

# ONTOLOGY FOR SEMANTIC MULTIMEDIA WEB

*Hiranmay Ghosh*

TCS Innovation Labs Delhi  
TCS Towers, 249 D&E Udyog Vihar Ph-IV  
Gurgaon 122015, INDIA

*Santanu Chaudhury*

Electrical Engineering Department  
Indian Institute of Technology, Delhi  
New Delhi 110016, INDIA

## ABSTRACT

This paper proposes a new multimedia ontology based scheme for semantic multimedia data processing on the web. The ontology language "Multimedia Web Ontology Language" (MOWL), is designed as an extension of OWL, the W3C recommended ontology language for the web. MOWL supports creation of and reasoning with perceptual modeling of concepts, and probabilistic evidential reasoning.

**Index Terms**— Multimedia systems, Ontology, Semantic Web, Bayesian network, Evidential reasoning

## 1. INTRODUCTION

The vision of semantic web proposes an environment where the data and services on the web can be semantically interpreted and processed by machines to facilitate human consumption. In today's cyberspace, audio-visual artifacts compete with traditional text and data in their information content. Machine interpretation of multimedia data is therefore essential for realization of the semantic web vision. While textual documents are effectively processed by semantic web technology, application of the technology for multimedia data is still at its infancy. In this context, we propose a new ontology based approach for contextual semantic interpretation of multimedia data.

Semantic web technology relies on ontology as a tool for modeling an abstract view of the real world and contextual semantic analysis of documents. Ontology languages like Web Ontology Language (OWL)[1] uses linguistic constructs for modeling the real-world and can be conveniently used for interpreting textual documents. An attempt to use ontology for interpreting multimedia contents is hindered by the *semantic gap* that exists between media features appearing in the documents and the linguistic structures representing the concepts in the ontology. To cope up with this deficiency, there have been some attempts to extend ontology with addition of media examples for multimedia data processing [2]. However, such extensions do not meet the specific demands for reasoning with media data. Semantic retrieval in multimedia repository is generally carried out by intelligent concept recognition tasks that are specific to the media types and repository

architectures. Multimedia libraries, pertaining to a common theme, can be built by several independent organizations in different ways and call for different retrieval strategies. Integration of such heterogeneous collection under a common thematic umbrella is a challenge with current semantic web technology.

In this context, we propose to extend Multimedia Web Ontology Language (MOWL) [3] as an extension to OWL, with additional capability of creating media-based perceptual models of real-life concepts and events and to reason with them. In contrast to crisp Description Logics (DL) based reasoning in OWL, we propose probabilistic evidential reasoning to cope up with the uncertainties that are inherent to multimedia data processing. Further, it is possible to create unique repository specific search strategies in response to a semantic query by reasoning with the perceptual model and the repository capabilities.

The rest of the paper is organized as follows. Section 2 provides an overview of the perceptual reasoning model and justifies proposal of MOWL. The new language constructs for MOWL are introduced in section 3. Section 4 explains the evidential reasoning scheme with MOWL. Finally, section 5 concludes the paper.

## 2. PERCEPTUAL REASONING AND MOWL

Ontology is a formal description of the abstraction of a domain. Human beings use natural languages to communicate an abstract view of the world. Natural language constructs are symbolic representations of human experience and is close to the conceptual model that Semantic Web technologies deal with. Thus, it seems quite *natural* to use natural language constructs to represent the ontology elements. As a result, it becomes convenient to apply semantic web technologies in the domain of textual information. In contrast, media artifacts are perceptual recording of human experience. An attempt to use the conceptual model to interpret these perceptual records gets severely impaired by the semantic gap that exists between the perceptual media features and the conceptual world. However, the concepts have their roots in perceptual experience of human beings and the apparent disconnect

between the conceptual and the perceptual worlds is rather artificial. The key to semantic processing of media data lays in harmonizing the seemingly isolated conceptual and the perceptual worlds.

Concepts are formed in human minds through a complex refinement process of personal experiences. Observations of the real world objects amounts to reception of large volumes of perceptual data through our sensory organs. The raw data go through a process of refinement to result in mental models. The models are further abstracted over a large number of observations, to give rise to *concepts*, which are labeled with linguistic constructs to facilitate communication. The fact that concepts are abstractions of perceptual observations has an interesting consequence. A concept gives rise to the expectation of some perceptible media properties on its embodiment in a multimedia artifact. Observation of those media properties forms the basis of concept recognition in a multimedia document. Figure 1 depicts the formation of the concept *Medieval Indian Monument* and the abstracted visual patterns that are expected on an embodiment of the concept in a multimedia artifact. Note that the different instances of the concept have significant variations and the perceptual model comprises an abstraction of their common visual properties.

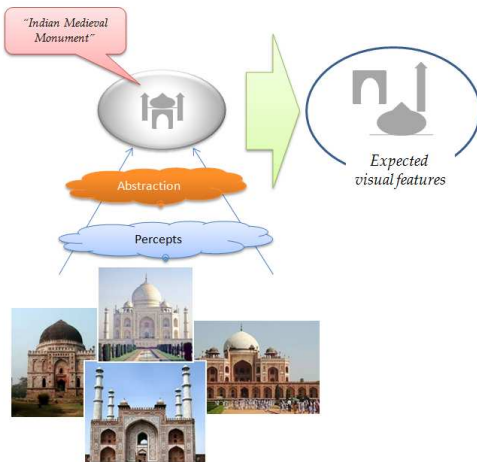


Fig. 1. Perceptual Model

Multimedia Web Ontology Language (MOWL) has been proposed to encode the domain knowledge pertaining to the expected perceptual properties of concepts and to reason with them, over and above the semantic properties that are encoded and reasoned with in OWL. Figure 2 depicts a small section of an ontology encoding media based description of concepts. The individual media patterns are often connected by some spatial or temporal relationships with each other, in context of a concept or an event. For example the dome of a medieval monument should occur above its other components, such as the minarets or the facade. moreover, the expected perceptual properties of concepts can be *inherited* by other concepts in

the domain, depending on their relationship. For example, a specific instance of a monument, the **Taj Mahal**, is made of a specific class of stones, **marble**, and therefore expected to "inherit" its color and texture properties. This form of media property inheritance is quite distinct from classical *property inheritance* rules that is supported by classical ontology models and requires distinct form of reasoning. This is required for creating a complete Observation Model of a concept using media properties of that concept and other related concepts.

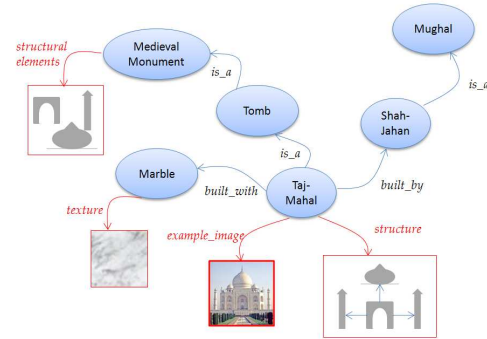


Fig. 2. Multimedia Ontology

Once an *Observation Model* for a concept is created, the presence of the expected media patterns and their spatio-temporal relationships can be verified in a multimedia artifact to detect the presence of the concept. However, there can be significant intrinsic differences between the instances of the concepts as well as the instances of the media artifacts depicting a specific instance of the concept. The latter can be attributed to variation in viewpoints, change in perspectives, occlusions and lighting conditions. This leads to a requirement of uncertain reasoning for semantic multimedia data processing. We have proposed a Bayesian Network based evidential reasoning scheme with MOWL.

### 3. MOWL LANGUAGE CONSTRUCTS

MOWL is designed as an extension of OWL to ensure compatibility with the W3C standards. It uses OWL constructs to define classes, individuals and properties. In addition, it proposes some language extensions for encoding media properties, property propagation rules and specification of conditional probabilities that characterizes uncertainty in association between concepts and their media properties.

The abstract class *mowl:MediaObject* in MOWL signifies media properties. In general, the expected media properties of concepts can either be specified as property constraints, e.g. *color = WhiteColor*, or be cited as media examples. Accordingly, MOWL defines two subclasses of *mowl:MediaObject* (see figure 3(a)):

1. Media property class *mowl:MediaFeature* can define a

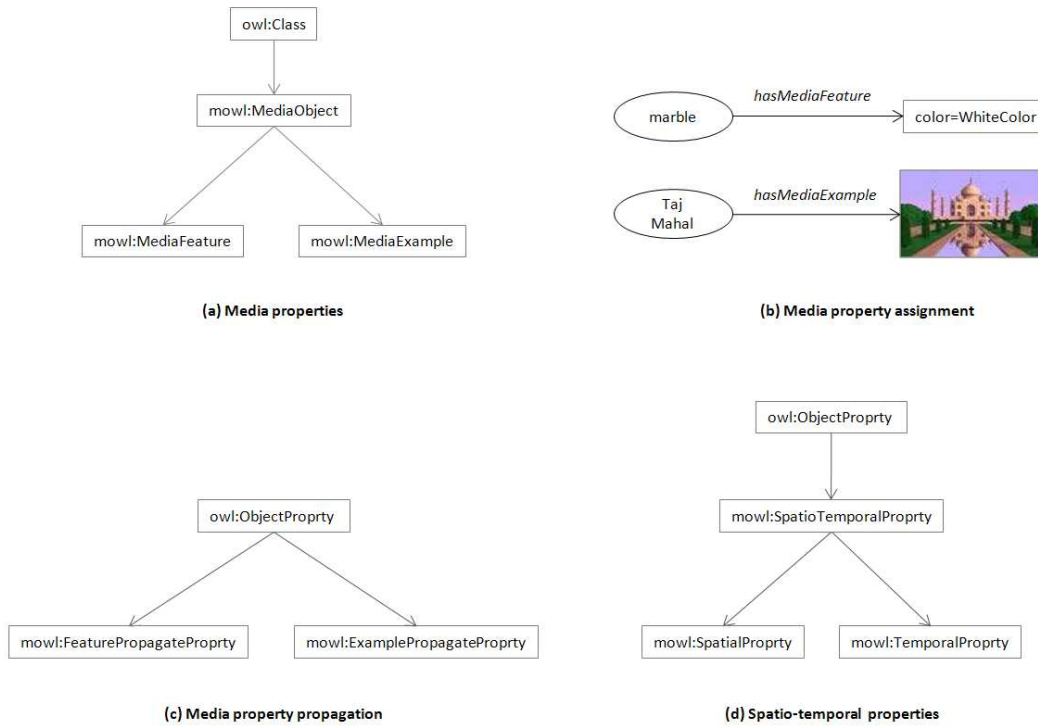


Fig. 3. MOWL Constructs

constraint on a media feature. For example color is *WhiteColor*.

- Media example class *mowl:MediaExample* can define a media example. For example, photograph of a monument.

We propose that the media feature specifications follow MPEG-7 [4] schema.

The OWL class *owl:ObjectProperty* can be instantiated to give individual properties with a defined range and a domain. Media features are associated with concepts. Hence the domain and the range of an instantiated property are set as a *mowl:MediaFeature* and the concept respectively. For example, we associate the media feature *color=WhiteColor* with a concept **marble** using an instantiated property *hasMediaFeature*. Similarly, we can associate an example image with the concept **TajMahal** using an instantiated property *hasMediaExample* (see figure 3(b)).

Media properties and examples can propagate across connected concepts in an ontology depending on the semantics of the relation in a specific domain. MOWL defines the properties *mowl:FeaturePropagationProperty* and *mowl:ExamplePropagationProperty* to indicate the two types of propagation. A media feature or an example will propagate from one concept to another if and only if the relation connecting the concepts has the corresponding property (See figure 3(c)). For example, if the relation *builtWith* has

*mowl:FeaturePropagationProperty* and connects the concepts **TajMahal** and **marble**, the properties of marble will flow into TajMahal.

The different media properties may be connected with different spatio-temporal relations. MOWL defines constructs for specifying such spatio-temporal properties (see figure 3(d)). Rather than restricting the relations to a few pre-specified values, such as *left*, *right*, *before*, *after*, etc., MOWL provides for defining such relations. One way to define such relations has been proposed in [5].

The probabilistic reasoning model of MOWL is based on Bayesian Network. Accordingly, MOWL treats the nodes (concepts) in the ontology as random variables and defines constructs for specifying Conditional Probability Tables (CPT) for the connected nodes. The approach is similar to [6] that proposes a probabilistic extension of OWL.

#### 4. REASONING WITH MOWL

In our approach, concept recognition in multimedia artifacts is based on observation of some expected media patterns in the artifact. We model concept recognition as an evidential reasoning problem. If a concept *c* causes a set of media patterns *M* to appear in a multimedia artifact, observation of a specific pattern *m*  $\in$  *M* provides some evidence towards the concept. A concept is recognized when there is sufficient cu-

mulative evidence for the concept as a result of observation of several media features,  $m_1, m_2, \dots, m_k \in M$ . There are three distinct stages of reasoning with MOWL for semantic query processing.

The first stage involves derivation of an Observation Model for a semantic query. A semantic query is mapped to one or more MOWL concept nodes based on some similarity measure between the query and the attributes of the node attributes. An Observation Model for each of these nodes is created from the media properties associated with that node and other nodes which are connected to it with a relation bearing *propagate* property. The CPT's for the connected nodes are derived from the corresponding nodes in the ontology. Finally, the Observation Models for all the nodes are merged. The resultant Observation Model is thus organized as a Bayesian Tree. The root node in the tree represents the concept and the leaf nodes represent its expected media properties. Figure 4 depicts a typical Observation Model for the concept **TajMahal**.

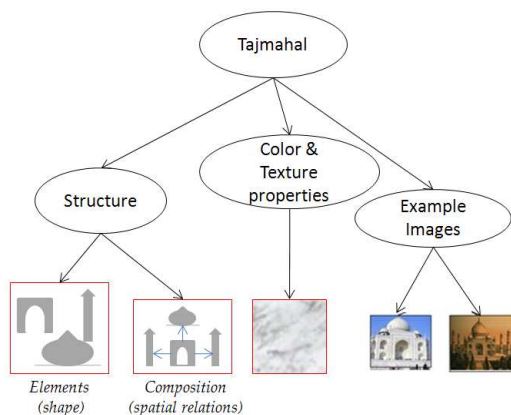


Fig. 4. Observation Model

The Observation Model is neutral to any specific media type or repository architecture and, in general, contains a redundant set of patterns for different types. It may not be possible or may be extremely computationally expensive to evaluate all of these media patterns at a given multimedia repository. For example, an audio pattern cannot be detected in a library of still images. However, the evidential reasoning scheme deployed with MOWL can produce robust results despite limited input data. This property is exploited to realize concept recognition with a subset of media patterns that is specified in an Observation Model. Identification of an optimal subset of the media patterns to formulate a search strategy for a multimedia artifact collection is described in [7].

A retrieval strategy which is a subgraph of an Observation Model, is also organized as a Bayesian Tree with the concept at the root and the expected media patterns in the leaf nodes. An *observation* with a leaf node in a retrieval strategy involves actuation of some feature detection mechanism to

discover the corresponding media feature in a multimedia artifact. In general, an observation results in a similarity score, which is normalized in the range  $[0, 1]$ . We interpret this similarity score as a virtual evidence for the leaf node. The assignment of the virtual evidences to the different leaf nodes result in belief propagation in the Bayesian Network. The posterior probability of the root node as a result of such belief propagation represents the degree of belief in the concept.

## 5. CONCLUSION

In this paper, we have proposed a new ontology representation suitable for processing multimedia data repositories distributed on the web. The ontology representation MOWL is a syntactic extension of OWL, the current standard of ontology language for the web proposed by W3C. It supports perceptual modeling of concepts and a probabilistic evidential reasoning scheme that is necessary for multimedia data processing.

## 6. REFERENCES

- [1] W3C Recommendation, "OWL Web Ontology Language: Overview," February 2004.
- [2] Marco Bertini, Alberto Del Bimbo, Carlo Torniai, Costantino Grana, and Rita Cucchiara, "Dynamic pictorial ontologies for video digital libraries annotation," in *MS '07: Workshop on multimedia information retrieval on The many faces of multimedia semantics*, New York, NY, USA, 2007, pp. 47–56, ACM.
- [3] Hiranmay Ghosh, Santanu Chaudhury, Karthik Kashyap, and Bridaduti Maiti, *Ontology Specification and Integration for Multimedia Applications. In Ontologies: A Handbook of Principles, Concepts and Applications in Information Systems*, pp. 265–296, Springer, 2007.
- [4] B S Manjunath, Phillipe Salembier, and Thomas Sikora, *Introduction to MPEG-7: Multimedia Content Description Interface*, John Wiley & Sons, 2002.
- [5] Sujal Subhash Wattamwar and Hiranmay Ghosh, "Spatio-temporal query for multimedia databases," in *MS '08: Proceeding of the 2nd ACM workshop on Multimedia semantics*, New York, NY, USA, 2008, pp. 48–55, ACM.
- [6] Zhongli Ding and Yun Peng, "A probabilistic extension to ontology language OWL," in *Proceedings of the 37th Hawaii International Conference On System Sciences*, January 2004.
- [7] Hiranmay Ghosh and Santanu Chaudhury, "Distributed and reactive query planning in r-magic: An agent based multimedia retrieval system," *IEEE Trans KDE*, vol. 16, September 2004.