MOWL: An Ontology Representation Language for Web based Multimedia Applications

ANUPAMA MALLIK, Indian Institute of Technology, Delhi
HIRANMAY GHOSH, Tata consultancy Services, Gurgaon
GAURAV HARIT, Indian Institute of Technology, Jodhpur
SANTANU CHAUDHURY, Indian Institute of Technology, Delhi

Several multimedia applications need to reason with concepts and their media properties in specific domain contexts. Media properties of concepts exhibit some unique characteristics that cannot be dealt with conceptual modeling schemes followed in the existing ontology representation and reasoning schemes. We have proposed a new perceptual modeling technique for reasoning with media properties observed in multimedia instances and the latent concepts. Our knowledge representation scheme uses a causal model of the world where concepts manifest in media properties with uncertainties. We introduce a probabilistic reasoning scheme for belief propagation across domain concepts through observation of media properties. In order to support the perceptual modeling and reasoning paradigm, we propose a new ontology language, Multimedia Web Ontology Language (MOWL). Our primary contribution in this paper is to establish the need for the new ontology language and to introduce the semantics of its novel language constructs. We establish the generality of our approach with two disparate knowledge-intensive applications involving reasoning with media properties of concepts.

Categories and Subject Descriptors: I.2.4 [Computing Methodologies]: ARTIFICIAL INTELLIGENCE—Knowledge Representation Formalisms and Methods

General Terms: Theory, Standardization

Additional Key Words and Phrases: Multimedia Web Ontology Language, OWL Ontology, Bayesian networks, Video Annotation

ACM Reference Format:
DOI: http://dx.doi.org/10.1145/0000000.0000000

1. INTRODUCTION

A large part of multimedia data in various web-based repositories and social networks remains to be adequately annotated. Such data need to be semantically interpreted in various application contexts. Domain knowledge plays a crucial role in semantic data processing in knowledge-intensive domains.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies show this notice on the first page or initial screen of a display along with the full citation. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, to republish, to post on servers, to redistribute to lists, or to use any component of this work in other works requires prior specific permission and/or a fee. Permissions may be requested from Publications Dept., ACM, Inc., 2 Penn Plaza, Suite 701, New York, NY 10121-0701 USA, fax +1 (212) 869-0481, or permissions@acm.org.
© 2013 ACM 1544-3558/2013/04-ART1 $15.00
DOI: http://dx.doi.org/10.1145/0000000.0000000

Application of semantic web principles for multimedia data interpretations have so far been restricted to loose integration of domain based reasoning and computer vision techniques, where concept recognition using computer vision techniques does not benefit from the domain knowledge. In order to overcome this short-coming, we have earlier presented a new ontological framework for multimedia data interpretation [Ghosh et al. 2007]. It uses a causal model for media properties of concepts and uses an abductive reasoning scheme for concept recognition. The fundamental nature of such modeling and reasoning is quite different from Descriptions Logic (DL) based crisp and deductive reasoning that has guided the semantic model of traditional ontology representation languages, such as OWL. This motivates us to propose a new ontology representation language, the Multimedia Web Ontology Language (MOWL) with distinct semantics from existing ontology languages for knowledge-based multimedia applications. While we have illustrated the use of MOWL in a heritage application [Mallik et al. 2011], a formal presentation of the language and its semantics has been lacking. This paper is motivated to address that gap. Our primary contributions in this paper are a formal description of the language and its semantics, and to establish its suitability for the new modeling and reasoning paradigm. We illustrate the generality of the approach in coping up with disparate multimedia applications and different media forms with two distinct application examples in the domains of heritage preservation and product recommendation.

There has been significant research on the use of ontology to interpret media data over the last decade. Use of ontology to relate media properties and objects has been proposed in [Gangemi et al. 2002; Jiang et al. 2004]. The representation schemes in these approaches had been ad-hoc and the systems could not scale up beyond the boundaries of closed applications. Standardization of multimedia content description and usage specification language (MPEG-7 and MPEG-21) motivated the researchers to create ontologies to correlate standard media feature descriptors with domain ontologies. There have been two major approaches. In the first approach [Arndt et al. 2007; Hunter 2003; Saathoff and Scherp 2010], multimedia ontologies are used to define generic media features and their relations. These ontologies are used in conjunction with domain ontologies for interpreting multimedia data in an application context. In the second approach [Garcia and Celma 2005; Tsinaraki et al. 2007], MPEG-7 descriptors are translated to a library of valid ontology classes, which can be used in a domain ontology to represent media properties of objects. Both of these approaches use crisp Description Logic based reasoning, which cannot cope up with uncertainties in the multimedia domain. For example, [Bai et al. 2007] combines MPEG-7 descriptors and Temporal Descriptions Logic (TDL) to describe sports events, such as a GoalScore, in soccer matches, but achieves a poor recall because of inflexibility in such crisp specifications. Another recent approach involves associating media properties to machine detectable objects in a domain ontology, either in the form of visual prototypes [Bertini et al. 2009], or as a set of media features organized as a naïve Bayesian network [Nikolopoulos et al. 2011]. Taking a step further, [Dasiopoulou et al. 2010] proposes an uncertain reasoning for structural composition of simple scenes, such as beach or landscape. In summary, traditional ontologies deal with conceptual models of domains, but do not provide tools to reason with the media properties of concepts. In order to overcome this shortcoming, the existing approaches for ontology based multimedia applications amalgamate computer vision techniques for concept recognition with ontologies for domain-specific conceptual reasoning. A major limitation in these approaches is that they cannot holistically reason with concepts and their media properties in the context of a domain. The disconnect between the conceptual and the perceptual worlds is responsible for what is described as the semantic gap in the literature. Our present work attempts to address this issue.

MOWL is a formal language to express and to reason with a novel representation scheme for multimedia ontology. We provide a philosophical background and establish the requirements for this language in section 2. In section 3, we introduce the MOWL language constructs and the semantics of
the language. A scheme for formal representation of spatio-temporal relations between concepts has been proposed in section 4. In section 5, we provide details of reasoning in the MOWL inferencing framework. Section 6 presents a few application examples for MOWL. In section 7, we present our conclusions.

2. NEED FOR A MULTIMEDIA ONTOLOGY LANGUAGE

In our view, the key to semantic processing of media data lies in being able to reason with media properties of concepts in a domain context. This amounts to extending the ontological description of a domain from conceptual level to the perceptual properties of concepts. Though apparently disconnected, the conceptual and the perceptual worlds are closely connected in reality. Concepts are formed in human minds through a complex refinement process of personal perceptual experiences [Kangassalo 1991]. An observation of a real world object amounts to reception of perceptual signals through our sense organs and creation of a mental model of the object. An abstraction of many such mental models, arising out of several observation instances, gives rise to concepts. A mental analysis of similarities and dissimilarities between the perceptual properties of concepts give rise to concept taxonomy, which is a key ingredient of a domain ontology. For example, observation of a number of historical monuments, and comparison of their structural similarities, gives rise to concept class “monument” and its sub-classes, such as “fort”, “palace” and “tomb”. The basic mechanism of concept recognition in the media world can be viewed as a corollary to the genesis of concepts. When we “look for” an instance of a concept in a multimedia document, we expect some perceptual patterns to appear. The observed media patterns in a multimedia document provide evidence for the concept to be recognized. For example, an Indian medieval monument is recognized by observing its characteristic visual patterns, such as a circular dome and long vertical minarets in an image instance.

This observation motivates us to use a causal model of the world for reasoning with concepts and their media properties. In our model, the real world concepts cause some media patterns to manifest in the media instances where they occur. The concepts can be recognized with the evidence provided by the observed media patterns. While it is possible to associate media properties with concepts through ontology languages like OWL [Garcia and Celma 2005; Tsinaraki et al. 2007], it does not imply the causal model and does not support the associated abductive reasoning. This motivates us to define the new language MOWL, where we provide specific language constructs with causal connotations to associate media properties to concepts. These relations are generally marked with uncertainties, because of inherent differences across the concept instances and their media manifestations. We have represented this uncertainty using Conditional Probability Tables (CPT’s) as in [Ding and Peng 2004]. However, there is a major difference in the way that the probabilities have been used with MOWL. We observe that the media properties of a concept are often influenced by those of some related concepts in the domain. For example, an ontology on heritage monuments may encode the knowledge that the “Tajmahal” is an instance of “Indian medieval monuments” and that it is built with “marble”. This background knowledge suggests that a visual portrayal of the Tajmahal is likely to exhibit the general structural composition of Indian medieval monuments, as well as the color and texture properties of the stone. On the other hand, the “Tajmahal” being a typical instance of a tomb from the Mughal era, an example image of Tajmahal is a valid example for a tomb of that period. Thus, media properties propagate across related concept nodes related in certain ways in a multimedia ontology as shown in figure 1. This is a unique requirement for multimedia data interpretation and is quite distinct from the property inheritance rule implied by hierarchical relationship in a traditional ontology. We have defined a subclass of properties in MOWL to indicate such relations. We reason with the ontology to derive a Bayesian network that relates concepts with their possible media manifestations, including
those propagated from the related concepts. This derived Bayesian network is called an *Observation Model* and is used for recognition and association of concepts in specific application contexts.

A further requirement for the multimedia ontology representation language is that it should be possible to represent the media properties of concepts in different ways and at different levels of abstractions. For example, the shape of a dome can be specified either with an abstract geometric description or with a set of example images. A dance posture may best be specified by the binary output of a specialized feature classifier. A multimedia ontology language needs to provide distinct semantics for such variations. Some of the complex media patterns characterizing a concept can be composed of simpler patterns arranged in space and time. For example, the event class “goal score” in a soccer match can be characterized by a temporal sequence of audio-visual units, where each of the visuals is a spatial composition of constituent media objects, such as a ball, the players and a goal-post. Like the media properties, the spatio-temporal relations for an event class are marked with uncertainties.

A multimedia ontology language should be capable of formal yet flexible description of the structural composition of complex media objects. We have adopted the representation for spatio-temporal relations as proposed in [Wattamwar and Ghosh 2008], which provides constructs for formal specification of spatio-temporal composition, to describe an event class. A fuzzy membership function is used to compute the belongingness of an event instance in an event class so described. This is a novelty as compared to the *crisp* temporal logic or *informal* MPEG-7 like spatial and temporal relations being used in some of the earlier works [Bai et al. 2007; Tsinaraki et al. 2007]. To illustrate the principles,

![Diagram of media properties and examples](image)

**Fig. 1:** Propagation of (a) media properties and (b) media examples

figure 2 depicts a small section of an ontology and a possible OM for a concept, the Tajmahal, derived from that ontology. The OM is organized as a Bayesian tree, with the concept Tajmahal placed at the root node. The expected media patterns, when the monument manifests in a media instance (image or video), appear at the leaf nodes of the tree. Note that the Tajmahal “inherits” the media properties of some related concepts in the domain. The structure of the Tajmahal has been specified as a spatial composition of its constituent media objects, namely a facade, a dome and the minarets. For recognizing the concept “Tajmahal” in a media document, the Bayesian tree is initialized and appropriate media detectors are run to discover the media patterns specified at the leaf nodes. When a pattern is discovered, the corresponding node is instantiated resulting in a belief revision in the network. The presence of the concept is inferred by virtue of the posterior probability of the root node as the cumulative result of such belief revisions.

The abductive reasoning system proposed by us is weaker than the deductive reasoning employed with conventional ontology schemes, but is essential for dealing with inherent uncertainties in the
observation of multimedia artifacts. The major advantage of the abductive reasoning system is that it can produce robust results with fewer and uncertain observations. Another advantage of this concept recognition scheme is the separation of the knowledge about media properties of concepts from the underlying collection characteristics. The media properties in the ontology should ideally cover varied example instances of concepts in different media forms. In general, multimedia features may not be restricted to visual and audio properties alone, but may include a wider variety of contextual data, such as textual annotations and sensor data. Thus, an OM, when constructed from the multimedia ontology, may exhibit a large degree of redundancy in terms of expected media patterns. A method to choose a set of observable nodes that can produce adequate belief value for a concept within the bounds of a specified computational cost leaf nodes, using a greedy algorithm has been discussed in [Chaudhury and Ghosh 2004]. This flexibility enables integration of multiple distributed collections under a common conceptual framework. We shall not deal with this aspect further in this paper.

3. LANGUAGE CONSTRUCTS IN MOWL

MOWL distinguishes between two types of entities, namely (a) the concepts that represent the real world objects or events and (b) the media objects that represent manifestation of concepts in different media forms. Detection of the media objects leads to concept recognition. Like in other ontology languages, the concepts and the media objects can be organized in class hierarchies and can have properties. A class of properties relate the concepts with the media objects. These relations are causal and uncertain in nature. The uncertainties are captured in MOWL as conditional probability tables. The language constructs to express the same are described in section 3.2.

Another class of properties that relates two concepts signifies media property propagation. It is possible to define domain specific properties and associate them to this class. For example, a property built_with in heritage architecture domain can belong to this class signifying propagation of visual properties from a type of building material (Marble) to a constructed monument (Tajmahal). These media propagate properties are causal and form an important part of the Observation Model, thus assisting in concept recognition through detection of media properties that propagate from related concepts and media objects.
A media object can be specified in different ways, the simplest being a low-level media feature specification (e.g. DominantColor = "yellow") using standard MPEG-7 tools. Complex media features, such as a dynamic body posture, generally require dedicated pattern detectors. These media objects can have a procedural specification, for example, the specification of a service. Another type of media objects are specified by spatio-temporal arrangement of simpler media patterns. MOWL provides for a formal definition of such arrangement, which is consistent with and can be executed with an extended MPEG-7 Query Engine proposed in [Wattamwar and Ghosh 2008]. Media examples can also be associated with media object instances, when an example-based search is necessary for their detection. For instance, a face image can be associated with the media object humanFace in context of a specific person, to be used by a face recognition service. The language constructs of the multimedia web ontology language are based on RDF/RDF-S. A few new datatypes are also defined, that have specific contextual semantics. These language constructs are described with examples in the next few sections.

3.1 Concepts, Media Observables and Media Relations

MOWL declares two named classes `<mowl:Concept>` and `<mowl:MediaObject>` to represent its basic entities: the concepts and the media objects, which have some media-based properties. We have defined a few causal relations to associate media objects to other (more complex) media objects and concepts. The relation `<mowl:hasMediaObject>` assigns a media object to a concept or to another media object. This is a transitive relation, i.e. when C, M_1 and M_2 represent a concept and two media objects respectively, (C hasMO M_1 hasMO M_2) => C hasMO M_2. Media objects, media examples and media detectors are related to each other using `<mowl:hasMediaExample>`, `<mowl:hasMediaDetector>` and `<mowl:usesExample>`. We define a class of properties `<mowl:propagateMedia>` to signify media property propagation across two concepts. For example, if a certain dance form is generally accompanied with a certain music form, then a performance of that dance form is likely to manifest in the audio patterns of that music form. We can define a domain specific property `accompWith` as a sub-property of `<mowl:propagateMedia>`, when danceFormX accompWith musicFormY implies that media properties of musicFormY propagates to danceFormX.

Figure 3a depicts these relations.

Figure 4a depicts the association of some concepts, media objects and media examples for ICD. Figure 4b illustrates the corresponding MOWL snippet. The format for ontology visualization is such that the concepts are shown as ellipses and the media objects as rectangles. The media objects that can

---

1. The abstract syntax specification and RDF schema of these constructs is given in appendices available in the online version of this paper.

be detected by pattern detectors are shaded gray. We have used ASN [Patel-Schneider and Horrocks 2004] to describe the MOWL snippet. We have followed the convention through rest of the paper.

3.2 Uncertainty Specification in MOWL

Knowledge in several domains like medicine, art, heritage has media manifestations which are inherently uncertain, e.g. a lesion in the brain generally (but not always) manifests in a particular observable shape in an MRI image; conversely, appearance of that shape in an MRI image is not a conclusive evidence for a lesion. Thus, it is necessary to specify both the belief and plausibility of observing a media object effect) given a set of concepts, which can potentially cause them.

MOWL provides for specification of such uncertainties by defining a class <mowl:CPTable> for defining Conditional Probability Tables (CPTs). A CPT is associated with an effect node (typically, a media object node), and is conditioned by one or more cause nodes (typically, concept nodes) using <mowl:hasCPT> and <mowl:conditionedOn> properties respectively. Each of the cause and the effect nodes can assume two states, 1 (present) or 0 (absent). Each data row in CPT specifies the parent state and the probability values for the two states of the effect variable using a specific XML schema. Figure 5 depicts the definition of CPT for media object WhiteColor conditioned by the concept Marble.

It may be noted that the CPTs in MOWL do not indicate the probabilities for the relation to be true, but the probabilities for media property propagation across the concepts. If a CPT is not specified, we assume near-certain media property propagation approximating the CPT entries with some
default values. It has been observed that inferencing in Bayesian network is more sensitive to network topology than the conditional probability tables [Pradhan et al. 1996], when there are a large number of nodes, which is true for the OMs of concepts in real-life multimedia applications. Thus this approximation does not affect the reasoning adversely.

4. SPATIO-TEMPORAL RELATIONS

An event instance can generally be characterized with many facets, typically the 4 W’s (what, where, when and who) and its relationship with other events and its compositional structure [Shaw et al. 2009; Rafatirad et al. 2009; Scherp et al. 2009]. In context of MOWL, we deal with formal description of event classes that can be recognized with perceptual attributes discovered in multimedia documents. An event class is a subclass of media objects and its extrinsic relations with other entities in the domain can be described in MOWL like any other media objects. Its intrinsic properties are described through its structural composition, i.e. the constituent media objects and the spatio-temporal relations that exist between them.

One of the established tools for structural description of multimedia events is the Multimedia Description Scheme (MDS) of MPEG-7 [Salembier and Smith 2001]. It enables description of an instance of an event in a media instance in terms of simpler objects contained in the scene and the spatial and temporal relations between them. The spatio-temporal relation descriptors in MDS are normative open to interpretation by a system implementer. They cannot be used for formal ontological specifications. Moreover, the formal spatio-temporal model of an event need to account for uncertainties arising out of intrinsic differences across different instances of the events as well as difference in viewpoints [Randell et al. 2001]. In MOWL, we have adopted the representation for spatio-temporal relations as proposed in [Wattamwar and Ghosh 2008] for extending MPEG-7 Query Language. The scheme enables definition of spatial and temporal relations with formal semantics and uncertainties. We describe the representation scheme briefly and provide the language constructs in MOWL to express the same in the following subsections.

4.1 Spatio-temporal Model of an Event

The relation between two events occurring in a 4-dimensional event space \((x, y, z, t)\) can be represented with the interval relations [Papadias and Delis 2001] on projections of the events on the axes represented by \(R_x, R_y, R_z\) and \(R_t\) respectively, where each of the elements \(R_x, R_y, R_z, R_t\) can be represented as a five tuple of binary numbers representing truth values of the propositions representing intersection of the secondary event with five regions of interest of the primary event. While such representation can uniquely represent the relations between two events, both with convex contours, ambiguities in representing events with concave contours on the 4-D event space (e.g. when a rigid object has a sinusoidal motion) need to be resolved. To resolve such ambiguity, a set of six containment relations, \(R_{1100}\) (outside), \(R_{1011}\) (contains), \(R_{1111}\) (inside), \(R_{1110}\) (overlaps), \(R_{1110}\) (touching) and \(R_{0110}\) (skirting) have been defined where the quadruple of binary numbers represent the truth values of the containment propositions: \(p := A \setminus (A \cap B) \neq \phi\) (A is not completely included in B), \(q := B \setminus (A \cap B) \neq \phi\) (B is not completely included in A), \(r := A \cap B \neq \phi\) (A and B have some intersection), \(s := A \cap B \neq \phi\). Thus, we can define the spatio-temporal relations between two events without ambiguity, using a quintuple \(R = (R_x, R_y, R_z, R_t, R_c)\), where \(R_x, R_y, R_z\) and \(R_t\) denote the relative positions of the projections of the events in \(x, y, z\) and \(t\) axes respectively and \(R_c\) denotes the containment relation between the two. In general, the relationship between two events can be expressed as a logical conjunction of a subset of these relations.

\[\text{The operator } \cap^* \text{ denotes the regularized intersection [Foley et al. 1982] and is defined as } A \cap^* B = \text{ closure(}\text{interior}(A \cap B))\]
The spatio-temporal relations so defined are crisp and evaluate to either “true” or “false”. Multimedia data processing however requires some uncertain handling of such relations. For example, figure 6 (a) depicts a visual model of the Tajmahal as “dome \text{ between } \rightarrow \text{ minaret pair}” and figures 6 (b) and (c) shows two views of the monument from different angles. While the model clearly holds good for the first example, it is not so apparent in the second view. In order to account for such situations, the binary truth values of the intersection and containment propositions are replaced with fuzzy membership values. These values are derived from five fuzzy membership functions $\langle T, U, V, W, X \rangle$, associated with the five intersection propositions along each of the axes $x, y, z$ and $t$ and four fuzzy membership functions $\langle P, Q, R, S \rangle$ associated with the four containment propositions. The functions are generally defined in piecewise linear fashions over normalized interval axes, for each of the intersection and containment propositions.

The resulting descriptions for complex media objects can be used for their recognition, despite differences in media manifestations, using MPEG-7 query engine as proposed in [Wattamwar and Ghosh 2008]. We illustrate the principles using a description for the structural composition of Tajmahal as depicted in figure 6. Let us define an 1D relation \text{ between} with the fuzzy functions $T, V$ and $X$ ($U$ and $W$ being don’t cares in this example) as follows:

$$T ::= (\infty, 1), (-0.1, 1), (0.1, 0), (\infty, 0); \quad V ::= (-\infty, 0), (-0.1, 0), (0.1, 1), (0.9, 1), (1.1, 0), (\infty, 0);$$

$$X ::= (-\infty, 0), (-0.1, 0), (0.1, 1), (\infty, 1)$$

These functions are pictorially depicted in figure 6 (d). In this figure the \text{minaret pair} has been chosen as the primary interval $[a, b]$. The intervals $[c, d]$ and $[e, f]$ represent the positions of the dome with respect to the \text{minaret pair} for the images in figures 6 (b) and (c) respectively. The membership value of the function \text{between} is defined as $b = \neg t \wedge v \wedge \neg x$, where $t, v$ and $x$ represent the fuzzy membership values computed from the functions $T, V$ and $X$ respectively.

While the value for $b$ evaluates to 1.0 for figure 6 (b), it evaluates to 0.6 for figure 6 (c). Thus, the second view is also recognized to conform to the description, though with some uncertainty. The flexibility in recognition is quite useful for recognizing complex media objects.

4.2 Language Constructs

Complex media objects are composed of simpler media objects in different media forms and interconnected through spatio-temporal relations. MOWL defines $<$mowl:ComplexObject$>$ as a subclass of $<$mowl:MediaObject$>$ to represent such objects. To define a complex object formally, we define three properties for this class as shown in figure 3b. The properties $<$mowl:hasSubject$>$ and $<$mowl:hasObject$>$ associate other media objects with a complex media object as its subject and object. The subject and the object of a complex object can themselves be other complex objects. This recursive definition helps in defining arbitrarily complex objects where more than one spatial and temporal relation are involved.

$^3$The relation \text{between} defined in this example is equivalent to $R_{01x101} = R_{000101}R_{001100}R_{011000}R_{011100}$. 

The property `<mowl:hasPredicate>` associates a spatial or temporal predicate with a complex object. MOWL defines a named class `<mowl:STPredicate>` to represent such predicates. Spatio-temporal relations like inside and followedBy can be defined as instances of this class. These predicates being relations, should ideally be defined as a property in an ontology language, but as discussed in the previous section, spatio-temporal relations in MOWL are represented by a 5-tuple \( \langle R_x, R_y, R_z, R_t, R_c \rangle \) and a set of fuzzy functions for each. Thus these relations need to be reified and described with the help of their own properties. As no elegant reification is available with current RDF/OWL representation, we have defined these predicates as a class instead of a property in MOWL. The class `<mowl:FuzzyRelation>` is used to define the relations \( R_x \ldots R_t \) and `<mowl:FuzzyContainment>` is used to define \( R_c \). The fuzzy membership functions \( T, U, V, W \) and \( \chi \) can be defined for each of the Fuzzy relations \( R_x \ldots R_t \), by data-type properties `<mowl:hasT>`, `<mowl:hasU>` and so on, and those associated with \( R_c \) with data-type properties `<mowl:hasP>`, `<mowl:hasQ>` and so on. The range of these data-type properties is a vector of tuples \((t, v)\) representing the five regions of interest, where \( v \) is the fuzzy membership value at the point \( t \) on the axis. While defining the constraints for the fuzzy functions, we find that \( t \) can take values \(-\infty, +\infty\) or any float value, whereas \( 0 \leq v \leq 1 \). Each fuzzy function like \( T \) can be defined as a vector of many such tuples. MOWL defines some special XML schema data types to define these fuzzy functions values: `<xsd:FuzzyVector>` defines a vector of tuples; `<xsd:tupleType>` defines a data type to represent a tuple \((t, v)\); `<xsd:axisPtType>` defines a data type which can be a string with values “infinity” or “+infinity” or a decimal for \( t \) value of the tuple \((t, v)\); and `<xsd:probValType>` data type defines a decimal value between 0 and 1, both inclusive, for \( v \) value of the tuple \((t, v)\).

Figure 7a gives a graphical view of the spatial event `dome between minarets`, which has been described in section 4.1. Figure 8a shows a more complex goal score event in a soccer match, involving a recursive definition with temporal and spatial predicates. Figures 7b and 8b depict MOWL encoding for the respective events.

5. MOWL INFERENCING FRAMEWORK

There are two stages of reasoning in the MOWL inferencing framework. In the first stage, the neighbourhood of the concept is explored and an OM is created with the relevant media objects. In the second stage, reasoning with the OM is done for concept recognition.

5.1 Constructing the Observation Model

To construct the OM for a concept $c$, we consider the ontology graph $\Gamma$ to be an overlay of two subgraphs: (a) a subgraph $\Gamma_h$ comprising all nodes from which media examples can flow to $c$ and (b) a subgraph $\Gamma_p$ comprising all nodes from which media properties can flow to $c$. $\Gamma_h$ includes $c$, all descendants of $c$ in concept hierarchy and their media examples. $\Gamma_p$ includes $c$, all ancestors of $c$ in concept hierarchy as well as the concepts connected to $c$ through media property propagate relations and all media properties associated with them. The OM is constructed recursively from $\Gamma_p$ and $\Gamma_h$ as shown in figure 9.

5.2 Reasoning for Concept Recognition using the OM

The OM $\Omega_c$ is generally organized as a Bayesian tree with the concept node $c$ at the root node and observable media properties and examples at the leaf nodes, with the exception when a complex media object (a spatio-temporal composition of other media objects) is present in the OM. We shall discuss this exception in the next paragraph. In all other cases, media detectors are run to detect the observable media properties and examples, which are present as the leaf nodes of the OM. This results in belief propagation in the Bayesian network and a belief revision at the root node. The concept $c$ is considered to have been recognized if the posterior property of the root node exceeds a threshold $\theta$ as a result of such belief propagation. We use fuzzy logic based rules to recognize complex media objects that comprise spatio-temporal arrangements of other media objects. Thus, the OM gets partitioned into multiple Bayesian networks as shown in figure 10 when there is a complex media object $mo(c)$ in the OM. The segment of the OM below each of the constituent media objects becomes an independent Bayesian network and contains that media object at the root node. Each of these BNs can be considered as an OM for that media object, which is independently recognized through belief propagation in the respective Bayesian networks. When all constituent media objects of a complex media objects are recognized, the
Anupama Mallik et al

(1) Initialize OM $\Omega$, with concept $c$ as the root (only) node.
(2) Use property propagation subgraph $\Gamma_p$ to add nodes below $c$ in $\Omega_c$ as follows:
   (a) For each media object $mo(c)$, associated with $c$ in $\Gamma_p$.
      i. Add $mo(c)$ below $c$ in $\Omega_c$.
      ii. Compute the CPT from $\Gamma_p$, as may be needed from joint probability distributions.
      iii. Expand $mo(c)$ as follows:
         If $mo(c)$ is an elementary media object, do nothing.
         If other media objects are associated with $mo(c)$ in $\Gamma_p$, add each of them below $mo(c)$. Copy corresponding CPT's from $\Gamma_p$.
         If $mo(c)$ is a spatio-temporal composition of other media objects, copy the spatio-temporal structure in $\Omega_c$.
      iv. Perform steps 2(a) recursively to expand all media objects to their constituent elementary components.
   (b) Media property propagation: For all concepts $c_p(c)$ from which media properties flow to $c$ in $\Gamma_p$.
      i. Add $c_p(c)$ below $c$ in $\Omega_c$.
      ii. Compute the CPT from $\Gamma_p$, as may be needed from joint probability distributions.
      iii. Add each media object $mo(c_p)$ associated with concept $c_p(c)$ in $\Gamma_p$, below $c_p(c)$ in $\Omega_c$ and expand them recursively following step 2(a).
      iv. Perform steps 2(b) recursively to include all concepts from which media properties flow into $c$ and to expand media objects associated with them.
(3) Media example propagation: Use hierarchy subgraph $\Gamma_h$ to add nodes below $c$ in $\Omega_c$ as follows:
   For each concepts $c_a(c)$ which are children of $c$:
   (a) Add all media examples of $c_a(c)$ below $c_a(c)$ in $\Omega_c$.
   (b) Compute the CPT from $\Gamma_h$, as may be needed from joint probability distributions.
   (c) Perform step 3 recursively to include all media examples from all descendant nodes that flow into $c$.
(4) Cleanup:
   (a) Remove all leaf nodes in $\Omega_c$ which are concepts (not observable). Do repetitively, till all such nodes are removed.
   (b) Remove the media objects which are specialized with other media objects present in $\Omega_c$.

compliance to their specified spatio-temporal composition is asserted, resulting in integration of the evidences to recognize the complex media object $mo(c)$. The segment of the OM above $mo(c)$ is also an independent Bayesian network and have that media object as a leaf node. When $mo(c)$ is recognized, this leaf node is instantiated to provide evidential support to the concept at the root node of the OM, just like any other leaf node in the OM. There can be multiple complex media objects in an OM, when the same decomposition and integration rules apply to every node.

6. DOMAIN MODELLING WITH MOWL

6.1 Digital Heritage Preservation

Intangible heritage resources like music and dance forms are preserved by recording the performances of various artistes and exponents in multimedia format. Significance of such a preservation lies in the numerous stories and expressions that have evolved in the prevailing social, cultural and geo-political milieu and the styles of their depiction. An archive of heritage artefacts needs to make use of an ontology that formally captures the complex relations between such entities and their depictions to establish...
the context for the multimedia artefacts. Recent projects on digital heritage preservation [Schneider 2003; Doerr 2012] use traditional ontology representation schemes to create a core ontology and to relate the metadata. While they can reason with the domain concepts, they cannot reason with the multimedia representations of the heritage artefacts and their properties.

We have described an ontology based digital heritage application in the domain of Indian classical dance (ICD) in [Mallik et al. 2011]. In this section, we elaborate on the need for MOWL to model the domain. ICD is based on a well-defined grammar for portraying mythical stories, for enacting the roles therein, and for choreographing dance sequences through a series of dance steps, gestures and postures. The audio-visual depictions (e.g. the dance steps, gestures and postures) in an instance of a dance performance are manifestations of the background concepts (e.g. mythical stories, episodes, roles, emotions, etc.). Concept recognition from the audio-visual recordings of dance performances requires the domain model to capture the relations between the concepts and their media manifestations. The media manifestations in a dance recital can be said to be caused by the background concepts that the artiste intends to portray, and thus provide evidence for the latter. Further, the individual dancers make their own experiments with the compositions and exercise some freedom, introducing a good deal of uncertainty in portrayal of the concepts. The capability of this abductive reasoning with uncertain causal relations of MOWL has a specific advantage over existing ontology languages, when it comes to robust concept recognition. Many of the dance steps are manifested through sequences of postures. The individuality of the artistes brings in uncertainty in such sequences. It is possible to define such uncertain sequences with the spatio-temporal relations in MOWL. Further, many of the concepts in the domain, e.g. the mythical stories and the roles, are interrelated, implying that recognition of one leads to increased belief in presence of another. The concept of media property propagation in MOWL facilitates such reasoning. Thus, the capabilities of MOWL in causal modeling, defining spatio-temporal compositions and reasoning with media property propagation, play an important role in recognizing the underlying concepts from audio-visual depictions of Indian classical dance forms. Such capabilities are not present in traditional ontology languages.

To illustrate the idea, let us consider a classical dance form Odissi, that is typically characterized by an opening act of Mangalacharan (invoking the gods). Mangalacharan is performed as a combination of three dance steps: Bhumipranam (salutation to the Earth), Pushpanjali (offering of flowers) and TrikhandiPranam (salutations). Each of these dance steps manifests in a series of postures. For example, Bhumipranam begins with a Pranam (salutation with folded hands), followed by a bending down with closed eyes. Each of these elementary postures can be detected using a trained set of classifiers. Thus, OdissiDance, Mangalacharan and its constituent steps can be modeled as concepts, evidenced by the observable postures and their sequences, which can be modeled as media objects. The concepts and their relationships are depicted in figure 11a and the corresponding MOWL encoding in figure 11b. The different artistes put different emphasis on these dance steps and bring in some variations in the postures, leading to uncertainties in these media manifestations. We associate CPT’s with the media manifestations based on the significance associated with the postures in the literature. We depict a few other concepts in the figure, for example, OdissiMusic, which generally accompanies OdissiDance; and two of the well-known artistes, Madhumita Raut and Yamini Krishnamurthy, for the dance form, together with their possible media manifestations. These concepts are likely to co-occur with instances of Odissi dance recital, and hence, their corresponding media manifestations can also be expected with an Odissi dance performance. Note that some of the properties, e.g. the artistes or the postures, may not necessarily be exclusive for a particular dance form. In reality, the ontology is much larger and consists of several hundred nodes and relations. Some of these nodes are indicated with dotted lines, which are not further elaborated in this paper.
The ontology can be used to create Observation Models for the different concepts in the domain following the algorithm given in section 5.1. The OM for the concept Mangalacharan (without the CPT's) is shown in figure 12. Note that the concept is recognized with a multitude of evidences and failure of a feature detector has little impact on the overall performance. In particular, the belief in concept Mangalacharan is reinforced with establishment of the context OdissiDance through recognition of related concepts, like the OdissiMusic or either of the two artistes, because of media property propagations.
tion. Thus, the concept is likely to be recognized even if some of the constituent dance steps are not recognized, when the context is established. Further, the constituent elementary postures of complex media objects, like BhumiPranam, have more definitive features, and it is possible to build more accurate classifiers for them. Deployment of such pattern detectors and reasoning with their spatio-temporal composition improves the performance of detection of complex media properties.

<table>
<thead>
<tr>
<th>Entity</th>
<th>SVM Precision</th>
<th>SVM Recall</th>
<th>MOWL Precision</th>
<th>MOWL Recall</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Elementary media objects</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ChawkPosture</td>
<td>0.84</td>
<td>0.87</td>
<td>0.86</td>
<td>0.88</td>
</tr>
<tr>
<td>PranamPosture</td>
<td>0.78</td>
<td>0.78</td>
<td>0.78</td>
<td>0.78</td>
</tr>
<tr>
<td><strong>Complex media objects</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FluteAction</td>
<td>1.00</td>
<td>0.50</td>
<td>0.95</td>
<td>0.89</td>
</tr>
<tr>
<td>BhumiPranam</td>
<td>0.73</td>
<td>0.85</td>
<td>0.86</td>
<td>0.89</td>
</tr>
<tr>
<td>Ardhacheera</td>
<td>0.91</td>
<td>0.83</td>
<td>0.90</td>
<td>1.00</td>
</tr>
<tr>
<td><strong>Abstract Concepts</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>KrishnaRole</td>
<td>0.70</td>
<td>0.60</td>
<td>0.84</td>
<td>0.89</td>
</tr>
<tr>
<td>Mangalacharana</td>
<td>0.47</td>
<td>0.47</td>
<td>0.84</td>
<td>0.84</td>
</tr>
<tr>
<td>OdissiDance</td>
<td>0.44</td>
<td>0.47</td>
<td>0.80</td>
<td>0.87</td>
</tr>
<tr>
<td>BattuDance</td>
<td>0.29</td>
<td>0.36</td>
<td>0.72</td>
<td>0.88</td>
</tr>
<tr>
<td>YashodaRole</td>
<td>0.67</td>
<td>0.55</td>
<td>0.78</td>
<td>0.85</td>
</tr>
<tr>
<td>MahabharatTheme</td>
<td>0.40</td>
<td>0.25</td>
<td>0.55</td>
<td>0.75</td>
</tr>
<tr>
<td>KrishnaTheme</td>
<td>0.70</td>
<td>0.32</td>
<td>0.75</td>
<td>0.83</td>
</tr>
</tbody>
</table>

Fig. 13: Results for concept recognition in ICD ontology

We have compared the concept recognition performance of the new approach with a traditional classifier based approach. We have chosen 63 entities in the ontology for experimentation. The chosen entities represent a mix of elementary media objects, complex media objects and abstract concepts. We have used SVM based classifiers for detecting elementary media objects, such as the elementary dance postures and the genre of the music [Mallik et al. 2011]. We created similar classifiers for the other experimental entities for sake of comparison. The results for some of the entities are shown in figure 13. As expected, the recognition performance in the MOWL-based approach remains the same for low-level media objects which are detected directly with specialized pattern detectors. The complex media objects, such as the dance steps and actions, are described as temporal sequences of elementary postures in our approach. We achieve superior results for these entities. The abstract concepts in the domain have no definitive features and it is difficult to design usable classifiers for them. The recognition performance with SVM based classifiers is extremely poor for such entities. Our approach that involves detecting elementary media objects and combining the evidences, provides significant improvement in results in these cases.

6.2 Product Recommendation for Feature-rich Commodities

Content based filtering technique for product recommendation involves semantic matching of user profile and product features. The semantics association of features with product categories are quite complex in many domains, such as fashion. An approach to recommendation of fashion goods involves encoding of expert opinion on semantics of garment properties as domain ontologies and reasoning with them [Liu et al. 2012], [Vogiatzis et al. 2012]. However, the interpretation of garment features in context of fashion is quite subjective. The crisp ontological classification and the reasoning rules are inflexible to capture such subjectivity. Moreover, a fashion ontology needs to deal with look-and-feel (visual and tactile properties) of garments, which the traditional ontology languages cannot reason with. In this section, we present an ontology based approach to fashion recommendation, where MOWL has been used as an ontology representation language. The probabilistic knowledge representation scheme of MOWL and its capability to reason with media properties make it more suitable for knowledge representation in media-rich domains, such as fashion, than the traditional knowledge based tools.

Figure 14 shows a part of our fashion ontology, which has been encoded in MOWL. The actual ontology has about 200 nodes, with approximately 75% having media related properties. In addition to
defining the taxonomy, the ontology describes uncertain causal relations across concept classes. The entities in our ontology represent garments and their related attributes. Garments can be classified into different categories like those for the \textit{Upper} body, like t-shirts, shirts, jackets, etc, or those worn as \textit{Lower} garments, for e.g. trousers, skirts or shorts. Different clothing attributes, like \textit{Collar} and length of \textit{Sleeves}, that need to be linked to garment classes can be simply attached to them as their media manifestations (dotted lines in figure 14) in MOWL. The ontology also includes entities for human users and their physical attributes. A person belongs to one of the \textit{human color styles} - (Spring, Summer, etc.) [Kentner 1979], which is manifested in the colors of her skin, eyes, hair, etc. The human color styles cause some garment colors to be more suitable than others for that person to wear. Similarly, the body shape of a person may belong to one of the classes like \textit{Normal}, \textit{Oval}, \textit{BroadAtTop} etc, which manifests into some body dimensions and causes preferences to some garment fits.

Another concept in the ontology is \textit{Occasion} to wear or \textit{Style} like \textit{SportsWear}, \textit{OfficeWear} and \textit{SmartCasual}. Physical attributes of a human being and the occasion to wear, cause some garments to be more appropriate than others. Such style advise is coded implicitly as hasRec links (dashed lines in figure 14) from an occasion class to a garment class, according to what is appropriate to wear for a certain occasion. For example, while a shirt or trouser is appropriate to wear to office, a pair of shorts would be an inappropriate choice for OfficeWear. Different garment categories may be appropriate for more than one kind of occasion, for e.g. a shirt can be worn for a \textit{Party} or to \textit{Office}. The \textit{Occasion/Style} classes make some garment patterns to be more suitable to wear than the others. The classification through body attributes and preferences as described above is often subjective and needs to be modeled with uncertainties. To illustrate, while \textit{plain shirts} may be more suitable as an office-wear, \textit{striped shirts} may also be acceptable with some lower probabilities. Similarly, a person with \textit{Normal} body type can generally be recommended to wear \textit{TightFit} clothes, but \textit{LooseFit} clothes may also be appropriate. The domain model proposes that some media properties such as \textit{SummerColor} and \textit{TightFit} flow from human properties \textit{SummerStyle} and \textit{NormalBody} type respectively to garments (suitable for that person). Similarly media properties like a pattern recommended for an occasion flow to the garment (suitable for that occasion).

For our garment image database (composed of garment images downloaded from the internet for illustrative purpose), we extracted low-level image, texture and shape-based features from garment
images. The feature vectors obtained were used to train SVM classifiers to detect generic garment attributes like color and pattern, as well as specific attributes like garment-length. Other attributes like TightFit and LooseFit are subjective and needed some manual tagging. All the garments available in our virtual garment store are semi-automatically annotated using concept-recognition, then curated and associated with the different ontology nodes following a process similar to the one described in the previous application. Fashion experts provide the conditional preferences for the different classes of products in the ontology.

We illustrate the effectiveness of such ontological representation with an example use-case of garments recommended for a user wishing to select clothes for a specific wear-occasion, say for office. As a first step, we construct an OM for the user as shown in figure 15b from the given ontology to profile the user based on the properties human color style and body type. Different human attributes, which are present as the leaf nodes of this OM, are either detected through appropriate sensors or are provided by the user. Instantiation of some of these nodes leads to updated posterior probabilities of the different possible human color-styles and body types, and the ones that best explain the observed human attributes are chosen. In this example, the user profile determined by reasoning with observations of her body colors and dimensions, is of a person with a normal body type and summer color style.

We obtain OMs for concepts OfficeWear, NormalBody and SummerStyle from the ontology, and construct an OM for the recommendation context $X$ by attaching these OMs to a root node denoting this context.
text. An OM for this use-case is shown in figure 15a.\(^4\). We find that leaf nodes in this OM represent clothing attributes. Appropriate detectors are run on a garment image (or available annotations are analyzed) to ascertain the media properties present in the image. For eg, a color histogram based color detector is run on the image, and it returns a similarity value 0.93 for white color, with which the leaf-node WhiteColor in the OM is instantiated. The garment annotation states that it has collar and long sleeves, so the nodes LongSleeve and Collar are instantiated with belief value 1.00. Accordingly, as other attributes are detected in the garment, their corresponding leaf-nodes are instantiated. We show the OM in figure 15a after some leaf-nodes pertaining to a plain, white, tight-fit, collared, long-sleeve shirt have been instantiated and belief has been propagated in the Bayesian network. The posterior probability at the root node for such a Shirt is 0.80, which is above a threshold value, and so this garment is a candidate for recommendation to the User in the current context. We repeat the above process for all the garment images in the database, and get the recommendation score for each garment, which is the posterior probability at the root node of the OM. Based on this score, some garments are recommended for the context represented by the OM. The score is also used to rank the recommended garments so that user can obtain a ‘best to worst’ list of garments recommended according to her profile and occasion to wear. Recommendation scores for some other garments for this context are shown in a table in figure 15c. We find that in this context, plain, tight-fit, long-sleeve shirts are highly recommended, and so are tight-fit T-shirts with collar and long-sleeves, but striped, loose-fit, half-sleeve shirts have a low score and will not be recommended. Plain, tight-fit trousers get high score, and long skirts are also recommended. Figure 17 shows some ranked results (in best to worst order) of garment images recommended by our recommender for the user profile in the example use-case, for different occasions.

<table>
<thead>
<tr>
<th>Recommendation Context</th>
<th>SVM Model NDCG score</th>
<th>OM NDCG score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simple Associations</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NormalBody</td>
<td>0.90</td>
<td>0.92</td>
</tr>
<tr>
<td>OvalBody</td>
<td>0.95</td>
<td>0.96</td>
</tr>
<tr>
<td>Complex Associations</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SummerStyle</td>
<td>0.71</td>
<td>0.86</td>
</tr>
<tr>
<td>SpringStyle</td>
<td>0.80</td>
<td>0.88</td>
</tr>
<tr>
<td>SportsWear</td>
<td>0.80</td>
<td>0.92</td>
</tr>
<tr>
<td>OfficeWear</td>
<td>0.71</td>
<td>0.86</td>
</tr>
<tr>
<td>More Complex Associations</td>
<td></td>
<td></td>
</tr>
<tr>
<td>OfficeWearNormalSummer</td>
<td>0.53</td>
<td>0.87</td>
</tr>
<tr>
<td>SportsWearOvalSpring</td>
<td>0.57</td>
<td>0.85</td>
</tr>
</tbody>
</table>

Fig. 16: Table showing comparison of Recommendation results by SVM models and Observation Models.

Fig. 17: Results of OM based Clothing Recommendation.

To evaluate the recommendation by our recommender system, we conducted experiments with SVM models for clothing suggestions based on labelling of garment images by experts for different contexts, which are as simple as NormalBody or more complex like OfficeWear. The classification score given by the SVM model to a particular garment for a context label, is its recommendation score for the context. These scores were used to rank the recommended garments. Then we obtained clothing recommendations from the ontology by reasoning with the OM for different contexts, and computed the

\(^4\)Nodes are shown with belief value, and CPT for node N with parent X is shown as a pair (\(P(N \mid X)\), \(P(N \mid \neg X)\)) attached to the connecting link

rankings based on the posterior probability in the OM, as explained in the example use-case. The two sets of ranked results were evaluated against expert rankings, using the performance evaluation parameter - Normalized Discounted Cumulative Gain (NDCG) [Busa-Fekete et al. 2012] commonly used to evaluate ranking systems. Table in figure 16 shows a comparison of the NDCG scores for the two models. We find that while SVM based recommendation is comparable to OM based recommendation for contexts which are linearly associated with the attributes, for e.g. association of TightFit clothes to a user having NormalBody; but it starts to perform worse as the associations become more complex, as in the case of combinations of occasions with user profiles. Observation models perform better in these cases, as they incorporate complex style rules, as well as the uncertain associations between high-level concepts and clothing attributes.

Our approach provides quite a few benefits as compared to SVM based [Liu et al. 2012] or SWRL rule-based [Vogiatzis et al. 2012] recommendation. In the first place, the domain rules need not be exhaustively enumerated and it is sufficient to encode the rules connecting the broad classes. MOWL helps in reasoning with the media properties of the concepts like garments, humans and occasions. More importantly, the causal probabilistic reasoning enables ranking of the recommendations allowing for user preferences, if required, which can be encoded in the CPT's. The reasoning scheme is robust. It can recommend garments when partial inputs are available. This is not possible with crisp ontology based reasoning. Further, inputs from expert rules and purchasing patterns of online commercial websites can be seamlessly integrated in the same reasoning framework.

7. CONCLUSION

Contextual and semantic interpretation of multimedia data poses a major challenge, unless they are collected and organized in some specific way. Extension of semantic web principles to multimedia data needs to overcome what is known as the semantic gap in the literature. In this paper, we have argued that multimedia applications need an ontology representation and reasoning scheme different from the traditional ontology languages. The difference primarily lies in the uncertain causal relations that exist between the domain concepts and their media manifestations, which calls for an evidential reasoning. Further, domain knowledge is required to determine the influence of related concepts in media manifestations of a concept. We have presented a new ontology representation language, MOWL, that can support such perceptual modeling and reasoning. Uncertainties in multimedia manifestations are dealt with probabilistic reasoning using Bayesian networks. While use of Bayesian network for multimedia concept recognition is not uncommon, the novelty in our approach lies in dynamic creation of the BN in a domain context, incorporating additional contextual information. We have experimented with two disparate knowledge intensive and media feature rich applications to establish the generality of the approach. We plan to experiment with more real life applications to investigate the sufficiency of the language features of MOWL and extend it, if necessary.

REFERENCES


APPENDIX

A. MOWL SYNTAX STRUCTURAL SPECIFICATION

A structural specification of MOWL provides the foundation for the implementation of MOWL tools such as parsers and reasoners. It describes the conceptual structure of MOWL ontologies and thus provides an abstract representation for all syntaxes of MOWL. This structural syntax specification is an extension of general definitions in OWL 2 syntax specification [Motik et al. 2012]. An ontology encoded in MOWL consists of the three different syntactic categories:

- **Entities** - Classes that represent sets of individuals including concepts and media objects;
- **Datatypes** - which are sets of literals such as strings or integers or more complex user-defined XSD and WSDL datatypes;
- **Object and data properties** - which can be used to represent relationships in the domain. These properties include some new kind of relations like media associations, media propagation, spatio-temporal relations and uncertainty specifications. MOWL specification here is written using the BNF notation, summarized in Table 1 in OWL-2 syntax [Motik et al. 2012]. In this document, examples assume the following namespace prefix bindings unless otherwise stated:

<table>
<thead>
<tr>
<th>Prefix</th>
<th>IRI</th>
</tr>
</thead>
<tbody>
<tr>
<td>rdf</td>
<td><a href="http://www.w3.org/1999/02/22-rdf-syntax-ns">http://www.w3.org/1999/02/22-rdf-syntax-ns</a></td>
</tr>
<tr>
<td>rdfs</td>
<td><a href="http://www.w3.org/2000/01/rdf-schema">http://www.w3.org/2000/01/rdf-schema</a></td>
</tr>
<tr>
<td>xsd</td>
<td><a href="http://www.w3.org/2001/XMLSchema">http://www.w3.org/2001/XMLSchema</a></td>
</tr>
<tr>
<td>owl</td>
<td><a href="http://www.w3.org/2002/07/owl">http://www.w3.org/2002/07/owl</a></td>
</tr>
</tbody>
</table>

OntologyDocument := { prefixDeclaration } MowlOntology
prefixDeclaration := 'Prefix' '(' prefixName '=' fullIRI ')' MowlOntology := "MowlOntology" '(' MowlOntologyIRI [ versionIRI ] ) directlyImports-...Documents OntologyAnnotations axioms ')' axioms := Axiom ...

// MOWL datatypes derived from standard XSD datatypes //
mowl:probValType
mowl:axisPtType
mowl:tupleType
mowl:probabilityVector

// Extending OWL2 Classes with MOWL Classes //
// Basic Classes of MOWL //
Concept := IRI
ComplexObject := IRI

// Media based classes of MOWL //
MediaExample := IRI
MediaDetector := IRI

// Spatio-temporal classes of MOWL //
STPredicate := IRI

// Media classes of MOWL //
MediaExample := IRI
MediaDetector := IRI

// Uncertainty specification classes of MOWL //
ExampleDetector := IRI
CPTable := IRI
CPRow := IRI

// Extending OWL2 Expressions with MOWL Class Expressions //
ClassExpression := ... | BasicClassExpression | MediaClassExpression | STClassExpression | CPTClassExpression
BasicClassExpression := 'Concept' | 'MediaObject'
MediaClassExpression := 'MediaExample' | 'MediaDetector'
STClassExpression := 'ComplexObject' | 'ST_Predicate' | 'ST.Predicate' | 'FuzzyRelation' | 'FuzzyContainment'
CPTClassExpression := 'CPTable' | 'CPRow'

// Extending Object Properties of OWL2 with MOWL Object Properties //
ObjectPropertyExpression := ObjectProperty | MediaPropertyExpression | STPropertyExpression | RelationPropertyExpression | Containment-PropertyExpression | CPTPropertyExpression

// MOWL Object Properties //
MediaPropertyExpression := 'hasMediaExample' | 'hasMediaDetector' | 'hasMediaObject' | 'propagateMedia'
STPropertyExpression := 'hasST_Predicate' | 'ComponentPropertyExpression'

ComponentPropertyExpression \(:=\) ‘Subject’ | ‘Object’
RelationPropertyExpression \(:=\) ‘RX’ | ‘RY’ | ‘RZ’ | ‘RT’
ContainmentPropertyExpression \(:=\) ‘RC’
CPTPropertyExpression \(:=\) ‘hasCPT’ | ‘hasCPRow’ | ‘conditionedOn’

//Extending Data Properties of OWL2 with MOWL Data Properties //
DataPropertyExpression \(:=\) DataProperty | MediaDataExpression | RelationDataExpression | ContainDataExpression | CPTDataExpression

//MOWL Data Properties //
MediaDataExpression \(:=\) ‘hasURI’ | ‘usesExample’
RelationDataExpression \(:=\) ‘hasT’ | ‘hasU’ | ‘hasV’ | ‘hasW’ | ‘hasX’
ContainDataExpression \(:=\) ‘hasP’ | ‘hasQ’ | ‘hasR’ | ‘hasS’
CPTDataExpression \(:=\) ‘ParentStates’ | ‘probValues’

//Extending OWL2 axioms by MOWL axioms //
axiom \(:=\) ... | ClassAxiom | MowlMediaAxiom | MowlSTAxiom | MowlCPTAxiom

//MOWL hierarchical relation is same as OWL //
ClassAxiom \(:=\) SubClassOf | ...
SubClassOf \(:=\) ‘SubClassOf’ '(' axiomAnnotations subClassExpression superClassExpression ')' | ...
subClassExpression \(:=\) ClassExpression
superClassExpression \(:=\) ClassExpression

//MOWL Media based axioms //
MowlMediaAxiom \(:=\) MediaObjectAxiom | MediaPropertyAxiom | PropagateAxiom

//Property to associate media manifestations to MOWL concepts //
MediaObjectAxiom \(:=\) ‘hasMediaObject’ '(' axiomAnnotations BasicClassExpression MediaObject ')' | ...
MediaPropertyAxiom \(:=\) MediaDetectorAxiom | MediaExampleAxiom

//Properties that allow association of media properties to MOWL media objects //
MediaDetectorAxiom \(:=\) ‘hasMediaDetector’ '(' axiomAnnotations MediaObject MediaDetector ')' | ...
MediaExampleAxiom \(:=\) ‘hasMediaExample’ '(' axiomAnnotations MediaObject MediaExample ')' | ...

//Property that allows propagation of media properties between MOWL concepts //
PropagateAxiom \(:=\) ‘propagateMedia’ '(' axiomAnnotations Concept MediaObject ')' | ...

//Data Properties that link WSDL specification to Pattern Detectors, and media examples to Example Detectors //
MediaDataAxiom \(:=\) HasURIAxiom | UseExampleAxiom

UseURIAxiom \(:=\) ‘hasURI’ '(' axiomAnnotations MediaClassExpression xsd:anyURI ')' | ...
UseExampleAxiom \(:=\) ‘usesExample’ '(' axiomAnnotations ExampleDetector MediaExample ')' | ...

//Spatio-temporal properties allow special encoding for spatial, temporal and spatio-temporal associations //
MowlSTAxiom \(:=\) STAxiom | RelationAxiom | ContainmentAxiom | FuzzyAxiom

STAxiom \(:=\) PredicateAxiom | ComponentAxiom

PredicateAxiom \(:=\) ‘hasST Predicate’ '(' axiomAnnotations ComplexObject ST_Predicate ')' | ...
ComponentAxiom \(:=\) ‘ComponentPropertyExpression’ '(' axiomAnnotations ComplexObject MediaObject ')' | ...

//Fuzzy aspects of Spatio-temporal properties //
RelationAxiom \(:=\) ‘RelationPropertyExpression’ '(' axiomAnnotations ST_Predicate FuzzyRelation ')' | ...
ContainAxiom \(:=\) ‘ContainPropertyExpression’ '(' axiomAnnotations ST_Predicate FuzzyContainment ')' | ...
FuzzyAxiom \(:=\) FuzzyRelationAxiom | FuzzyContainAxiom

FuzzyRelationAxiom \(:=\) ‘RelationDataExpression’ '(' axiomAnnotations FuzzyRelation FuzzyVector ')' | ...
FuzzyContainAxiom \(:=\) ‘ContainDataExpression’ '(' axiomAnnotations FuzzyContain FuzzyVector ')' | ...

//Uncertainty Specification related Axioms //
CPTAxiom \(:=\) CPTTableAxiom | CPTRowAxiom | ConditionAxiom | CPTDataAxiom
CPTTableAxiom \(:=\) ‘hasCPT’ '(' axiomAnnotations BasicClassExpression CPTable ')' | ...
CPTRowAxiom \(:=\) ‘hasRow’ '(' axiomAnnotations CPTable CPTRow ')' | ...
ConditionAxiom := 'conditionedOn' '(' axiomAnnotations CPTable BasicClassExpression ')' 
CPTDataAxiom := ParentAxiom | ValueAxiom 
ParentAxiom := 'ParentStates' '(' axiomAnnotations CPRow xsd:string ')' 
ValueAxiom := 'probValues' '(' axiomAnnotations CPRow mowl:probabilityVector ')' 

// Generic OWL2 Assertions apply to MOWL per se //
MOWL - An Ontology Representation Language

• 1:25

<owl:ObjectProperty rdf:ID="hasMediaExample">
  <rdfs:domain rdf:resource="#MediaObject" />
  <rdfs:range rdf:resource="#MediaExample" />
</owl:ObjectProperty>

<owl:ObjectProperty rdf:ID="hasURI">
  <rdfs:comment> URI associated with Media examples and detectors for their location. </rdfs:comment>
  <rdfs:domain rdf:resource="#MediaObject" />
  <rdfs:range rdf:resource="&xsd;anyURI" />
</owl:ObjectProperty>

<owl:ObjectProperty rdf:ID="usesExample">
  <rdfs:comment> Property of example detector to specify the class of media examples it can use. </rdfs:comment>
  <rdfs:domain rdf:resource="#ExampleDetector" />
  <rdfs:range rdf:resource="#MediaExample" />
</owl:ObjectProperty>

<!-- Spatio-temporal constructs -->
<owl:Class rdf:ID="ComplexObject">
  <rdfs:comment> MOWL class to represent spatio-temporal composition of media objects. </rdfs:comment>
  <rdfs:subClassOf rdf:resource="#MediaObject" />
  ...
</owl:Class>

<owl:Class rdf:ID="ST_Predicate">
  <rdfs:comment> Spatio-temporal predicate to encode spatial and temporal relations between media objects. </rdfs:comment>
  <rdfs:subClassOf rdf:resource="#FuzzyRelation" />
  ...
</owl:Class>

<owl:Class rdf:ID="FuzzyRelation">
  ...
</owl:Class>

<owl:objectProperty rdf:ID="RX">
  <rdfs:domain rdf:resource="#ST_Predicate" />
  <rdfs:range rdf:resource="#FuzzyRelation" />
</owl:objectProperty>

<!-- Datatypes needed for MOWL constructs -->
<owl:DatatypeProperty rdf:ID="probValType">
  <rdfs:isDefinedBy rdf:resource="&mowlDatatypes;" />
  <rdfs:comment> type to encode decimal values from zero to one </rdfs:comment>
</owl:Datatype>

<owl:DatatypeProperty rdf:ID="axisPtType">
  <rdfs:isDefinedBy rdf:resource="&mowlDatatypes;" />
  <rdfs:comment> string or decimal type to accommodate "-infinity", "+infinity" as well as other decimal points on the axis. </rdfs:comment>
</owl:Datatype>

<owl:DatatypeProperty rdf:ID="tupleType"/>
  <rdfs:comment> tuple (t,v) type i.e. a pair indicating a point t on the axis with the fuzzy membership value v </rdfs:comment>
</owl:Datatype>
