the relevant information, including the different modalities, as well as the dynamic nature of the knowledge involved. In the following, we present the ontology infrastructure developed in the context of the BOEMIE

---

<sup>8</sup> <http://kaon.semanticweb.org/>

<sup>9</sup> <http://wordnet.princeton.edu/>

project in order to address such issues and provide support for enriched content interpretation as well as management services.

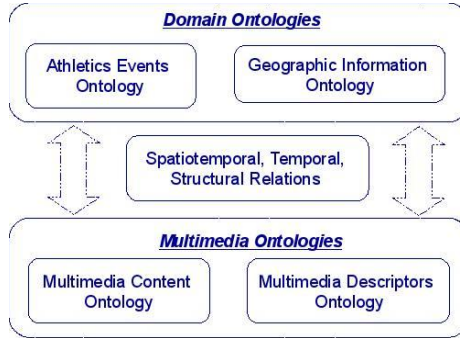
## 4 Representation of Multimedia Semantics in BOEMIE

In the current section we present the architecture and design choices followed in the context of the BOEMIE project in order to construct an ontology infrastructure. This infrastructure is developed in such a way that it can provide the means to manage and combine multimedia specific information and domain-specific one in order to enable:

- The semantic labelling of multimedia documents after the detection of concepts and relations from low-level analysis modules.
- The enrichment of the annotation of multimedia documents by providing definitions for complex (high-level) concepts utilised by reasoning services.
- Presentation and retrieval of multimedia documents w.r.t. the information that they convey.
- The evolution and learning process by providing a modular and pattern based ontology infrastructure which can be (semi)automatically evolved.

In order to account for the different types of knowledge involved and meet the different requirements imposed by the different modules which use the ontology infrastructure, the developed ontology model consists in practice of several interrelated and interlinked ontologies that can be divided into two categories. The first category consists of the multimedia ontologies, while the second one of the so called domain ontologies. Each of these two categories further contains two also independent ontologies. More precisely, the domain ontologies include the *Athletics Events Ontology* (AEO), describing our domain of interest which is public athletics events, and the *Geographic Information Ontology* (GIO), describing geographic information. On the other hand, the multimedia ontologies consist of the *Multimedia Content Ontology* (MCO), representing content structure information, and the *Multimedia Descriptors Ontology* (MDO), representing low-level numerical information extracted by analysis modules. An advantage of the proposed architecture is that it is highly modular, as the multimedia structure-related information is independent of the content and common for all multimedia documents, whereas the information about the content of a multimedia document depends totally on its subject. Furthermore, this discrimination can significantly improve the response time of the system to content related end-user queries, since the multimedia structure-related information is usually larger than the domain specific one, but also much less interesting for the end-user.

The four individual ontologies are interconnected and therefore can be used by applications that need to combine information and knowledge from different resources. Thus, the developed ontologies do not stand alone but are interlinked through proper structural, spatial, temporal, or any other kind of relations, of



**Fig. 1.** Architecture of the Multimedia Semantic Model

which the domain or range might be defined in different ontologies. This interconnection finally provides a global and modular ontology infrastructure which is called *Multimedia Semantic Model* (MSM). Figure 1 depicts the overall architecture of the MSM model with the various ontologies and their interconnections. As we can see besides the interconnections between ontologies of the same categories there are also interconnections between ontologies of the multimedia and the domain category.

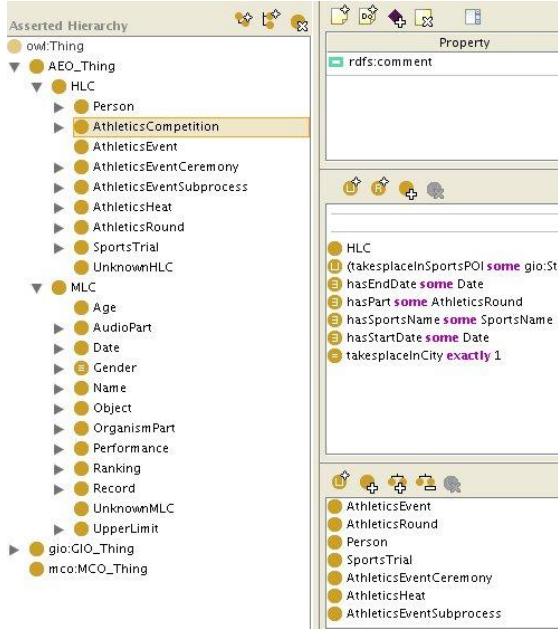
The knowledge representation formalism that we adopted for the construction of the ontologies of the MSM is Description Logics (DLs) [2]. DLs belong to the family of concept-based representation formalisms and actually consist of expressive fragments of First Order Logic (FOL), providing decidable and empirically tractable reasoning services, like logical consequence (entailment) and concept subsumption, i.e. checking if a concept (class) is a sub-concept (sub-class) of another one.

In the following, an overview of the ontologies of the MSM is provided.

#### 4.1 Domain Knowledge Representation

**Athletics Events Ontology.** The *Athletics Events Ontology* (AEO) is a formal conceptualization of the domain of interest of the BOEMIE use case scenario which is public athletics events, i.e. jumping, running and throwing events held in European cities. The concepts and relations of the AEO are used for annotation and retrieval of multimedia documents on the subject of athletics events, i.e. on information relevant to athletics competitions and their constituents events as well as information about athletes and performances gained in such events.

During the knowledge acquisition phase of the ontology development process, and taking into consideration the results of analysis, a discrimination has been established between the representation of concepts (semantic entities) that can be immediately instantiated by analysis modules, such as concrete objects, or names of athletes and locations, also called *Mid Level Concepts* (MLCs) in the framework of BOEMIE, and the representation of more abstract concepts that cannot be detected automatically by analysis, also called *High Level Concepts*

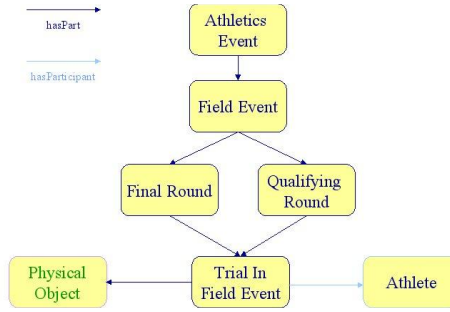


**Fig. 2.** The root concepts of the Athletics Events Ontology

(HLCs), such as composite events. This discrimination is required both in the ontology evolution process, in terms of applying different patterns for the definition of a new concept, accordingly to its substance, i.e. whether it is a MLC or a HLC as well as in image interpretation and reasoning. As a consequence, the root classes of the AEO hierarchy are the MLC concept and the HLC concept, as shown in Figure 2.

Two different design patterns have been implemented, one for the definition of MLCs and one for the definition of HLCs. MLCs are formalised as atomic concepts, subclasses of the MLC root concept of the AEO hierarchy (e.g.  $\text{Object} \sqsubseteq \text{MLC}$ ). Every modality provides its own MLCs. Thus, the subclasses of the MLC concept are *Age*, *Date*, *Gender*, *Audiopart*, *Performance*, *Ranking*, *Name*, *OrganismPart*, etc. Among these, the concepts *Age*, *Date*, *Gender*, *Performance*, *Ranking* and *Name* can be instantiated by text analysis whenever a relevant string is detected. On the other hand, image analysis instantiates mainly concepts that are subclasses of the concepts *Object* and *OrganismPart*, whenever a relevant image region is detected.

HLCs are formalised as complex concepts that appear in the left-hand side of terminological axioms built using DL concept constructors such as  $\exists, \forall, \sqcup, \sqcap$ . HLCs are designed as aggregates, which consist of multiple parts that can be either MLCs or other HLCs, and are constrained by relations representing spatial, temporal and other kinds of logical relations between these multiple parts, based on the approach described in [34].



**Fig. 3.** Conceptualisation of field athletic events and their partonomical relations

The subconcepts of the HLC concept conceptualise the complex concepts of the domain of athletics based on descriptions provided by IAAF Competition Rules and IAAF Technical Regulations<sup>10</sup>. Thus, the most important subconcepts of the HLC concept are the following:

- The concept **AthleticsCompetition**, which conceptualises series of events held over one or more days, i.e it conceptualises whole athletics competitions, such as the Olympic Games are, which are composed of different kinds of events.
- The concept **AthleticsEvent**, which conceptualises a single race or contest in a competition that takes place in a specific point of space and time. An athletics event might be a track, a field, a roadrace, a racewalking, a cross country or a combined event. Moreover, track events and field events consist of either one final round or more qualifying rounds.
- The concept **AthleticsRound**, which conceptualises a single round, final or qualifying, in an event that takes place in a specific point of space and time. A qualifying round consists of more than one athletics heats.
- The concept **AthleticsHeat**, which conceptualises a single heat held in a track or field event that takes place in a specific point of space and time, whenever the number of athletes is too large to allow the event to be conducted satisfactorily in a single round (final).
- The concept **AthleticsTrial**, which conceptualises a single trial in a field event that takes place in a specific point of space and time.
- The concept **Person**, which conceptualises persons that participate in very different ways in an athletics competitions. Therefore, its subclasses are not only **Athlete** but also **TechnicalPersonnel**, **Judge**, **Coach** and **Referee**.

The partonomical relation that we have used in order to represent the fact that competitions are composed of events, and events are composed of rounds, etc., is the transitive relation **hasPart**, as can be seen in Figure 3.

<sup>10</sup> <http://www.iaaf.org>



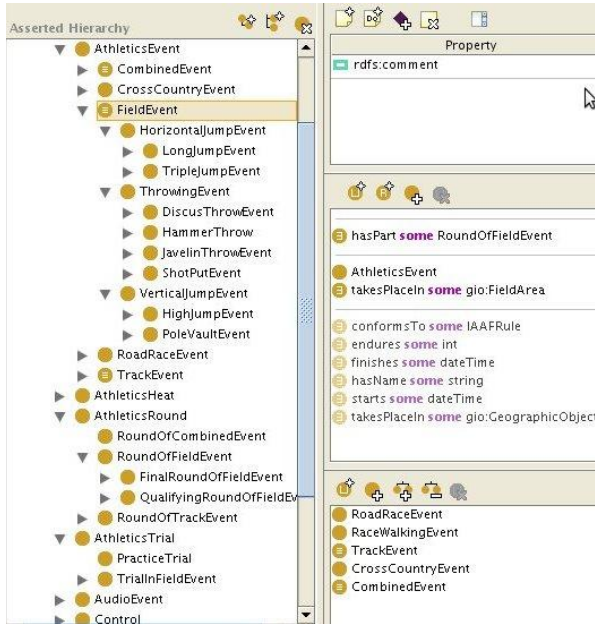


Fig. 4. Conceptualisation of the concept **FieldEvent**

Finally, in order to address the several characteristic aspects of athletics events, corresponding specialisation concepts have been introduced. In Figure 4 the subclasses of the **AthleticsEvent** concept are illustrated as well as the definition of the specialisation concept **FieldEvent**. We can observe that the necessary conditions for an instance of the **FieldEvent** concept are that it is composed of rounds and that it takes place in the field area of a stadium. In addition, it inherits necessary conditions by its superconcept **AthleticsEvent**, i.e. it must start and finish on a specific date, it must have a specific duration, it must conform to a specific IAAF rule and it must have a specific name. In the same way, all events are defined with respect to their specific attributes.

**Geographic Information Ontology.** The context of usage of the *Geographic Information Ontology* (GIO) within BOEMIE consists in providing the representation of the relevant geographic information in order to associate events/objects from the annotated multimedia content to the respective place/location they take place in (e.g., the stadium and city in which a given athletics competition takes place). In this way, the GIO enables visualisation and navigation on enriched with domain specific information maps (e.g., visualisation of a marathon route on a city map). Moreover, the GIO can provide assistance in the interpretation process through the exploitation of geographic information. To accomplish the aforementioned, the GIO needs to provide support for the representation of the following types of information:

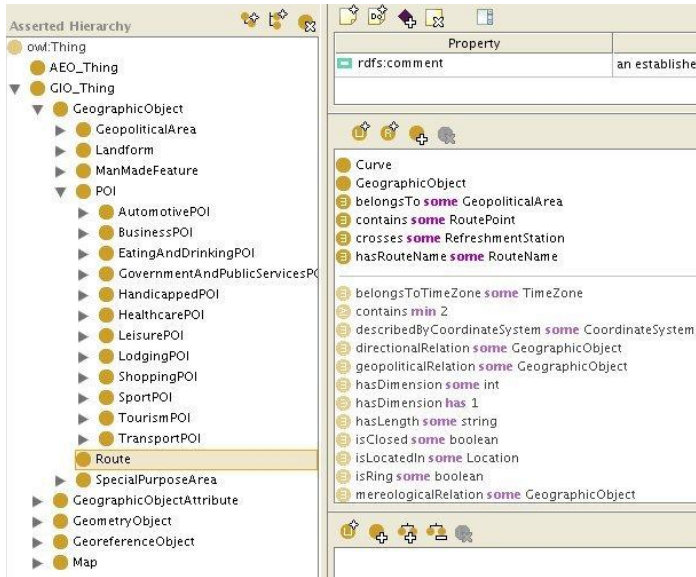
- Geopolitical information, i.e. information about geographic areas, which are associated with some sort of political structure, such as continents, countries and cities.
- Geographic information regarding places and locations of interest.
- Position related information, i.e. coordinates and respective coordinate systems, so that the considered objects can be linked/projected to corresponding map positions.
- Spatial relations, so that from an initial set of geometry-based calculated relations, further ones may be obtained automatically through inference services.

For the development of the GIO, the TeleAtlas database schema model<sup>11</sup> has been used as a guideline, especially for the identification of the types of information that should be covered. TeleAtlas database provides extremely rich, hierarchically structured, thematic information in the form of *Points Of Interest* (POI) and an underlying geometry features' model that enables equally rich functionalities in terms of calculating spatial relations holding among the given geographic objects. Considering the purely geographic information, such as coordinate systems and units of measures, this choice is also justified by the fact that TeleAtlas has followed the corresponding OpenGIS standard specifications. With respect to the thematic information, we observed again compliance to a high degree with the ontologies and vocabularies employed in the relevant literature, so we used the TeleAtlas taxonomy as the basis and applied modifications and further enrichments where necessary. The top level concepts of the developed GIO, illustrated in Figure 5, are the following:

- **GeographicObject**: The **GeographicObject** concept is used to represent any type of object used for referring to geographically related information. Each geographic object is associated with some map, on which it is projected, and some coordinates that identify its position within this map. In addition, it is related to other geographic objects through spatial relations, it belongs to a specific timezone and is located in some location. Moreover, the **GeographicObject** class comprises the **GeopoliticalArea**, **Landform**, **ManMadeFeature**, **POI** (Point of Interest), **Route** and the **SpecialPurposeArea** classes. The **GeographicArea** concept accounts for the different categories of geographic areas, such as countries and cities. The **POI** concept models in a hierarchical manner locations / places of general interest. Some indicative subclasses of the **POI** concept are **SportPOI**, **LeisurePOI** and **TransportPOI**. Subclasses of the **SportPOI** that are mainly used for representing the locations where athletic competitions take place are the concepts **Stadium**, **SwimmingPool**, **TennisCourt**, etc. In addition, although not included in the Teleatlas database schema, we have defined the concept **Route**, as a subclass of the concept **GeographicObject** to represent geographic information relevant to the route of road race events.

---

<sup>11</sup> <http://www.spatialinsights.com/catalog/product.aspx?product=95>



**Fig. 5.** A part of the GIO hierarchy

- **Map:** The **Map** concept is a symbolised depiction of a space which highlights relations between components of that space. To identify the referred map, a string denoting its location (file, url, etc.) is associated with it.
- **GeoreferenceObject:** The **GeoreferenceObject** concept is used to represent information for referring to the location of a specific geographic object by means of coordinates. The subclasses **Coordinate**, **CoordinateSystem** and **CoordinateValue** are used to represent coordinate related information.
- **GeometryObject:** It is used to provide geometry-dependent information about geographic objects. The **GeometryObject** class has subclasses the following concepts **Point**, **Curve**, **Surface** and the concept **GeometryCollection**. Each geometric object has specific important features, which can be inherited by geographic objects and provide important information about them. For example, since a route of a race event is a curve, and a curve has a certain length, a starting and an ending point, then a route should also have a certain length and a starting and ending point.
- **GeographicObjectAttribute:** This concept represents important attributes of geographic objects, such as their address, their official name, etc.

Additionally, with respect to the different types of geographic areas included, corresponding sets of spatial relations have been defined. More specifically, the properties **geopoliticalRelation**, **topologicalRelation**, **directionalRelation** and **mereologicalRelation** have been introduced and appropriate sub-properties have been defined.

## 4.2 Structure and Low-Level Descriptor Representation

**Multimedia Content Ontology.** The Multimedia Content Ontology (MCO) addresses structural aspects (i.e. decomposition semantics) pertaining to the different multimedia content types. Such knowledge is required to enable attaching annotations to the corresponding content parts (e.g. to annotate a specific still region of an image as depicting an athlete or a video segment as depicting a pole vault trial) and handle part-whole semantics (e.g. an image is comprised of the set of its constituent still regions to which it is segmented, thus if one still region depicts an athlete, the image itself depicts this athlete as well). Providing the means to capture and represent such knowledge, the MCO aims to support for unambiguous multimedia annotation, retrieval, exchange, and sharing of meta-data addressing media related aspects, as well as the application of inference. Therefore, its construction is based on the distinct representation of:

- the different types of multimedia content (e.g. images, captioned images, web pages and video),
- the possible logical relations among them (e.g. a web page may consist of a text extract, two images, and an audio sample),
- the semantics of the decomposition of the corresponding media types into their constituent parts according to the level of the produced annotations, e.g. a video can be decomposed into video segments based on shots, each of those segments further decomposed into constituent frames or moving regions when more detail with respect to localization is required,
- and the relations that associate multimedia content to the semantic entities conveyed (e.g. a still region depicts a person face).

As such, the MCO is strongly related to semantics extraction task, since during fusion information, about the provenance of the annotations extracted by the individual modalities is utilised. Furthermore, providing the means to represent the decomposition of multimedia documents into constituent parts, it supports the information retrieval and presentation tasks. The main top level classes include the `mco : MultimediaContent` class, which captures through its specialisation the various single and multiple modality content types of interest, the `mco : MultimediaSegment` class, which comprises the different segment types to which the various media items can be (spatially, temporally or spatiotemporally) decomposed to, and the `mco : SegmentLocator` class, (see Figure 6) which includes information about the various ways for identifying and designating a particular segment. The implemented MCO follows to a large extent the guidelines

```

SingleMediaItem  $\sqsubseteq$   $\exists$ hasMediaDecomposition.MultimediaSegment
Image  $\sqsubseteq$   $\exists$ SingleMediaItem
Image  $\equiv$   $\forall$ mediaHasDecomposition.StillRegion
StillRegion  $\equiv$   $\forall$ segmentHasDecomposition.StillRegion  $\sqcap$ 
 $\forall$  hasSegmentLocator.VisualLocator

```

**Fig. 6.** Part of StillImage definition in the MCO

specified in the MPEG-7 structure of content Multimedia Description Scheme, while enhancing it in order to avoid its inherent ambiguities. To accomplish this, the definition of the various content and segment types is logically grounded on the applicable decomposition schemes and the localisation information required for the identifications; thereby, and contrary to the respective definitions in the relevant literature, MCO models unambiguously the semantics of the notions involved.

**Multimedia Descriptors Ontology.** The Multimedia Descriptor Ontology (MDO) captures knowledge related to low-level representation of multimedia content, i.e. information about the descriptors employed by the different modalities to characterise content at feature (signal) level. The MDO is strongly related to the semantics extraction task, since it supports the individual modalities analysis in the detection of mid-level concepts (MLCs) through the linking of descriptors to domain specific concepts, as well as in the enhancement of their performance, enabling clustering of feature-level similar objects, and thus supporting the handling of unknown MLCs. The MDO has been designed based on two principles:

1. compliance with the respective MPEG-7 Visual and Audio parts to ensure wide coverage and interoperability in case of modalities processing enrichment with additional analysis modules, and
2. support for the requirements specific in the BOEMIE project with respect to the addressed modalities and the used tools.

As a result of the latter for example, since analysis focuses on quantitative descriptions, i.e. numerical representations of the analysed visual properties, quantitative descriptors (e.g. such as bright/dark, smooth/coarse) have not been addressed. The top level concept of MDO is the `mdo : MultimediaDescriptor` concept which is subclassed with respect to the different modalities into the concepts `mdo : VisualDescriptor`, `mdo : AudioDescriptor`, and `mdo : TextualDescriptor`. In addition, the `Adds` concept, also subclassed with respect to the different modalities, has been introduced to provide the means to capture information required for representing the corresponding modality descriptors. Each of the latter serves as the root of the ontology component representing the respective

$$\begin{aligned}
 \text{DominantColorDescriptor} &\sqsubseteq \forall \text{hasDominantColor}.\text{DominantColorComboValue} \\
 &\quad \sqcap \geq 1\text{hasDominantColor} \\
 \text{DominantColorComboValue} &\sqsubseteq \\
 &\quad \forall \text{hasColorQuantizationComponent}.\text{ColorQuantizationDescriptor} \\
 &\quad \sqcap \geq 1\text{hasColorQuantizationComponent} \\
 &\quad \sqcap \forall \text{hasColorSpaceComponent}.\text{ColorSpaceDescriptor} \\
 &\quad \sqcap \forall \text{hasColorValuesComponent}.\text{ColorValuesElement} \\
 &\quad \sqcap \leq 8\text{hasColorValuesComponent} \\
 &\quad \sqcap \forall \text{hasSpatialCoherencyComponent}.\text{SpatialCoherencyElement}
 \end{aligned}$$

**Fig. 7.** The definition of the Dominant Color Descriptor in the MDO

modality descriptors. Visual descriptors include color, texture, shape, motion and localization descriptors as for example the concepts: `mdo : DominantColor`, `mdo : HomogeneousTexture`, `mdo : TrajectoryType`, etc., while auditory descriptors address basic audio signal features as for example the following descriptors: `mdo : FundamentalFrequency`, `mdo : ZeroCrossingRate`, etc. Similarly, the defined properties are organised in a hierarchical way. For example, the relation `mdo : hasDominantColorDescriptor` is subsumed by `mdo : hasColorDescriptor` which in turn is subsumed by `mdo : hasVisualDescriptor`.

### 4.3 The Multimedia Semantic Model

Although that for the sake of ontology design we have considered the four ontologies as separate ontological modules, their borders are in fact vague. While developing an ontology, we confronted often the situation in which we needed to define a new relation the domain of which belonged to the ontology that we were developing at that time but the range belonged to another ontology of our framework. Thus, and through the definition of appropriate relations spanning across multiple ontologies, a network of structural, spatial and temporal relations, of which the domain and range belonged to different ontologies, emerged gradually. This network of relations comprises the so called Multimedia Semantic Model (MSM) that realises the integration of the different ontological modules into an interlinked and interconnected ontology infrastructure.

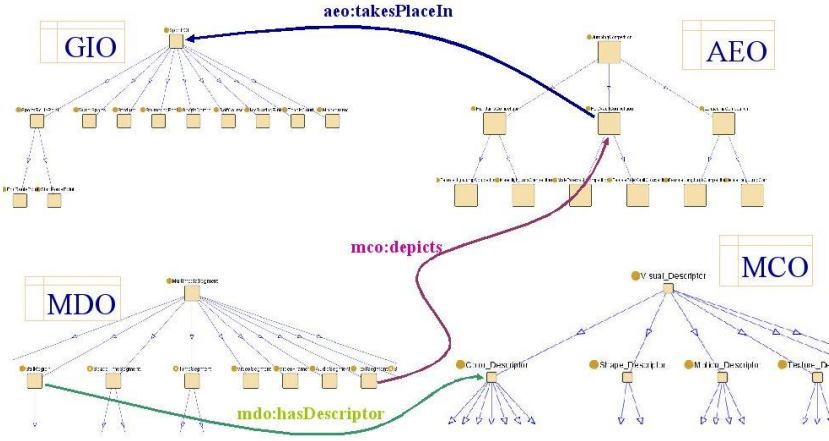
We note again, that all four ontologies, as well as the MSM model of interrelations have been manually engineered, while the specifications and requirements for new relations and concepts, as well as for the revision and enhancement of existing definitions, have issued from the feedback received regarding the use of the ontologies in the tasks of multimedia analysis, interpretation, management, and ontology evolution addressed within the BOEMIE project.

The Multimedia Semantic Model is illustrated in Figure 8, where we can observe examples of these interlinking relations, which can be divided in the following three categories according to our ontology architecture:

- Relations among concepts of the multimedia ontologies: These relations combine information about structural aspects of multimedia documents with information about low-level features of multimedia objects and can be helpful for the presentation of multimedia objects as well as learning algorithms of new concepts from unknown objects. An indicative example of this kind of relations is the `mdo : isDescriptorOf` relation which connects instances of descriptors, defined in the MDO, with instances of the multimedia segments that they describe, defined in MCO. For example, in order to represent the fact that an instance of a still region has a certain color descriptor we would use the following assertion:

`mdo : isDescriptorOf(mdo : ColorDescriptor1, mco : StillRegion1)`

- Relations among concepts of the domain ontologies: These relations connect information about events of the domain of interest with map data and are



**Fig. 8.** Interconnections of the ontologies of the MSM

extremely helpful for presentation and retrieval of multimedia documents with respect to the geographic information that they convey by linking the annotated parts of the multimedia documents to geographical map data. In particular, they combine information about athletics events with information about the geographic/geopolitical area that they have taken place. A characteristic example of this category of relations is the `aio : takesPlaceIn` relation which connects instances of concepts like `aio : AthleticsEvent`, `aio : AthleticsRound`, `aio : AthleticsTrial`, defined in AEO, to the location that they have taken place, e.g. to instances of concepts `gio : Stadium`, `gio : StadiumArea`, `gio : City`, `gio : Country` of the GIO. For example, in order to represent the fact that an instance of a Marathon event has taken place in a specific city, we would use the following assertion:

`aio : takesPlaceIn(aio : MarathonEvent1, gio : City1)`

- Relations among concepts of the multimedia and the domain ontologies: These relations connect structural aspects of multimedia objects with their domain specific content and are really indispensable for presentation and retrieval purposes of multimedia objects or entire documents with respect to end-user queries on the domain of interest. One characteristic relation of this kind is the `mco : depicts` relation which connects instances of multimedia segments, defined in the MCO, with instances of concepts defined in AEO or GIO. For example, we could use the relation `mco : depicts` to declare that a specific segment of a text denotes an instance of a stadium, or that a specific region of a still image denotes an instance of a person's face, using the following assertions:

`mco : depicts(mco : TextSegment2, gio : Stadium1)`

`mco : depicts(mco : StillRegion2, aio : PersonFace1)`

## 5 Representation of Uncertainty

In the previous section, we have shown how to provide a formal representation of multimedia semantics using ontology languages, and more precisely OWL and its underlying technology of Description Logics (DLs). Although DLs are significantly expressive, they feature limitations when it comes to modelling domains where *imperfect*, like *uncertain* or *vague/fuzzy* information is apparent. This is often the case with the task of knowledge-based multimedia processing and interpretation. More precisely, image and video analysis algorithms are usually based on statistical criteria, thus the results they provide also contain *confidence degrees*. Moreover, it is also usual that the information that exists in a multimedia document is inherently vague, like for example the color (red, very red, blue, etc.), the size (large, small, etc) or the shape (long, circular, rectangular, etc.) of a specific object.

The representation and management of imperfect, uncertain and/or vague knowledge, is a huge topic that has received tremendous interest in AI (expert systems, natural language processing and understanding, etc.), in database management systems (relational schemata, deductive databases, etc.), in the field of knowledge representation and reasoning in general (probabilistic logic, Dempster-Shafer theory, Bayesian inference, subjective logic, etc.), and so forth; see [29] for a list of applications of fuzzy sets and fuzzy logic. Corresponding approaches have been developed in the context of ontology languages that extend the underlying mathematical frameworks so as to allow the formal handling of imperfect knowledge. Relevant proposals in the literature, include probabilistic DLs [16], probabilistic OWL [12], possibilistic DLs [40], as well as fuzzy DLs and fuzzy OWL [50–52].

As the aforementioned extensions model different types of imprecision, their appropriateness for a given application depends on the particular semantics involved. In the case of confidence degrees encountered in image and video analysis, the imprecision semantics lie in the nature of “confidence” captured in the computed degrees. Approaches where concepts are detected on the grounds of perceptual similarity, imply a prototypical set of feature values that constitute a visual/perceptual definition of the concept. As the presence of a concept is determined based on the similarity of those values, concepts can be considered as fuzzy sets, where the similarity (distance) function plays the role of the membership function. Contrariwise, approaches that utilise concepts’ co-occurrence and correlation, pertain to a probabilistic/possibilistic interpretation of the associations between visual features and semantic concepts. Support Vector Machines [7] constitute a popular example of the former category, while Bayesian Nets [18] and Hidden Markov models fall in the latter.

Apparently, both types of imperfection pertain to the case of multimedia processing and interpretation, while the complementary aspects addressed, render each of them a crucial component towards complete and robust solutions. In this chapter though, we focus solely on handling the vagueness encompassed in the processing of multimedia content. Specifically, in the following, we go through the theory of fuzzy Description Logics, in order to provide an insight on how such



extended theories could be used to represent and reason with the imperfection of the processed multimedia documents. We will provide examples on how fuzzy DLs can be used and a short overview of tools that can be used in practical applications.

### 5.1 Fuzzy Extensions of OWL and DLs

As is the case with classical OWL and Description Logics, fuzzy Description Logics provide the notions of concepts (**C**), roles (**R**) and individuals (**I**) in order to represent the primitive concepts of our domain knowledge. So for example one can use the atomic (primitive) concepts **Blue**, **Large**, **Arm**, **Person**, **Car** in order to represent entities that are depicted in an image or video, primitive roles **hasColor**, **hasPart** to describe binary relations or individuals  $car_1$ ,  $person_2$  in order to represent the specific objects of a specific image. Then concepts, roles and individuals are used together with the constructors in order to devise more complex concepts. For example using the construction of conjunction ( $\sqcap$ ) we can describe the concept of blue cars by writing  $Car \sqcap BlueColored$ , or we can use the constructor of existential restrictions ( $\exists$ ) together with the conjunction constructor to describe the notion of a clouded sky as  $ClearSky \sqcap \exists contains.Cloud$ . More formally, fuzzy-*SHOIN*-concepts and roles are defined as follows.

**Definition 1.** Let  $RN \in \mathbf{R}$  be a role name and  $R$  be an *f-SHOIN*-role. *f-SHOIN*-roles are defined by the abstract syntax:  $R ::= RN \mid R^-$ , where  $R^-$  denotes the inverse of the role  $R$ . The inverse relation of roles is symmetric, and to avoid considering roles such as  $R^{--}$ , we define a function *Inv* which returns the inverse of a role, more precisely  $Inv(RN) := RN^-$  and  $Inv(RN^-) := RN$ . The set of *f-SHOIN*-concepts is the smallest set such that

1. every concept name  $CN \in \mathbf{C}$  is an *f-SHOIN*-concept,
2. if  $o \in \mathbf{I}$  then  $\{o\}$  is an *f-SHOIN*-concept,
3. if  $C$  and  $D$  are *f-SHOIN*-concepts,  $R$  an *f-SHOIN*-role,  $S$  a simple<sup>12</sup> *f-SHOIN*-role and  $p \in \mathbb{N}$ , then  $(C \sqcup D)$ ,  $(C \sqcap D)$ ,  $(\neg C)$ ,  $(\forall R.C)$ ,  $(\exists R.C)$ ,  $(\geq pS)$  and  $(\leq pS)$  are also *f-SHOIN*-concepts.

As we can see, *f-SHOIN*-concepts are fairly standard with respect to classical *SHOIN*-concepts and roles [2].

Similarly to classical DLs, in fuzzy DLs one can also define new concepts using the notion of concept axioms. Let  $C$  and  $D$  be *f-SHOIN*-concepts. Concept axioms of the form  $C \sqsubseteq D$  are called *inclusion axioms*, while concept axioms of the form  $C \equiv D$  are called *equivalence axioms*. Thus, we can describe intentional knowledge in the same way as the standard OWL language. For example we can provide the axiom:

$$CloudedSky \equiv ClearSky \sqcap \exists contains.Cloud$$

<sup>12</sup> A role is called *simple* if it is neither transitive nor has any transitive sub-roles. Allowing only simple roles to participate in number restrictions is crucial in order to get a decidable logic [22].

that defines the new concept of clouded sky. A similar case can be made about roles, where we can capture partonomic relations with the aid of inverse roles, transitive role axioms, and role inclusion axioms.

The power of fuzzy Description Logics comes into play when one wants to represent instance assertions (individual axioms). More precisely, fuzzy ontology languages allow one to represent the degree to which an individual belongs to a concept. For example we could state that object  $obj_1$  is **Blue** to a degree 0.9, or that it is **Large** to a degree 0.7. For these reasons in fuzzy ontologies, the notion of an assertion (or fact) is extended to that of a *fuzzy assertion* (or fuzzy fact) [52]. Fuzzy assertions are of the form  $(a : C) \geq n_1, (a : D) = n_2 ((a, b) : R) \geq n_3$  and so on, where  $C, D$  are concepts (classes) and  $n_1, n_2, n_3$  are degrees from the unit interval  $([0,1])$ .

A *fuzzy ontology*  $\mathcal{O}$  consists of a set of the above axioms.

As with classical DLs, fuzzy-DLs provide for a formal meaning to their building blocks, thus they constitute a well-defined and semantic way of representing (vague) knowledge. Such fuzzy semantics are provided with the aid of the (relatively) standard notion of *fuzzy interpretation* introduced in [52]. Roughly speaking, concepts are interpreted as fuzzy sets and roles as fuzzy relations [29]. For example, considering the object  $Rome^{\mathcal{I}}$ , that denotes the city, and the fuzzy set  $HotPlace^{\mathcal{I}}$  that denotes hot places, a fuzzy set has the form  $HotPlace^{\mathcal{I}}(Rome^{\mathcal{I}}) = 0.7$ , meaning that rome is a hot place to a degree equal to 0.7. Fuzzy interpretations can be extended to interpret complex *f-SHOIN*-concepts and roles, with the aid of the fuzzy set theoretic operations defined and investigated in the area of fuzzy set theory [29]. The interested reader can refer to the wealth of fuzzy DL literature for the complete set of semantics [49, 51–53].

As with classical DLs, fuzzy DLs provide a set of inference services which can be used to query fuzzy ontologies. Interestingly, today there exist reasoning algorithms [50, 52] as well as practical reasoning systems. One such a system is FiRE (Fuzzy Reasoning Engine) which can be found at <http://www.image.ece.ntua.gr/~nsimou/FiRE> together with installation instructions and examples. FiRE currently supports  $f_{KD}$ -*SHIN*, i.e. fuzzy-*SHOIN* without the nominal constructor.

Let us now see a specific example of the use of fuzzy DLs in the task of knowledge based multimedia processing. Consider for example pictures that depict athletics, like athletes performing high jump, pole vault, discus throw attempts etc. A segmentation algorithm is applied on such images to identify the different objects that are depicted as image segments. For each segment we can then extract their MPEG-7 visual descriptors. These are numerical values which provide information about the texture, shape and color of a region. One could use such values in order to *move* from low-level descriptions to more high-level ones. For example, if the green component in the RGB color model of region 1 ( $reg_1$ ) is equal to 243, we can be based on a mapping (fuzzy partition) function [29] and deduce that  $reg_1$  is **GreenColored** to a degree at least 0.8. On the other hand another region with a green component of 200 could be **GreenColored** to a degree 0.77. Similarly, we can extract additional fuzzy assertions using other MPEG-7

descriptors, like texture or shape. Subsequently, we can construct an ontology which could be used to provide semantic descriptions (definitions) of the optical objects that exist in our image. A sample ontology could be the following:

HorizontalBar  $\equiv$  RectangularShaped  $\sqcap$  Elongated  $\sqcap$  HorizontallyDirected,  
 LandingPit  $\equiv$  BrownColored  $\sqcap$  CoarseTextured  $\sqcap$  RectangularShaped,  
 PoleVault  $\equiv$  AthleticEvent  $\sqcap$   $\exists$ hasPart.HorizontalBar  $\sqcap$   $\exists$ hasPart.Pole

Finally, using concept axioms such as the above ones together with fuzzy assertions created by mapping MPEG-7 features to fuzzy concepts and inference services of fuzzy DLs, we can extract all the implied knowledge for a specific image. The following table provides a few examples of initially extracted concepts from MPEG-7 descriptors and inferred concepts using fuzzy-DL reasoning.

**Table 1.** Semantic labelling

Region	Extracted Concept	Degree	Inferred Concept	Degree
<i>region<sub>1</sub></i>	RectangularShaped	0.69	HorizontalBar	0.69
	Elongated	0.85		
	HorizontallyDirected	0.80		
<i>region<sub>2</sub></i>	BrownColored	0.85	LandingPit	0.73
	CoarseTextured	0.73		
	RectangularShaped	0.91		

More extended examples on the use of fuzzy-DLs in the context of multimedia processing and interpretation can be found in [9, 46].

## 6 Conclusions and Open Issues

Today a vast amount of multimedia documents exist in multimedia databases of TV channels, production companies, museums, film companies, sports federations, etc. But all this cultural heritage is almost completely lost or never reused since accessing them is highly inflexible, inefficient and extremely expensive. In most cases these multimedia documents lay in legacy systems free of content descriptions and searching for documents which depict particular content may take hours or even days. To solve this problem one has to provide appropriate ways to represent the multimedia content in a semantically rich and machine understandable way.

Representation of multimedia content semantics is one of the most important issues in the multimedia research community. Firstly, having the description of the content in a semantically rich form enables us to provide semantic access to multimedia documents. Moreover, with the advent of the semantic web publishing such content on the web enables interoperability and reuse of multimedia information. Additionally, the use of semantic technologies gives new possibilities in using inference and reasoning services for the tasks of assisting several

multimedia related tasks, like multimedia analysis. Several proposals for representing the semantics of multimedia documents or for using semantic technologies for performing knowledge-based multimedia processing and interpretation have been proposed in the literature. All these approaches have followed different modelling choices due to the fact that the resulting ontologies were used in different application scenarios or domains.

In the current chapter we have reported on our results of developing ways to represent multimedia content semantics within the BOEMIE project. We have presented four, interconnected ontologies, namely the Athletics Events Ontology (AEO), the Geographic Information Ontology (GIO), the Multimedia Content Ontology (MCO) and the Multimedia Descriptor Ontology (MDO). These ontologies are purposed to capture and represent the information that exists in different parts of multimedia documents. More precisely, the MCO ontology is purposed to represent the structural information of multimedia documents, the MDO ontology the low-level numerical information that is extracted by multimedia analysis modules, while AEO and GIO high-level knowledge about the domain that the specific multimedia documents depict. All aforementioned ontologies, although independently developed, are interlinked using several spatiotemporal relations in order to provide a global framework for representing the semantics of multimedia content. Furthermore, given the imprecision inherent both in the information conveyed by multimedia content and in multimedia analysis and processing, non-standard technologies based on fuzzy extensions to DLs, have been presented as possible means to represent and manage such type of information.

Compared with the relevant literature, the proposed Multimedia Semantic Model, and the opportunities for its extension through the use of fuzzy DLs for the formal handling of uncertainty, brings a number of additional advantages. First, the proposed framework addresses in an integral manner the core issues involved in the interpretation and semantic management of multimedia content, namely the representation and linking of domain with media specific notions in a manner that enables the utilisation of reasoning in a semantically rich way, the handling of imperfect knowledge in terms of vagueness, and the seamless interchange, sharing and reuse of both the background knowledge as well as the resulting semantic interpretations. The specialised ontology patterns proposed for the representation of primitive concepts extracted through analysis and of more complex ones, derivable by means of reasoning, constitute a significant contribution towards the first issue. The clean modelling and axiomatised media specific ontologies, especially with respect to the representation of content structure, constitute the main contribution compared to the existing MPEG-7 multimedia ontologies. Moreover, the advantages from the integral, multiple modalities, view taken on the issues involved, is further strengthened by the modular architecture and the extensible design followed.

Finally, based on the experiences drawn, future research directions and open issues may be summarised in the following.

- The multimedia ontologies have been developed with the aim to live in an evolving environment where apart from representation and reasoning, they will be used for the tasks of presentation, retrieval, learning and evolution. Thus, it remains to evaluate if the proposed architecture is sufficient to support also such tasks.
- First results have shown that DL based ontologies together with rule language, like DL-safe rules are expressive enough to be used for the task of multimedia interpretation and reasoning. On the other hand more extensive evaluation has to be performed in order to estimate the deficiencies and assess the value of DLs for such tasks.
- Currently, although a number of spatiotemporal relations have been used inference services do not go beyond traditional DLs. In other words true spatiotemporal reasoning is not supported. Obviously, such services are important for video analysis and representation as well as for representing image relations. It is an open issue on how existing spatiotemporal extensions to DL languages can be used for representing such multimedia content.

## References

1. Arndt, R., Troncy, R., Staab, S., Hardman, L., Vacura, M.: COMM: Designing a Well-Founded Multimedia Ontology for the Web. In: Aberer, K., Choi, K.-S., Noy, N., Allemang, D., Lee, K.-I., Nixon, L.J.B., Golbeck, J., Mika, P., Maynard, D., Mizoguchi, R., Schreiber, G., Cudré-Mauroux, P. (eds.) ASWC 2007 and ISWC 2007. LNCS, vol. 4825, pp. 30–43. Springer, Heidelberg (2007)
2. Baader, F., Nutt, W.: Basic description logics. In: Baader, F., Calvanese, D., McGuinness, D., Nardi, D., Patel-Schneider, P.F. (eds.) *The Description Logic Handbook: Theory, Implementation, and Applications*, pp. 43–95. Cambridge University Press, Cambridge (2003)
3. Berners-Lee, T., Hendler, J., Lassila, O.: The semantic web. *Scientific American* (2001)
4. Bertini, M., Del Bimbo, A., Torniai, C.: Enhanced ontologies for video annotation and retrieval. In: *Multimedia Information Retrieval*, pp. 89–96 (2005)
5. Bloehdorn, S., Petridis, K., Saathoff, C., Simou, N., Tzouvaras, V., Avrithis, Y., Handschuh, S., Kompatsiaris, Y., Staab, S., Strintzis, M.G.: Semantic annotation of images and videos for multimedia analysis. In: Gómez-Pérez, A., Euzenat, J. (eds.) *ESWC 2005*. LNCS, vol. 3532, pp. 592–607. Springer, Heidelberg (2005)
6. Brickley, D., Guha, R.V.: OWL Web Ontology Language Overview, W3C Recommendation February 10 (2004), <http://www.w3.org/TR/owl-features/>
7. Burges, C.J.C.: A Tutorial on Support Vector Machines for Pattern Recognition. *Data Mining and Knowledge Discovery* 2(2), 121–167 (1998)
8. Dasiopoulou, S., Heinecke, J., Saathoff, C., Strintzis, M.G.: Multimedia reasoning with natural language support. In: 1st IEEE International Conference on Semantic Computing (ICSC), Irvine, CA, USA (2007)
9. Dasiopoulou, S., Kompatsiaris, I., Strintzis, M.G.: Investigating fuzzy DLs-based reasoning in semantic image analysis. In: *Multimedia Tools Appl., S.I. Semantic Multimedia* (2009), doi:10.1007/s11042-009-0393-6

10. Dasiopoulou, S., Mezaris, V., Kompatsiaris, I., Papastathis, V.K., Strintzis, M.G.: Knowledge-assisted semantic video object detection. *IEEE Trans. Circuits Syst. Video Techn.* 15(10), 1210–1224 (2005)
11. Dasiopoulou, S., Tzouvaras, V., Kompatsiaris, I., Strintzis, M.G.: Enquiring MPEG-7 based multimedia ontologies. *Multimedia Tools and Applications* 46(2-3), 331–370 (2010)
12. Ding, Z., Peng, Y.: A Probabilistic Extension to Ontology Language OWL. In: *Proceedings of the 37th Hawaii International Conference on System Sciences (HICSS-37)*, Big Island, Hawaii, p. 10 (January 2004)
13. Gangemi, A.: Ontology design patterns for semantic web content. In: Gil, Y., Motta, E., Benjamins, V.R., Musen, M.A. (eds.) *ISWC 2005*. LNCS, vol. 3729, pp. 262–276. Springer, Heidelberg (2005)
14. Gangemi, A., Guarino, N., Masolo, C., Oltramari, A., Schneider, L.: Sweetening ontologies with DOLCE. In: Gómez-Pérez, A., Benjamins, V.R. (eds.) *EKAUW 2002*. LNCS (LNAI), vol. 2473, pp. 166–181. Springer, Heidelberg (2002)
15. Gangemi, A., Mika, P.: Understanding the Semantic Web through descriptions and situations. In: Chung, S., Schmidt, D.C. (eds.) *CoopIS 2003, DOA 2003, and ODBASE 2003*. LNCS, vol. 2888, pp. 689–706. Springer, Heidelberg (2003)
16. Giugno, R., Lukasiewicz, T.:  $P\text{-}\mathcal{SHOQ}(\mathbf{D})$ : A probabilistic extension of  $\mathcal{SHOQ}(\mathbf{D})$  for probabilistic ontologies in the semantic web. In: Flesca, S., Greco, S., Leone, N., Ianni, G. (eds.) *JELIA 2002*. LNCS (LNAI), vol. 2424, pp. 86–97. Springer, Heidelberg (2002)
17. Gruber, T.R.: Toward principles for the design of ontologies used for knowledge sharing. In: *Inter. Workshop on Formal Ontology*, Padua, Italy (March 1993)
18. Heckerman, D.: A tutorial on learning with bayesian networks. *Learning in Graphical Models*, 301–354 (1998)
19. Hollink, L., Little, S., Hunter, J.: Evaluating the application of semantic inferencing rules to image annotation. In: *3rd International Conference on Knowledge Capture (K-CAP)*, Banff, Alberta, Canada, pp. 91–98 (2005)
20. Hollink, L., Worring, M.: Building a visual ontology for video retrieval. In: *Proc. 13th ACM International Conference on Multimedia*, Singapore, November 6–11, pp. 479–482 (2005)
21. Horrocks, I., Patel-Schneider, P.F.: Reducing OWL entailment to Description Logic satisfiability. *J. Web Sem.* 1(4), 345–357 (2004)
22. Horrocks, I., Sattler, U., Tobies, S.: Practical reasoning for expressive description logics. In: Ganzinger, H., McAllester, D., Voronkov, A. (eds.) *LPAR 1999*. LNCS (LNAI), vol. 1705, pp. 161–180. Springer, Heidelberg (1999)
23. Hu, B., Dasmahapatra, S., Lewis, P.H., Shadbolt, N.: Ontology-based medical image annotation with Description Logics. In: *Proc. Inter. Conference on Tools with Artificial Intelligence (ICTAI)*, Sacramento, California, November 3–5 (2003)
24. Hudelot, C., Thonnat, M.: A cognitive vision platform for automatic recognition of natural complex objects. In: *ICTAI*, pp. 398–405 (2003)
25. Hunter, J.: Adding Multimedia to the Semantic Web: Building an MPEG-7 Ontology. In: *Proc. The First Semantic Web Working Symposium, SWWS 2001*. Stanford University, California (July 2001)
26. Hunter, J., Drennan, J., Little, S.: Realizing the hydrogen economy through Semantic Web technologies. *IEEE Intelligent Systems* 19(1), 40–47 (2004)
27. Hyvonen, E., Styrman, A., Saarela, S.: Ontology-based image retrieval. In: *XML Finland Conference*, October 21–22, pp. 15–27 (2002)

28. Jaimes, A., Tseng, B.L., Smith, J.R.: Modal keywords, ontologies, and reasoning for video understanding. In: Bakker, E.M., Lew, M., Huang, T.S., Sebe, N., Zhou, X.S. (eds.) CIVR 2003. LNCS, vol. 2728, pp. 248–259. Springer, Heidelberg (2003)
29. Klir, G.J., Yuan, B.: Fuzzy Sets and Fuzzy Logic: Theory and Applications. Prentice-Hall, Englewood Cliffs (1995)
30. Goodall, S., Grimwood, P., Kim, S., Lewis, P., Martinez, K., Stevenson, A., Addis, M., Boniface, M.: Sculpteur: Towards a new paradigm for multimedia museum information handling. In: Fensel, D., Sycara, K., Mylopoulos, J. (eds.) ISWC 2003. LNCS, vol. 2870, pp. 582–596. Springer, Heidelberg (2003)
31. Maillot, N., Thonnat, M.: A weakly supervised approach for semantic image indexing and retrieval. In: Leow, W.-K., Lew, M., Chua, T.-S., Ma, W.-Y., Chaisorn, L., Bakker, E.M. (eds.) CIVR 2005. LNCS, vol. 3568, pp. 629–638. Springer, Heidelberg (2005)
32. Martínez, J.M.: MPEG-7: Overview of MPEG-7 Description Tools, Part 2. IEEE MultiMedia 9(3), 83–93 (2002)
33. Meghini, C., Sebastiani, F., Straccia, U.: A model of multimedia information retrieval. *Journal of the ACM* 48(5), 909–970 (2001)
34. Möller, R., Neumann, B.: Ontology-based reasoning techniques for multimedia interpretation and retrieval. In: *Semantic Multimedia and Ontologies: Theory and Applications* (2008) (to appear)
35. MPEG-21. Multimedia Framework (MPEG-21) - ISO/IEC TR 21000-1:2004 (2002)
36. Nack, F., van Ossenbruggen, J., Hardman, L.: That obscure object of desire: Multimedia metadata on the Web, Part 2. IEEE MultiMedia 12(1), 54–63 (2005)
37. Neumann, B., Möller, R.: On scene interpretation with description logics. Technical Report FBI-B-257/04 (2004)
38. Oberle, D., Ankolekar, A., Hitzler, P., Cimiano, P., Sintek, M., Kiesel, M., Mougouie, B., Baumann, S., Vembu, S., Romanelli, M.: DOLCE ergo SUMO: On foundational and domain models in the SmartWeb Integrated Ontology (SWIntO). *J. Web Sem.* 5(3), 156–174 (2007)
39. Petridis, K., Anastasopoulos, D., Saathoff, C., Timmermann, N., Kompatsiaris, Y., Staab, S.: M-ontoMat-annotizer: Image annotation linking ontologies and multimedia low-level features. In: Gabrys, B., Howlett, R.J., Jain, L.C. (eds.) KES 2006. LNCS (LNAI), vol. 4253, pp. 633–640. Springer, Heidelberg (2006)
40. Qi, G., Pan, J.Z., Ji, Q.: Extending description logics with uncertainty reasoning in possibilistic logic. In: Mellouli, K. (ed.) ECSQARU 2007. LNCS (LNAI), vol. 4724, pp. 828–839. Springer, Heidelberg (2007)
41. Celma, O., García, R.: Semantic Integration and Retrieval of Multimedia Metadata. In: *Proc. International Semantic Web Conference (ISWC)*, Galway, Ireland, November 6–10 (2005)
42. Celma, O., Halaschek-Wiener, C., Mannens, E., Troncy, R., Boll, S., Burger, T.: Multimedia vocabularies on the Semantic Web. In: *W3Cu Incubator Group Report*, July 24 (2007)
43. Schober, J.P., Hermes, T., Herzog, O.: Content-based image retrieval by ontology-based object recognition. In: Haarslev, V., Lutz, C., Möller, R. (eds.) *Proc. Workshop on Applications of Description Logics*, Ulm, Germany (2004)
44. Schreiber, A.T., Dubbeldam, B., Wielemaker, J., Wielinga, B.J.: Ontology-based photo annotation. *IEEE Intelligent Systems* 16(3), 66–74 (2001)
45. Di Sciascio, E., Donini, F.: Description logics for image recognition: a preliminary proposal. In: *Proceedings of the International Workshop on Description Logics (DL 1999)* (1999)

46. Simou, N., Athanasiadis, T., Tzouvaras, V., Kollias, S.: Multimedia reasoning with *f-SHIN*. In: 2nd International Workshop on Semantic Media Adaptation and Personalization, London, United Kingdom, December 17-18 (2007)
47. Smeulders, A.W.M., Worring, M., Santini, S., Gupta, A., Jain, R.: Content-based image retrieval at the end of the early years. *IEEE Trans. Pattern Anal. Mach. Intell.* 22(12), 1349–1380 (2000)
48. Staab, S., Studer, R. (eds.): *Handbook on Ontologies*, 2nd edn. Springer, Heidelberg (2009)
49. Stoilos, G., Stamou, G.: Extending fuzzy description logics for the semantic web. In: *Proceedings of the 3rd International Workshop on OWL Experiences and Direction (OWL ED 2007)* (2007)
50. Stoilos, G., Stamou, G., Tzouvaras, V., Pan, J.Z., Horrocks, I.: Reasoning with very expressive fuzzy description logics. *Journal of Artificial Intelligence Research* 30(5), 273–320 (2007)
51. Stoilos, G., Stamou, G., Tzouvaras, V., Pan, J.Z., Horrocks, I.: Fuzzy OWL: Uncertainty and the semantic web. In: *Proc. of the International Workshop on OWL: Experiences and Directions* (2005)
52. Straccia, U.: Reasoning within fuzzy description logics. *Journal of Artificial Intelligence Research* 14, 137–166 (2001)
53. Straccia, U.: Towards a fuzzy description logic for the semantic web. In: Gómez-Pérez, A., Euzenat, J. (eds.) *ESWC 2005. LNCS*, vol. 3532, pp. 167–181. Springer, Heidelberg (2005)
54. Troncy, R., Celma, O., Little, S., GarcíaGarcía, R., Tsinaraki, C.: Mpeg-7 based Multimedia Ontologies: Interoperability Support or Interoperability Issue? In: *Proc. 1st Workshop on Multimedia Annotation and Retrieval enabled by Shared Ontologies (MARESO)*, Genova, Italy, pp. 2–16 (2007)
55. Tsinaraki, C., Polydoros, P., Christodoulakis, S.: Integration of OWL ontologies in MPEG-7 and TV-anytime compliant semantic indexing. In: Persson, A., Stirna, J. (eds.) *CAiSE 2004. LNCS*, vol. 3084, pp. 398–413. Springer, Heidelberg (2004)
56. Tsinaraki, C., Polydoros, P., Christodoulakis, S.: Interoperability support between mpeg-7/21 and OWL in ds-mirf. *IEEE Trans. Knowl. Data Eng.* 19(2), 219–232 (2007)
57. van Ossenbruggen, J., Nack, F., Hardman, L.: That obscure object of desire: Multimedia metadata on the web, part 1. *IEEE MultiMedia* 11(4), 38–48 (2004)