

---

# An Ontology Representation Language for Multimedia Event Applications

---

Nisha Pahal\*, Brejesh Lall and Santanu Chaudhury

*Department of Electrical Engineering, Indian Institute of Technology, Delhi, India*  
*E-mail: nisha23june@gmail.com; brejesh.lall@gmail.com;*  
*schaudhury@gmail.com*

*\*Corresponding Author*

Received 28 February 2020; Accepted 12 November 2020;  
Publication 05 March 2021

## Abstract

This paper presents formalization of a new Multimedia Web Ontology Language (E-MOWL) to handle events with media depictions. The temporal, spatial and entity aspects that are implicitly linked to an event are represented through this language to model the context of events. The language E-MOWL provides a rich method for representing knowledge corresponding to a specific domain wherein the context specifies the intended meaning of each element of the domain of discourse; an element in different context may correspond to different functional role. The context information associated with an event ties the audiovisual data with event related aspects. In this work, we have extended E-MOWL to model the geographic properties associated with an event by exploiting the geospatial knowledge. This facilitates in identifying the geographic context of an event. All these aspects when considered altogether provide the evidence and contribute towards recognizing an event from multimedia documents. The language also enables reasoning with the uncertainty associated with the events and is organized in the form of Bayesian Network (BN). The media items that are semantically relevant can be assimilated together on the basis of their association with events. We have demonstrated the efficacy of our approach by utilizing an ontology

*Journal of Web Engineering, Vol. 20.2, 217–252.*

doi: 10.13052/jwe1540-9589.2021

© 2021 River Publishers

for the entertainment category in news domain to offer an application *news aggregation* and event-based book recommendations.

**Keywords:** Multimedia web ontology language (MOWL), event, multimedia event ontology language (E-MOWL), Bayesian network (BN), inference.

## 1 Introduction

Events are things that happen. In contrast, physical objects are said to exist. Events are occurents – they occupy time periods and may persist by having different stages at different times. Typically events are located in space but may not have clear spatial boundaries. Also events can be co-located in space but unlikely to move over space. Semantic analytics require, among other things, explicit modeling of spatial and temporal relationship among entities. This can be achieved by considering event or occurent as the basic entity and modeling event-to-event and event-to-object relationships. Event oriented ontology can be used for describing events and related concepts. Traditional ontology representation schemes have been built around Web Ontology Language (OWL) with description logic as the underlying formal logical model that can express classification and properties but is not suited for temporal and spatial relations.

Typically events are recorded on media. A news report or tweet can be a textual representation of an event. A video sequence can typically capture a possible dynamic event over a temporal interval. Event record indicates where event has occurred and which objects/entities are participant in the event. Photos with associated time, GPS location and camera parameters are nothing but records of events in our lives. Hence, understanding and mining of multimedia contents requires a conceptual model of events. The already existing Multimedia Web Ontology Language (MOWL) can be leveraged for perceptual modelling of a domain, where the concepts manifest into media patterns in the multimedia document and helps in semantic processing of the contents. Providing semantic tag and reasoning with events recorded in multimedia require such specialized multimedia ontology which consider events as fundamental entities. The multimedia depiction of an event amounts to assimilation or composition of images of atomic level sub-events that are parts of higher-level composite event. Since context is the key to event detection, the event models should have the capability to correctly record the event aspects that is, the time, date and the agent/ entity related information. Another aspect which is considered especially important for

event characterization is the geospatial knowledge. It ties the events to the geographic entities thus increasing the expressive power of a multimedia web ontology language to tackle events. This further helps to reason about the geographic context present in multimedia document.

### **1.1 Related Works**

This section briefly describes the related work corresponding to the ontology, event models and multimedia. The classical ontology models can not capture dynamic features of the event. Liu et al. (Liu et al., 2010) extended the existing OWL by introducing some new constructs and axioms for representing multimedia event ontology. Westermann and Jain (Westermann and Jain, 2006) have discussed elementary design considerations for a suitable event model-E for multimedia applications. For inference they have developed an event query algebra and rule language based on formal definition of E. A framework for video event representation and annotation based on definition of an ontology, suitable for video content is described by Francois et al. (Francois et al., 2005). They proposed a scheme for representing video events called Video Event Representation Language (VERL). But this model fails to capture uncertainty associated with an event. Westermann and Jain (Westermann and Jain, 2007) presented a set of requirements for a common event model. They also considered uncertainty associated with an event and elementary taxonomic dimensions for the comparison of event models and their capabilities. Scherp and Mezaris (Scherp and Mezaris, 2014) had presented a very illustrative explanation of events, their relations and mining the aspect of an event. They have conducted a detailed analysis on an extensive set of existing event-based systems and event models with respect to the different event aspects. Shaw and Larson (Shaw and Larson, 2008) examined a number of standards for organizing collections of archival, historical, news and personal information to see what resources they offer for modelling events. They modelled the events in the lives of individuals and connect them with particular places and time periods. An ontological structure with explicit and formal descriptions about the concepts have been proposed by Wu (Wu, n.d.) for monitoring changes in tropic forest. The author also considered spatial, temporal and thematic information associated with an event and used them to construct ontology which is further used to infer implicit events from the deforestation data. Pongpaichet et al. (Pongpaichet et al., 2013) introduced a computing framework called EventShop to recognize evolving situations from massive web streams in real-time. They had considered these

web streams as spatio-temporal thematic streams and combined them using a set of generic spatio-temporal analysis operators to recognize evolving situations.

The significant work has been carried out to model, represent and extract event related data from multimedia resources. Many existing event ontology languages make use of OWL to depict an event but doesn't consider media manifestations while specifying an event from media data. The authors in [(Chaudhury et al., 2015), (Ghosh and Chaudhury, 2013)] established the need for constructing a multimedia web ontology language and introduced the semantics of its novel language constructs. Also, the Bayesian Network based probabilistic reasoning scheme to cope up with the inherent uncertainties was introduced. They validated their approach with two disparate knowledge-intensive applications involving reasoning with media properties of concepts. Their approach helped in bridging the gap between perceptual and conceptual world. Although, MOWL can be utilized for perceptual modeling of a domain, where the concepts manifest into media patterns in the media document and helps in semantic processing of the contents. But, the key to semantic processing of events in multimedia data lies in its ability to reason with the media properties of concepts in a domain. The requirement for relating the digital objects with context is essential in order to visualize the correct perspective of the multimedia document. So, apart from the ones existing in MOWL the authors in (Pahal et al., 2013) had proposed additional constructs to take care of events with media depictions. In this, the MOWL definition of spatio-temporal relations (Wattamwar and Ghosh, 2008) has been extended to represent the context of an event. The context is any kind of information that gets defined with respect to an event instance wherein context has its own structure and involves parameters like time, location, and actors/ entities involved in the event.

## **1.2 Our Contribution**

The present work extends the previous work (Pahal et al., 2013) to specify constructs that model the geospatial knowledge associated with an event. The geospatial knowledge has been exploited to recognize the relatedness between the geographic entities that exist in various documents. For this, it is required that the ontology encodes the geographical terminology and explicit spatial relationship between these geographic entities. The behavior of coordinate and topological relations are predictable in Geographic Information System (GIS), but relative locations such as near, or touches are much more difficult to model. Therefore, we require geospatial knowledge

that maps the geographical entities onto the multimedia event ontology concepts incorporating various geo-relations. Many scenarios, e.g. mountainous terrain, implicitly contain geo-tags which are ignored by standard web ontology languages. These geo-tags contain information that can be exploited to perform more complete semantic processing. The present work facilitates identification of locations and the associated relationship which can be referred to using the place names. This helps in identifying the geographical context of an event that may refer to some location exactly and in some cases approximately. For instance, if Mumbai is not named in a news related to Maharashtra, then to capture geographic relations like Mumbai *is in* Maharashtra, geospatial knowledge can be utilized. To deal with the uncertainty associated with the event properties of the low-level multimedia concept we have incorporated probabilistic inference scheme. To completely describe the plethora of events that are taking place in a video we have leveraged a multimedia language that facilitates comprehensive support to represent various events and the relations that exists between them. The extended multimedia event ontology language not only considers the relation between two events or between an event and event object but also takes into account the geographical context of the event and the associated uncertainty. Using the geographic relations as present in the multimedia event ontology we extract the spatial patterns corresponding to an event present in documents. This aids in identification of event related information and their geographical context to efficiently retrieve events present in multimedia documents. Basically, such extension is useful in cases where we have a database of videos and depiction of event-centric information is of concern. For instance, media documents captured at some event by different users share the same context. The additional constructs and uncertainty specification are presented in Section 4.

The rest of the paper is structured as follows. In the next section the need for event representation and modelling in multimedia domain has been discussed. In Section 3, the formal model for representation of an event is defined. The multimedia event ontology has also been introduced to tackle events in media data. In Section 4, the semantics of various constructs for specifying the events, various relationships, their geographical context and uncertainty associated with them are presented. The inference scheme for event recognition and retrieval is discussed in Section 5. In this, the algorithm for construction of the OM is provided which further facilitates in reasoning process. In Section 6, we have discussed an application scenario corresponding to entertainment category in news domain. It utilizes the

proposed approach to support news aggregation and event-based recommendation by exploiting the knowledge encoded in E-MOWL based ontology. Finally, Section 7 concludes the paper by summarizing the capabilities of E-MOWL.

## 2 Event Representation and Modeling in Multimedia Domain

An event comprises of number of activities (or, sub-events) that takes place at specific time and interval and shares the same context. Event analysis contributes to enriching the events and stories with rich multimedia content. It provides insight and guidance by identifying and classifying the events. In the space of multimedia content semantic and taxonomic relations have been modelled for observables. Spatial relations and temporal relations are represented as implied observations. However, ability to model events with reference to the multimedia content can provide a mechanism to build systems with events as basic units of analysis. An ontological model of events can make taxonomic relation between events and relationship between events and event objects explicit. Observers in a multimedia record (like photo, video, audio, text) perceive, in many cases, an event which is associated with location, time, temporal/spatial interval and participants. An example is a video of a *music concert*. These instances are not necessarily same as examples of other multimedia visual concepts like desert, forest, hit etc.

Events manifest in multimedia content in various forms. But events invariably record occurrence at a temporal instance or over a temporal interval. Events can be grouped into event types. *Delivering talk* is an event class. An instance of the class would be *Santanu Sir's Talk at IITG*. A possible parent class of delivering a talk will be *Professional Event*. On the other hand, sub-event of *Delivering Talk* will be *Question and Answer*. It is obvious that like any other concept taxonomy of events are needed for high-level reasoning. Further, each event is associated with specific media property and spatial, temporal and entity related features for detection and instantiation of the events in the media stream. A multimedia event ontology requires to represent these features of an event. Any ontological formalism for representation of events, therefore, needs to address following issues:

- Informational Aspect: Entities, location, time, actors and/ or participants involved in an instance of an event distinguishes the instance from other instances.

- **Structural Aspect:** An event can be composed of sub-events spanning over time and/ or locations.
- **Causal Aspect:** Events can be causally related in a way such that occurrence of an event alter state of one or more other events.
- **Temporal Aspect:** Temporal relation between events indicate, among other things, synchronization and constraint relationships.
- **Spatial Aspect:** Spatial relation between events represent distribution of events over geographical regions.
- **Uncertainty Aspect:** Inherent uncertainty involved in causal association between events and incompleteness of the information about location, temporal interval, role and function of participants in an event as depicted in the multimedia rendering of the event requires uncertainty handling mechanisms for reasoning with events.
- **Domain Aspect:** Background knowledge about the event which can provide the semantics of vocabulary describing event classes, entities involved and relationship between events.

An ontology representation scheme for events in multimedia requires constructs to handle all the issues highlighted here. To automatically identify and classify some predefined events in multimedia document it is crucial to extract the context embedded within it. The context of the content in any document, text or multimedia, depends on the understanding of the sense. To capture the semantics of an event from multimedia content it is indispensable that the appropriate multimedia centric event model be developed and utilized. The event model representation binds an event and media, and exploits multiple modalities (image, text and audio). It further fuses information procured from each modality for event detection. The related condition that enables an event to occur is the context of an event. Context is inherent in multimedia documents, which is defined with respect to an event instance where the event context has its own structure. It involves parameters like time, geographical location or the actor/ entities involved. Event detection using ontological model and analysis of corresponding event statistics have lead to emergence of applications like tagging, aggregation and trend discovery.

### **3 Modelling of Events**

In this section, we have developed formalism for representation of events depicted in multimedia stream. We need to first define model for events because multimedia devices can only record an event but cannot reason with its media and event-related aspects.

### 3.1 Event Model

1. Definition of an event: An event is a collection of real-world objects/ entities/ sub-events that shares certain attributes. These attributes could be time, geographical location or the actors/ entities. It is defined as 5-tuple of attributes ( $Att^i$ ).  $E = \langle A, E, V, T, S \rangle$ 
  - (a)  $A$  represents informational aspect of the event - event name and action description.
  - (b)  $E$  represents entities or participants in the event.
  - (c)  $V$  refers to background or domain ontology linked with the event.
  - (d)  $T$  represents the time instance of occurrence of the event and the temporal interval of the event.
  - (e)  $S$  represents the geographical location of the event; events can also be distributed over multiple locations.

Each attribute is a structured data entity with multiple component field. Equivalence of attribute is attribute dependent. We can define *Event Class* ( $EC$ ) as a set of events which may share certain event attributes. For example, all events occurring at a spatial location i.e. having the same value of  $S$  can form an event class. Hence, all events in a class will have atleast one attribute with equivalent value. Taxonomic relationship between event classes can be constructed if one event class is subset of the other. For example, all events with same name as defined by domain ontology  $V$  will be superset of the class of the same event at a location. Formally, an event class  $EC_1$  is superset of  $EC_2$ , if there exist atleast an attribute  $Att^i$  such that  $Att^i_{EC_1} = Att^i_{EC_2}$  for all events in the event classes and  $\exists Att_j$  such that  $Att^j_{EC_1} \neq Att^j_{EC_2}$ . For example, *Earthquake* is an event class and *Earthquakes at Assam* are sub-class of the earthquake class.

There would also be non-taxonomic relations:

1. Composite Relation: An event can be composed of multiple sub-events which may be of smaller granularity over space and time. Typically, sub-events can span sub-intervals spanned by the attribute  $T$  of the event. Similarly, sub-events can span over sub-regions contained in the spatial extent defined by the attribute  $S$ . These sub-events may or may not be elements of the same event class. Hence, we have temporal as well as spatial composition of events giving rise to global events. Formally, we define composition operators  $\otimes_{time}$  and  $\otimes_{space}$ .

$$e_1 \otimes_{time} e_2 \dots \otimes_{time} e_n \rightarrow E_1$$

$$e_i \otimes_{space} e_j \dots \otimes_{space} e_r \rightarrow E_2$$



There can be also combination of temporal and spatial compositions. Obviously, attributes of the composite event will be a function of attributes of sub-events. These functions can be event specific, however sub-events have to satisfy the constraints defined for  $T$  and  $S$  attributes of the composite event.

2. Causal Relation: If occurrence of an event has a causal dependence on another event, then two events are considered causally related. More formally,  $E_1 \xrightarrow{\text{causes}} E_2$ .
3. Temporal and Spatial Relation: Events can have spatial and temporal relations among them. All standard spatio-temporal relations can be specified between events. For example, if one event occurs after another event in time then events satisfy temporal relation- *follow*:  $E_1 \succ E_2$  i.e.  $E_2$  follows event  $E_1$ . Similarly, if event  $E_2$  occurs in the same geographical region as  $E_1$  then  $E_2 \gamma_{in} E_1$ . An event instance *House Crash* occurred in Kathmandu located in Nepal where *Earthquake* event has occurred then  $\text{HouseCrash} \gamma_{in} \text{Earthquake}$ .

### 3.2 Event and Content

A model for representation of an event is defined in the previous section. This model enables us to consider events as basic entity for any reasoning or analytics applications. These events are typically recorded through multimedia devices and stored in different forms like image, text, video, audio or combination. Event attributes are mapped to content associated with events through application of multimedia processing techniques and/ or manual annotation. For example, given a video of an event, shots can correspond to sub-events. Automatic content analysis techniques can interpret content of shots to associate with informational attributes of events like participants in the event.

Multimedia event ontology, linking event to multimedia content, is built upon model of the event, association of the event with the conceptual specification of the event type and mapping attributes of the events to the observable features in the multimedia content. In other words, ontology specifies the observation model of the conceptual representation of the event factoring in the inherent uncertainty of association between detected observables and expected observations of the event attributes. In Section 4, we present a new ontology language for representation of the conceptual model of content based observation of an event. The ontology also provide semantic specification of relations between events and event attributes. We

extend multimedia web ontology language (MOWL) to build *multimedia event ontology language* (E-MOWL).

### 3.3 Basic Features of MOWL

Multimedia Web Ontology Language (MOWL) works with a causal model of the world, in which the real-world concepts lead to manifestation of media features in multimedia documents. This causal modeling is the main difference between MOWL and other knowledge representation languages like OWL. The causal model is the basis for abductive reasoning for concept recognition in multimedia data, where the media features observed in a multimedia document are causally explained as manifestations of concepts. Another important aspect is the formal definition of spatio-temporal relations between observable media patterns. MOWL supports perceptual modelling of a domain, wherein the concepts exhibit into media patterns in a media document and helps in detection of media objects. A media object can have some media properties which are typically like a constraint on a low-level media feature (e.g. color = yellow); or a composite media feature at a higher level. The latter can be a media pattern say an audio pattern, a body posture, a simple human action or a face pattern. These patterns can be recognized in the media data, with the help of some media feature classifiers or media pattern detectors. Thus, detection of a media object in a media instance requires some pattern detectors for its media properties.

Uncertainty is inherent in media observations and so, abductive mode of reasoning with uncertainty handling results in a robust concept recognition scheme. For this, the OM is utilized which is a probabilistic graph model that depicts the constraints in a graph model. Reasoning for derivation of the OM for a concept requires exploring the neighborhood of a concept and collating media properties of neighboring concepts, whenever media propagation is implied. The resultant OM of a concept is then organized as a Bayesian Network. Once the OM for a concept is created, it can be used for concept recognition and abductive reasoning scheme exploits the causal relations captured in the OM. MOWL has many advantages when it comes to multimedia ontology representation. It provides for the following aspects:

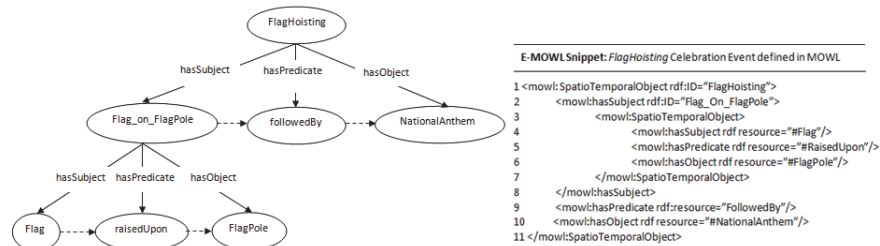
- perceptual modeling of concepts through their association with observable media patterns;
- spatio-temporal relations between concepts which can represent objects, by providing constructs for their modeling;

- uncertainty specification in concept relations by allowing Conditional Probability Tables (CPTs) to be encoded in the ontology;
- a robust inference mechanism to support probabilistic reasoning for concept recognition and recommendation based on Bayesian networks.

### 3.4 Multimedia Event Ontology

An event occurs inevitably in multimedia content, thus the semantic analysis and interpretation of multimedia content is required to identify the context of an event. The context information is implicitly linked to multimedia content and it comprises of timestamp, location and actor/ entities involved. So, the key insight in our proposed approach is to infer event related aspects from the lower-level multimedia content. This requires exploitation of domain *multimedia event ontology* handcrafted by a group of domain experts wherein, the nodes are the event/ entity classes and edges between nodes indicate a variety of event/ entity relationships, namely, spatio-temporal, spatial, temporal, causal and others. To describe an event class we have utilized the representation scheme for specification of spatio-temporal relations proposed in (Wattamwar and Ghosh, 2008). Any complex media object constitutes of several simpler media objects belonging to multiple media forms which are interconnected with Spatio-Temporal and/or temporal relations (Papadias et al., 2001). To capture the linkage between the primitive level sub-events so as to identify a complex event requires identification of semantic events from audio-visual data with spatio-temporal support. Figure 1 shows the MOWL encoding for the respective event. The *multimedia event ontology* can formally be defined as the conceptual specification of meaningful relations between the various concepts and can be represented as follows:

Multimedia Event Ontology = {C, R, A, I, O}, where C (concepts) represents the concept set comprising of real world events; R (relation) is a



**Figure 1** MOWL specification for spatio-temporal event *FlagHoisting*.

relation set and it mainly describes the event specific relations as observed in lower-level multimedia concepts; A (attribute) represents the various attribute set of events; I (instances) is a set of definition about an event instance; O (Observation Model) is a probabilistic graph model depicting the relation between concept or events with associated observable in terms of probabilistic conditional dependency. In other words, O is the belief network (Neapolitan, 2012) and reasoning is through belief propagation in the OM. In general, multimedia attributes are not restricted to just audio-visual properties but can include wider variety of contextual content. Events in audio-visual media like videos have Temporal and Geographical connotations which can be combined under an umbrella term *Context of the event*. Context of an event can extend to the Actor/Entities (and their role etc.) involved in it. But it lacks effective representation schemes. The lack of formal model for events hampers the identification and retrieval of various aspects that are associated with an event. Therefore, a representation language and event model, in terms of media-based properties of events provides a mechanism for representation and recognition of events.

#### 4 Language Constructs

To formalize the concepts and relationships among various multimedia and event related aspects in multimedia document we need to extend the existing MOWL language so as to represent the collective semantics using these relationships. We have designed E-MOWL as an extension to MOWL. Therefore, apart from MOWL constructs some additional language constructs have also been defined. This has been done so as to incorporate event information in classes, individuals and properties for realisation of the event model described in section 3. Also, it is important to account for the fact that an event takes place at certain location and if a user wishes to search for some event that is related to some place, geospatial knowledge plays a crucial role. The geospatial knowledge assists in identifying the locations and the relationship between them which can be referred to using the place names. So, along with the event related aspects it is required that the ontology incorporates the geographical terminology and explicit spatial relationship between these geographic entities. This facilitates in specifying the geographical context of an event that may refer to some location exactly and some cases approximately. The other relations that bind concepts and event objects are causal and uncertain in nature. To capture these uncertainties we have used BN based probabilistic reasoning scheme.

The language E-MOWL offers extensive support for constructing multimedia event ontology and facilitates recognition of high-level complex event, based on the hierarchy of sub-events present in the ontology. Although E-MOWL provides constructs to store substantial knowledge about an event like temporal, spatial and agent/entity related aspects but to pinpoint the geographical context of an event, geospatial knowledge plays a crucial role. The meta-data that geographically identifies a concept/event, termed as geo-tag, is associated with various media data (*say* video or geo-tagged photograph). The spatial class of E-MOWL has the geographic concept that is linked to multimedia content and it assists in identifying the geospatial relations that exists between the geographic entities. To extract the geographical context of a concept the information has to be presented in a structured way that is absent in MOWL. These geo-tags contain information that, if incorporated into MOWL can be exploited to perform more complete semantic processing. The practicality of geospatial knowledge is not limited to extracting out the geographic entities but can also be used to draw the geographic relation that exists corresponding to an event. For instance, if Turkey is not named in a news related to Istanbul, then to capture geographic relations like *Istanbul is\_in Turkey*, geospatial knowledge can be utilized.

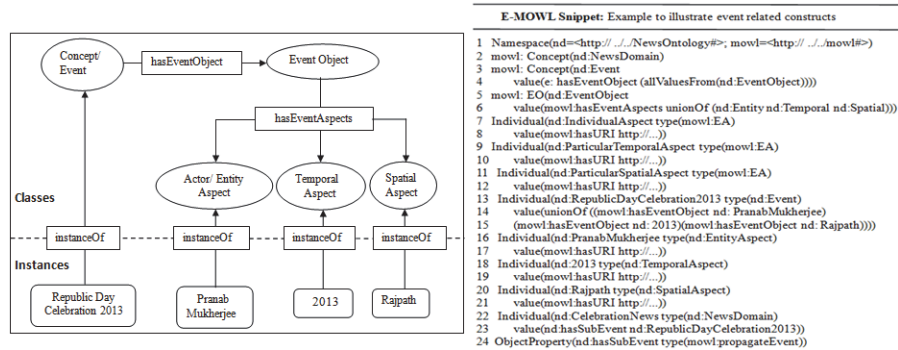
#### **4.1 E-MOWL: Concepts, Event Objects and Event Relations**

The language E-MOWL has been developed by building on the framework provided by MOWL. Although MOWL has the capability for describing media manifestations of the ontology concepts by associating media properties with these manifestations, but MOWL needs to be enriched to represent events. Moreover, it is required that the language must incorporate probabilistic reasoning methods so as to provide the capability of probabilistic inference while detecting an event in the presence of uncertainty. Encoding of event properties of the concepts is accomplished using an extended multimedia web ontology language E-MOWL. The language allows specification of event properties, their geographical context and supports uncertainty reasoning as observed in events. Identification of contextual information present in multimedia documents further facilitates automatic multimedia tagging. We have, unlike other approaches, exploited our ontological framework for providing automatic context-based multimedia tagging and have aggregated diverse documents corresponding to a query video. The document collection facilitates in providing appropriate contextually relevant information associated with the video comprising of other videos and text news. Thus providing complete description of the event for the user.

Apart from the ones that exist in MOWL, E-MOWL extends and includes two new classes to represent event related aspects – *Concept* and *EventObject*. E-MOWL is an extension of the MOWL with Spatio-Temporal relations and uses the MOWL language constructs to define the classes and properties related to an event. The class  $\langle mowl: Concept \rangle$  represents real-world events. The other class  $\langle mowl: EventObject \rangle$  is the manifestation of events in different media forms. It comprises of various sub-events that are collated to form a higher-level event. The class  $\langle mowl: EventObject \rangle$  is linked to following classes representing the three event related aspects:

- Spatial property class  $\langle mowl: SpatialAspect \rangle$  stipulates various spatial properties associated with an event. This includes information like city, district, latitude and longitude etc. about a particular place in which the event is occurring or has already occurred.
- Temporal property class  $\langle mowl: TemporalAspect \rangle$  specifies the time related aspects of an event. This associates the date, day and time related information to the event description.
- Actor property class  $\langle mowl: ActorAspect \rangle$  that addresses the people or things that are involved in an event. For example, information about person involved and other related aspects.

The snippet depicting the association of concepts and event objects for an event in political category is shown in Figure 2 where, EO represents Event Object Property and EA represents Entity Aspects. To describe this snippet ASN (Patel-Schneider et al., 2004) has been referred. The spatial class in E-MOWL also provides conceptualization of the *geographic* terms. The geographic data is considered as a class of spatial data which uniquely



**Figure 2** E-MOWL specification illustrating event related constructs.

identifies the location and/ or the geographical boundaries of spatial entities. The geographic knowledge plays a key role in providing the representation of appropriate geographic data/information to associate events from the multimedia content to respective location or place of their occurrence. The semantic processing of geographical knowledge facilitates in identifying the geographic context corresponding to events in multimedia applications. To elucidate, let's assume that *gunshot* event has taken place in a bank. There are multiple spatial locations associated with an event gunshot. For example, a video can record the *gunshot* which occurred at a spatial location. Geographical or spatial knowledge can help to interpret the location. Ontology based reasoning may establish the location as entrance to a known region.

In E-MOWL, to identify the geographical context of an event one needs language constructs to specify geographical/spatial properties, property propagation rules and specification of conditional probabilities. We define a new abstract class *<mowl:GeographicEntity>* for specifying the geographic properties of lower-level media objects. The property constraints can also be associated with the geographic properties of the concepts. The class *GeographicEntity* is further divided into 2 sub-classes. These are *GeoPhysicalEntity* and *GeoPoliticalEntity*.

- The class *<mowl:GeoPhysicalEntity>* specifies the information related to physical features of the geographic entity say its position, orientation etc. For example, the geographic entity, Pulicat Lake *is\_in EasternIndian* coast, *is\_in* direction *NorthOfEquator* and *is\_near Sriharikota* island etc. will be specified by an instance of this class.
- The class *<mowl:GeoPoliticalEntity>* considers the information related to political features of a geographic entity say information about state, country etc. For example, Pulicat Lake *is\_in* district *Nellore* which *is\_in* state *AndhraPradesh* and Andhra Pradesh *is\_located\_in* country *India* will be represented through this class.

Various event-based relations which relates a concept to event objects or relates two event objects are described as follows:

1. *<mowl:hasEventObject>* relation associates a concept with an event object or binds two event objects. This is a transitive relation which means that if for any concept *C* and event objects *E1*, *E2*, *E3* there exists a *hasEventObject* relation denoted as *hasEO*, then  
 $\text{hasEO}(C, E1) \text{ and } \text{hasEO}(E1, E2) \Rightarrow \text{hasEO}(C, E2)$   
 $\text{hasEO}(E1, E2) \text{ and } \text{hasEO}(E2, E3) \Rightarrow \text{hasEO}(E1, E3)$

This is also a causal relation, so E-MOWL allows for probabilities to be attached to the entities in this relation.

2. <mowl:hasSpatialAspect> property associates an event with the geospatial location where the event took place. Eg. *occurred\_at*, *is\_in* etc.
3. <mowl:hasTemporalAspect> relation binds an event with the time at which any particular event has occurred. Eg. *occurred\_on*, *has\_date*, *has\_day* etc.
4. <mowl:hasAgentAspect> relation is for the agents or the entities that are involved in an event. Eg. *done\_by*, *has\_Role*.

The geospatial location of the *hasSpatialAspect* property further binds the spatial location of occurrence of an event with another geospatial locations and usually falls into following categories:

- Inclusion: *is\_in*, *is\_part\_of*
- Neighbour/Adjacency: *is\_near*, *touches*
- Direction: *is\_in\_dir*, *has\_coordinates*

The language constructs have been extended to incorporate two important features namely, *Event Property Propagation* and *Uncertainty Reasoning*.

#### 4.1.1 Event propagate property

To specify relations which do not imply a concept hierarchy, but allow propagation of event properties across connected concepts in the ontology, E-MOWL defines a <mowl:propagateEvent> property. It specifies that if a particular event occurs at any place and time then its spatial and temporal properties i.e. the time and location properties is propagated to its sub-events also. Since, sub-event is not a part of taxonomic hierarchy, wherever applicable we need to explicitly indicate property propagation because in some cases sub-events can occur at geographically distinct locations and at different time instances.

#### 4.1.2 Uncertainty specification

The E-MOWL constructs that support uncertainty specification are described in the following schemas:

1. <mowl:CPTable> class allows defining CPT in the ontology. The CPT must state the concept or event object to which it is associated. The probability values associated with each row of the CPT is based on the parent-child combination as shown in Figure 3.



```

<mowl:ObjectProperty rdf:ID="hasCPT">
  <rdfs:comment> Property to associate a CPT with a Concept
</rdfs:comment>
  <rdfs:domain>
    <mowl:unionOf>
      <mowl:Class rdf:resource="#Concept"/>
      <mowl:Class rdf:resource="#EventObject"/>
    </mowl:unionOf>
  </rdfs:domain>
  <rdfs:range rdf:resource="#CPTTable"/>
</mowl:ObjectProperty>

```

**Figure 3** Snippet: Property to associate a CPT with a concept.

```

<mowl:Class rdf:ID="CPTTable">
  <rdf:label>Conditional Probability Table</rdf:label>
  <rdfs:comment>Class for encoding the uncertainty and encapsulating the CPTs.
</rdfs:comment>
  <rdfs:subClassOf>
    <mowl:Restriction>
      <mowl:onProperty rdf:resource="#conditionedOn" />
      <mowl:minCardinality rdf:datatype="&xsd;nonNegativeInteger">
        1 </mowl:minCardinality>
    </mowl:Restriction>
    <mowl:Restriction>
      <mowl:onProperty rdf:resource="#hasCPRow" />
      <mowl:minCardinality rdf:datatype="&xsd;nonNegativeInteger">
        1 </mowl:minCardinality>
    </mowl:Restriction>
  </rdfs:subClassOf>
</mowl:Class>
  <mowl:ObjectProperty rdf:ID="conditionedOn">
    <rdfs:comment> Specifies the parents of a concept or an EventObject conditioned on
    which the CPT entries are based.
  </rdfs:comment>
    <rdfs:domain rdf:resource="CPTTable"/>
    <rdfs:range>
      <mowl:unionOf>
        <mowl:Class rdf:resource="#Concept"/>
        <mowl:Class rdf:resource="#EventObject"/>
      </mowl:unionOf>
    </rdfs:range>
  </mowl:ObjectProperty>
  <mowl:ObjectProperty rdf:ID="hasRow">
    <rdfs:domain rdf:resource="CPTTable"/>
    <rdfs:range rdf:resource="CPRow"/>
  </mowl:ObjectProperty>

```

**Figure 4** Snippet: Class for encoding the uncertainty and CPTs.

2. `<mowl:hasCPT>` property associates a CPT as defined by `<mowl:CPTTable>` with an E-MOWL concept or event object as shown in Figure 4.

The uncertainty specification has been depicted in the snippet shown in Figure 5. It defines the CPT for the event object `FlagHoisting` conditioned on the concept `RepublicDayCelebration`. The CPT encodes the uncertainty related with the observation that `FlagHoisting` is a sub-event of `RepublicDayCelebration`. While assigning conditional probabilities it should be noted that a concept or event object in the multimedia event ontology can have only two states. Accordingly, it can have values true/false, or 1/0. If each parent node has 2 states then for  $n$  parents the number of states will be  $2^n$ .

```

<mowl:Concept rdf:ID="RepublicDayCelebration">
  ...
</mowl:Concept>
<mowl:EventObject rdf:about="#FlagHoisting">
  <mowl:hasCPT>
    <mowl:CPTTable>
      <mowl:conditionedOn rdf:resource="#RepublicDayCelebration"/>
      <mowl:hasRow>
        <mowl:CPRow>
          <mowl:ParentStates> "1" </mowl:ParentStates>
          <mowl:probValues> 0.75 0.25 </mowl:probValues>
        </mowl:CPRow>
      </mowl:hasRow>
      <mowl:hasRow>
        <mowl:CPRow>
          <mowl:ParentStates> "0" </mowl:ParentStates>
          <mowl:probValues> 0.30 0.70 </mowl:probValues>
        </mowl:CPRow>
      </mowl:hasRow>
    </mowl:CPTTable>
  </mowl:hasCPT>
</mowl:EventObject>

```

Figure 5 Snippet: Uncertainty specification.

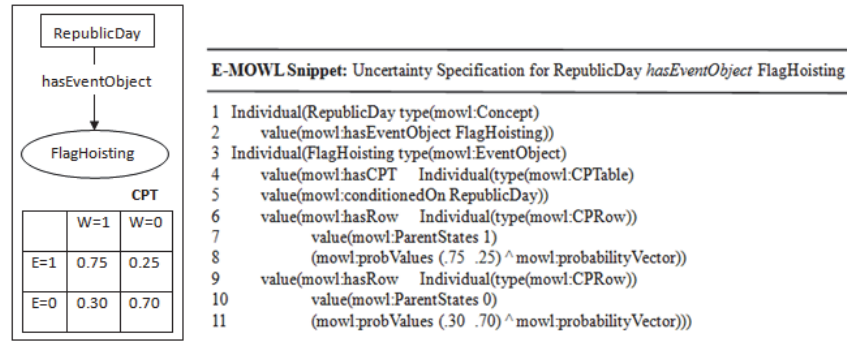


Figure 6 Uncertainty specification in E-MOWL for RepublicDay hasEventObject FlagHoisting.

For instance, the CPT for a concept with single parent ( $n=1$ ) will have 2 rows; CPT with two parents will have 4 rows and so on. Figure 6 depicts the definition of CPT for Event Object *FlagHoisting* conditioned by the concept *RepublicDay*.

The proposed representation scheme offers the following benefits:

1. Apart from encoding media features, media instances and spatio-temporal relations between objects the language allows a representation mechanism to link events to their media depictions.
2. The additional language constructs have been defined that facilitates in representing the complex events by embodiment of event aspects along with the relations that associate event properties with these aspects in the multimedia ontology.

3. The language also has the capability to specify the geospatial knowledge in the ontology so as to identify the geographical context of an event.
4. It specifies event property propagation rules that allows related events to share the context.
5. It offers BN based probabilistic reasoning scheme to deal with uncertainty associated with events and to infer higher-level concepts (or events) on basis of some evidences. The inference framework is different from the one that exist in MOWL as in present case apart from media observables event aspects are considered additionally for event detection in multimedia content.

## 5 Inference Scheme

Events are uncertain in nature and always occur in some context. The context information associated with an event is knowledge-based and it has to be guided and interpreted by ontology. To fully exploit the features of the extended multimedia ontology for events, one requires an inference framework. Its apparent that the higher-level event can be inferred more accurately by utilizing the media aspects which additionally act as an evidence of an inferred event. For instance, an image of *Taj Mahal* may witness the event *Performance* at Taj Mahal which in turn may serve as an evidence that an event *Concert* took place at Taj Mahal. Besides the multimedia properties specified in MOWL, in E-MOWL, recognition and retrieval of a concept is based on the event properties of the observable lower-level multimedia concept. For assigning the conditional probabilities between various concepts, belief network can be used. Let  $E$  be a set of event related information which contains elements such as temporal, spatial and agent/entity aspect i.e.  $\{e_1, e_2, \dots, e_l\} \in E$ . Further, let  $M$  be a set of multimedia information with elements as multimedia features i.e.  $\{m_1, m_2, \dots, m_k\} \in M$ . In the traditional setting of MOWL, if a concept  $c$  contains a multimedia feature  $m_i$ , then observation of  $m_i$  provides some evidence towards presence of the concept. When such a setting is extended to E-MOWL, the observation of event aspects  $e_j$  provides further evidence towards the presence of higher-level event (provided the concept contains the event related aspects  $e_j$ ). It enables the construction of the OM in the form of Bayesian Network, which is the event-centric description of a lower-level multimedia concept. The weights of the links that connect the concepts to the event aspects in the Bayesian network represent the uncertainties associated with the semantic interpretation of event specific information. This information can be exploited for drawing

information about the presence/absence of an event. The inference framework for event recognition and retrieval is different from the one that exists in MOWL since in present case, the event objects are considered additionally for the purpose of reasoning. E-MOWL handles the uncertainty in multimedia documents by providing various constructs to define the CPTs and binds them with concepts and event objects. We have utilized BN based reasoning scheme which facilitates in finding the probability of presence of an event in the multimedia data on the basis of context information. In our scheme, BN is a probabilistic acyclic graphical model which represents the causal relation between the lower-level multimedia concept and the expected event aspects. We fix the probabilities of the links, i.e. each parent-child pair combination and based on these, individual probabilities of each node is calculated using BN for inferencing.

The Inference framework involves two stages of reasoning for event recognition. In the first stage the OM (a probabilistic graph model) with the relevant event objects and media patterns is created by exploring the appropriate subgraph of the ontology graph. In the second stage, reasoning for event recognition is done using the OM. In this, the probability of the parent node is calculated based on the findings of some of the child nodes. If the probability of the parent node exceeds a certain threshold value then we say that an event has been detected. In the following section, we detail the inference framework that facilitates in dealing with uncertainty associated with events.

### 5.1 Constructing the OM

The OM is a probabilistic graph model depicting the constraints in a probabilistic framework, and the entire reasoning is done using it. The recursive steps involved in constructing the OM for a concept ( $\Upsilon$ ) using an ontology graph ( $\Pi$ ) are given in Algorithm 1. The OM is initialized with the root concept. The neighbour nodes of the root concept are considered which could be either *Event Object* or *Sub-event*.

- If the found node is an event object then the procedure *addEventObject* is called. If the event object is a leaf node then return it. If event object is not a leaf node then find its neighbour. If the found neighbour is ancestor to the event object then add a link from the neighbour to the event object else add the link from event object to its neighbour. Recursively call the same procedure for other event objects as well.

- If the found node is sub-event then the procedure *addSubEvent* is called. If the sub-event is a leaf node then return it. If sub-event is not a leaf node then find its neighbour. If the found neighbour is ancestor to the sub-event then add a link from the neighbour to the sub-event else add the link from sub-event to its neighbour. If there exists an event object then call the procedure *addEventObject*. Recursively call the same process for other sub-events as well.

## 5.2 Reasoning for Event Recognition using OM

Here, we try to recognize a conceptual event based on the properties of the concept. The OM created in the above step is used for event recognition. For assigning the conditional probabilities between concepts, BN is used. The BN reflects the causal relations between the various events and media patterns that can be expected in a manifestation of an event in a multimedia document. The root node represents the event (or sub-event) and the leaf nodes represent the contextual event instance and observable media patterns.

The steps involved in recognizing a concept are given in Algorithm 2. The input to this reasoning process is the multimedia document  $D$ . For each sub-event or event object in the OM, a procedure (detailed in Algorithm 2) is called to detect its presence. If the detection of a sub-event or an event object returns a TRUE value then, that node is instantiated in the Bayesian Network. In order to achieve this, the nodes in the graph are traversed and an initial probability value is assigned to the concept to be recognized. These traversed nodes could be either sub-event or an event object. In case the traversed node is sub-event then from the total number of sub-events only those concepts are considered that helps in recognizing an event. Their belief is then computed and the corresponding nodes are instantiated in the BN. On the other hand, if the traversed node is event object then, we use Natural Language Processing parser (<http://nlp.stanford.edu/software/lex-parser.shtml>) to identify the event object which could be place, time or an entity (person), depending on which we compute their belief values and the corresponding nodes are instantiated. Based on these cases the leaf nodes are processed and belief propagation takes place in the OM and posterior probability of the root node is computed. If this value exceeds a certain threshold  $t_v$ , then a concept is present in  $D$  and is returned. Initially, the concept (event) to be recognized is assigned a probability value which specifies the initial belief it has in a multimedia document.

**Algorithm 1:** Construction of an OM

---

**Input:** a) Multimedia Ontology for Events Graph  $\Pi$ ,  
b) Root Concept  $\Upsilon$   
**Output:** Observation Model  $\Theta$   
**notations:** EventObject- $\xi_c$ , Sub-event- $\eta_c$

**procedure** MAIN

- 1: Initialize  $\Theta$  with  $\Upsilon$  as the root node;  $\Theta \leftarrow \Upsilon$
- 2: Read neighbor node  $\xi$ , associated with  $\Upsilon$ , where  $\xi$  could be either  $\xi_c$  or  $\eta_c$
- 3: addEventObjects  $\xi_c$  to  $\Upsilon$
- 4: addSubEvents  $\eta_c$  to  $\Upsilon$
- 5: Compute CPTs from  $\Pi$  and attach it to nodes in  $\Theta$
- 6: **procedure** addEventObjects  $\xi_c$
- 7:   **if**  $\xi_c$  is a leaf node in  $\Pi$  **then**
- 8:     RETURN
- 9:   **end if**
- 10:   **else** do the following:
- 11:     **if**  $\xi$  is an Event Object  $\xi_c$
- 12:       **for** other Event Objects  $\xi_c(k)$  in  $\xi_c$ , where,  $1 \leq k \leq n$
- 13:         **if**  $\xi_c(k)$  is an ancestor to  $\xi_c$  **then**
- 14:         add a link from  $\xi_c(k)$  to  $\xi_c$
- 15:         **else** add  $\xi_c(k)$  as new child of  $\xi_c$
- 16:         addEventObjects  $\xi_c(k)$
- 17:         **end if**
- 18:       **end for**
- 19:     **end if**
- 20:   **end procedure**
- 21: **procedure** addSubEvents  $\eta_c$
- 22:   **if**  $\eta_c$  is a leaf node in  $\Pi$  **then**
- 23:     RETURN
- 24:   **end if**
- 25:   **else** do the following:
- 26:     **if**  $\xi$  is a Sub-Event  $\eta_c$
- 27:       **for** other related concepts  $\eta_c(k)$  (sub-event or event object) in  $\eta_c$ ,  
where,  $1 \leq k \leq n$
- 28:         **if**  $\eta_c(k)$  is an ancestor to  $\eta_c$  **then**
- 29:         add a link from  $\eta_c(k)$  to  $\eta_c$
- 30:         **else** add  $\eta_c(k)$  as new child of  $\eta_c$
- 31:         **end if**
- 32:         addSubEvents  $\eta_c(k)$
- 33:         addEventObject  $\eta_c(k)$
- 34:       **end for**
- 35:     **end if**
- 36:   **end procedure**

**end procedure**

---

---

**Algorithm 2:** Reasoning using OM

---

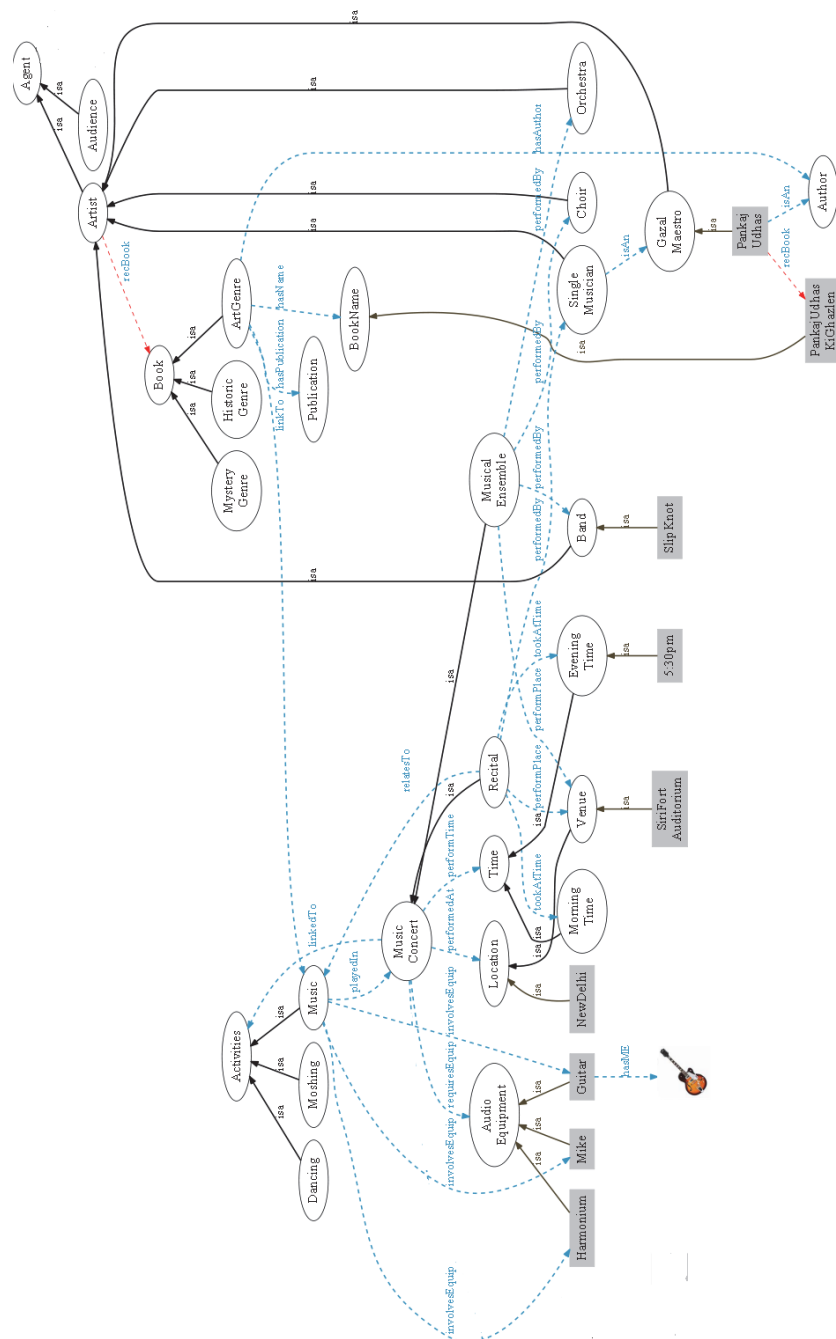
**Input:** a) Multimedia Document  $D$ b) Let the total number of concepts in the OM be  $v$ **Output:** Recognition of a concept  $\Upsilon$  in multimedia document  $D$ **procedure** MAIN

```

1: Assign a-priori probability value  $p_v$  to  $\Upsilon$  to get Belief ( $\Upsilon_p$ )
2: While traversing nodes could be either sub-event ( $\eta_c$ ) or event object ( $\xi_c$ )
3: if traversed node is sub-event  $\eta_c$  then
4:    $\eta_k = \{k: k \text{ is a sub-event in } v\}$ 
5:   for  $i=1$  to  $k$  in  $\eta_c(k)$  do
6:     compute Belief( $c_k$ ) in  $D$ 
7:     if Belief( $c_k$ ) is TRUE then
8:       Instantiate node ( $c_k$ )
9:     end if
10:  end for
11: else
12: end if
13: if traversed node is  $\xi_c$  then
14:    $\xi_k = \{k: k \text{ is a event object in } v\}$ 
15:   for  $i=1$  to  $k$  in  $\xi_c(k)$  do
16:     run NLP parser to identify  $\xi_c(k)$  as place  $p$ , time  $t$  and entity  $e$ 
    associated with events
17:     if  $\xi_c(k)$  is  $p$  then
18:       compute Belief( $c_k$ ) for  $p$  in  $D$ 
19:     else if  $\xi_c(k)$  is  $t$  then
20:       compute Belief( $c_k$ ) for  $t$  in  $D$ 
21:     else
22:       compute Belief( $c_k$ ) for  $e$  in  $D$ 
23:       if Belief( $c_k$ ) is TRUE then
24:         Instantiate node ( $c_k$ )
25:       end if
26:     end if
27:   end for
28:   Propagate Belief in  $\Theta$ 
29:   Belief  $\Upsilon_p \leftarrow$  Posterior Belief of  $\Upsilon_p$ 
30:   if Belief  $\Upsilon_p \geq t_v$  then
31:     return TRUE
32:   else
33:     return FALSE
34:   end if
35: end if
end procedure

```

---



**Figure 7** Multimedia event ontology snippet.



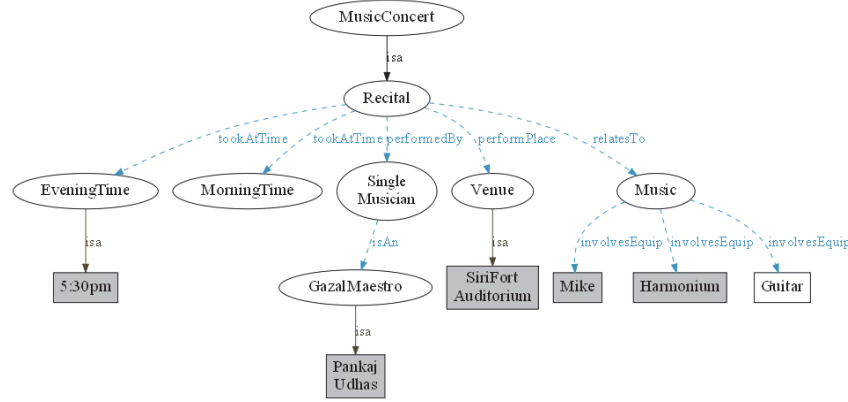
## **6 Application Scenario**

Initially, we provide a brief introduction to the News domain since we have used this domain as test bed to check the effectiveness of our ontology based approach. News act as a record of different kinds of events typically recorded on media and may be categorized into national, international, business, sports, entertainment etc. These are available to the users through various sources like newspapers, television and internet.

We provide an illustration of an application *News Aggregation* that comprises of multi-modal documents integrated from diverse sources to provide us with a complete description of news event. The selection of documents is based on context information identified automatically from the input video. Every event has context-related information associated to it such as time of occurrence, location and entities involved. The context corresponding to the input video facilitates collection of diverse documents from multiple sources using Google custom search engine. The retrieved results are further filtered using our proposed approach and only contextually relevant results (referred to as document collection) are returned to the user. The background knowledge corresponding to the news domain is encoded using the extended language E-MOWL. The obtained features and event related aspects are mapped onto the ontology concepts and there is belief propagation in the Bayesian network. We have considered video of a music concert event wherein, the higher-level event is identified from the input video based on multimedia event ontology that involves reasoning with event properties of the concepts. To draw an inference about the higher-level event and to deal with the uncertainty associated with events, BN based probabilistic reasoning scheme has been utilized. In belief updation and propagation, the probability of the root node is calculated based on the findings of some of the descendant nodes. If the probability of the parent node exceeds a certain threshold, then we say that an event has been detected. Apart from this the semantics of the multimedia content and the meta-data further facilitates recommendation process. The proposed system dynamically suggests other event related resources say books, movies to enhance the knowledge on the subject thus defining a complete study material.

### **6.1 Multimedia Ontology Based News Aggregation**

To achieve this, first we need to create the knowledge base. The background knowledge corresponding to the present event is encoded using the extended language E-MOWL and a snippet of the same is shown in 10. This snippet



**Figure 8** OM construction for the concept recital.

shows the ontology for the event MusicConcert belonging to entertainment category consisting of approximately 160 concepts. The event is linked to other media and event related aspects such as the date and time of music event (13th Feb 2016, 5:30 pm), the actor involved (Pankaj Udhas) and location (New Delhi) of the event. In this ontology, the domain concepts, spatial aspects, temporal aspects and agent aspects are shown as ellipses and the observable media properties and event instances are shown in rectangles. The arrows represent sub-class relations between concepts and with media or event properties. For example, the link which connects MusicConcert with Recital represents that Recital is a sub-event of MusicConcert. This ontology is further utilized to construct OM for the different concepts in the entertainment domain by following the algorithm discussed in Section 5. The OM for the concept Recital, sub-event of MusicConcert is shown in Figure 8. This OM is obtained by considering the root concept Recital linked to various other concepts through media and event related properties. In particular, the belief in concept Recital is supported with formation of the event-related concepts Single Musician, Evening Time and Venue through recognition of related event instances like Pankaj Udhas, Harmonium, Mike etc. which are obtained by running several media detectors and classifiers such as Corr-LDA.

### 6.1.1 Correspondence LDA (Corr-LDA)

The Corr-LDA model extracts conditional relationships between the set of image regions and set of words. Here, the features of an image are first

obtained and then the associated words are generated. Consider the size of dictionary to be  $M$  and assume that an image comprises of  $N$  finite regions. While annotating the images, initially descriptors for various image regions are generated and then for each of the textual annotations, a region is selected. That is, for each of the  $M$  words, one of the regions is chosen from the image and subsequently a word is drawn conditioned on the topic that generates the selected region (Blei and Jordan, 2003). The Corr-LDA model has been used to model the content information of multimedia data and is trained using the dataset of the type that one expects to find in entertainment domain such as Music Concert, Celebrity Award Function etc. A total of 30 annotations were used to train the Corr-LDA model. Example of the annotations are guitar, car, person, stage, screen, bat, sky, tree, ball, trophy, pitch, wall, ground etc. The test images are then provided as input to predict the basic annotation based on image content. The basic idea behind our model is to annotate the frames corresponding to a video. The training data comprises of manually annotated images. For example, an image in the *entertainment* category can be more likely annotated with *guitar*, *stage*, *screen*, *mike* etc. Now, for testing, the new unlabeled video frames are given as input which gets annotated based on this training data. If  $r = \{r_1, r_2, \dots, r_N\}$  denotes the set of image features,  $w$  denotes the set of associated words,  $z = \{z_1, z_2, \dots, z_N\}$  is the set of latent variables,  $y = \{y_1, y_2, \dots, y_M\}$  is the set of equiprobable indexing variables and  $\theta$  is the Dirichlet random variable, then the joint probability distribution is given as follows:

$$p(r, w, \theta, z, y) = p(\theta \mid \alpha) \left( \prod_{n=1}^N p(z_n \mid \theta) p(r_n \mid z_n, \mu, \sigma) \right) \cdot \left( \prod_{m=1}^M (p(y_m) \mid N) p(w_m \mid y_m, z, \beta) \right) \quad (1)$$

The calculated histogram corresponding to each image feature vector along with the manual annotations were given as an input for training the Corr-LDA (Pahal et al., 2015). The histogram for the test images were similarly calculated and using the image-annotation probabilities for the test images, only those annotations with probability greater than certain threshold were considered. We have sampled videos into frames at the rate of 1 frame per 2 seconds to generate the images. The annotations for the various video frames whose  $p(\text{annotation} \mid \text{image})$  is more than a threshold value (0.1) are considered in our case. We have computed the confusion matrix

		Predicted						
		Person	Guitar	Harmonium	Mike	Screen	Trophy	Theatrical Smoke
Actual	Person	78	0	0	0	0	0	0
	Guitar	0	50	0	2	0	0	0
	Harmonium	0	0	59	0	0	0	0
	Mike	0	0	0	58	0	2	0
	Screen	0	0	0	0	48	0	0
	Trophy	0	0	0	3	0	57	0
	Theatrical Smoke	0	0	0	0	0	0	62

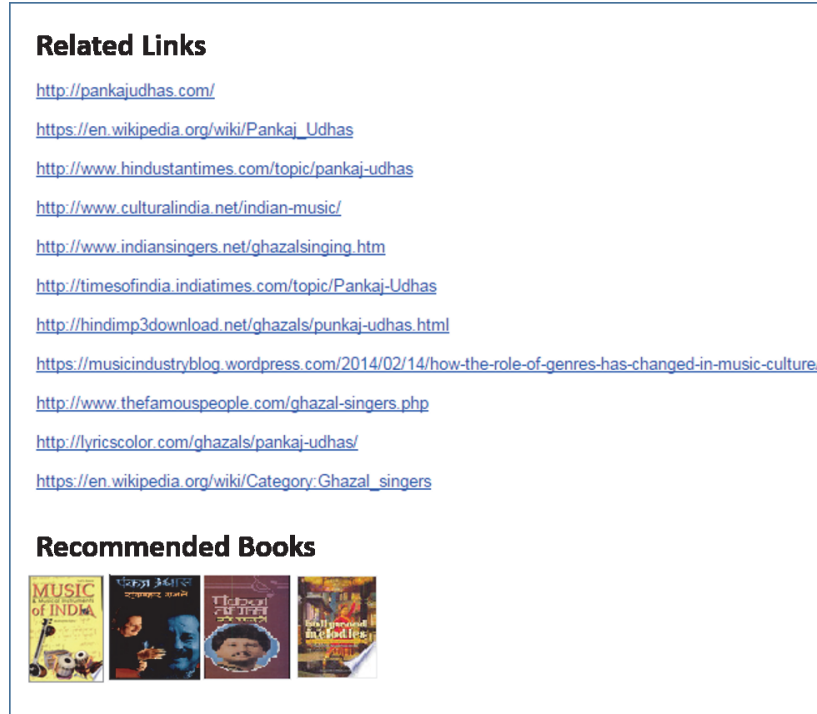
**Figure 9** Confusion Matrix for entertainment category.

from the annotations (meta-data) that are available with the videos. Each column of the matrix represents the annotation in the predicted class while each row represents the actual tags. The confusion matrix between actual and assigned annotations for video frames corresponding to entertainment category is shown in Figure 9. Although, we have observed that there is limited inaccuracy in determining the annotations but these do not have a major impact on the context identification.

### 6.1.2 News aggregation

Depending on the context of input video, we have collected information from multiple sources like blogs, text news and other news sites by utilizing Google custom search engine (<https://cse.google.com/cse/>). The corpus of around 100–125 documents have been crawled and the same is stored in a temporary database. The HTML parser has been utilized to extract the text content from these web pages and Natural Language Processing (NLP) technique is then followed to identify the named entities. The semantic processing exploits our E-MOWL based ontology reasoning scheme to identify the context of the named entities. The input video is then tagged with text news, blogs and other videos on the basis of semantics of the content. News aggregation thus provides us with a complete description of news in the form of collection of documents comprising of videos and text news. It creates a multi-modal repository linked to the news event which can then be successfully integrated to form the contextually relevant documents shown in Figure 10.

We have compared the performance of E-MOWL based concept recognition approach with traditional SVM based classifier and context-based



**Figure 10** Results of OM based news aggregation and book recommendation.

tagging approach (CBTA) as given in (Pahal et al., 2015). The comparison results for some of the abstract entities showing how correctly these entities have been classified is summarized in Table 1. The complex event objects, such as recording session and musical expressions are described as temporal sequences in our approach. The event recognition performance in the ontology based approach is superior and scalable as compared to SVM and CBTA which limits in terms of implementation. Basically, the abstract concepts in the domain have no definitive features that makes designing of classifiers difficult. The proposed approach considers the shreds of evidence to infer the overall context thus providing significant improvement in retrieved results.

## 6.2 Multimedia Event Based Book Recommendation

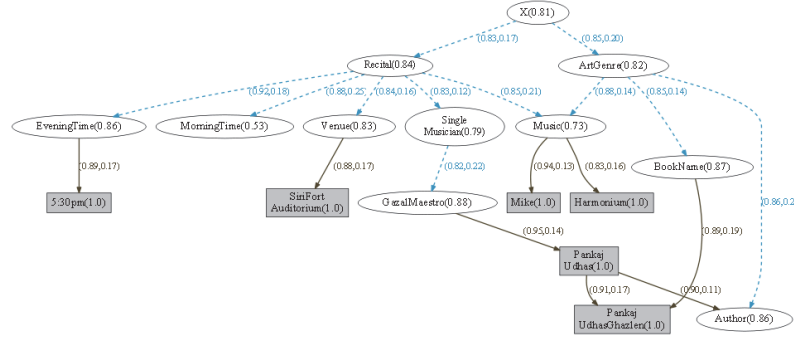
This section includes *Event-based Recommendation* where we focused on recommending books to the user based on the semantic meta-data and contextual information obtained from the input video. Besides the event related

**Table 1** Comparison of event detection results using SVM, CBTA and proposed approach

Abstract Events	SVM	CBTA	Proposed Approach
Recital	0.65	0.81	0.86
Music Ensemble	0.71	0.76	0.84
Recording Session	0.66	0.70	0.78
Release Status	0.69	0.75	0.83
Musical Expression	0.68	0.72	0.85

aspects the knowledge base contains concepts that are related to *Book Genre*. Based on this information the system then dynamically suggest books related to inferred event to enhance the knowledge thus defining a complete subject material. The probabilistic reasoning scheme of E-MOWL that reasons with the semantics of the multimedia content as well as with the metadata has been utilized for making book recommendations. Figure 10 also shows a part of the book ontology encoded in E-MOWL along with the concepts related to music concert. Books can be classified into different types which are further divided into genres such as drama, art, autobiographies etc. The different attributes like author, publication are associated with the book classes as their media manifestations and these textual features of book are related to different event aspects like location, agent etc. to facilitate recommendation. For instance, the concept *Artist* is linked to the concept *Book* through relationship *recBook*. Thus, the book named *PankajUdhasKiGhazlen*, a sub-concept of *Book* is the candidate for recommendation. The artist PankajUdhas is the author of this book and the relation between these instances has been captured using the relation *recBook*.

We created a database comprising of various genre of books and maintained a record of metadata corresponding to each book in a file. Since the amount of textual information available online lacks accurate or sufficient metadata so we manually annotated the book database to improve recommendations. To determine the recommendation, annotation of books to ascertain the particular genre in the OM is instantiated and their belief is propagated in BN. So, to achieve this, the obtained OM's for concepts *Recital* and *ArtGenre* from the ontology are attached to the *recommendation context X*. An OM for this scenario is shown in Figure 11. The book and recital attributes are represented in the OM as leaf nodes which leads to updated posterior probability in the BN. The posterior probability at the root node gives the recommendation measure and the book with a value greater than a certain threshold is a candidate for recommendation in the current context. This



**Figure 11** OM for recommended context  $X$ .

process is iterated for all the metadata stored corresponding to book database and the recommendation score is computed. Following this score, books are recommended for the context represented by the OM. Figure 10 also shows the books recommended to the user in identified context. It comprises of books that are related to the author/ agent involved in an event along with the ones that share the same event context.

## 7 Conclusion

In this paper, we have presented a new specification for representing events using Multimedia Web Ontology Language. Various conditions that enable an event to occur can be viewed as the context of an event which gets defined with respect to an event instance. The multimedia event ontology language (E-MOWL) allows representation of information related to various aspects that are associated with an event. This includes temporal, spatial and agent/entity related information. To identify the geographic context of an event we have utilized geospatial knowledge. Using this representation scheme the event relations in media objects are detected. We have developed an inference framework for detecting the highest-level event based on the spatio-temporal aspects, event aspects, relations and their geographic context present in multimedia document.

## References

- Blei, D. M. and Jordan, M. I. 2003. Modeling annotated data. *Proceedings of the 26th annual international ACM SIGIR conference on Research and development in informaion retrieval*, ACM, pp. 127–134.

- Chaudhury, S., Mallik, A. and Ghosh, H. 2015. *Multimedia Ontology: Representation and Applications*, CRC Press, 2015.
- Francois, A. R., Nevatia, R., Hobbs, J., Bolles, R. C. and Smith, J. R. 2005. Verl: an ontology framework for representing and annotating video events, *MultiMedia, IEEE* **12**(4):76–86.
- Ghosh, H. and Chaudhury, S. 2013. Ontology for semantic multimedia web.
- Liu, W., Liu, Z., Fu, J., Hu, R. and Zhong, Z. 2010. Extending owl for modeling event-oriented ontology, *Complex, Intelligent and Software Intensive Systems (CISIS), 2010 International Conference on*, IEEE, pp. 581–586.
- Neapolitan, R. E. 2012. *Probabilistic reasoning in expert systems: theory and algorithms*, CreateSpace Independent Publishing Platform.
- Pahal, N., Chaudhury, S. and Lall, B. 2013. Extending mowl for event representation (e-mowl), *Proceedings of the 2013 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT)-Volume 03*, IEEE Computer Society, pp. 171–174.
- Pahal, N., Chaudhury, S. and Lall, B. 2015. Context-based semantic tagging of multimedia data, *International Conference on Pattern Recognition and Machine Intelligence*, Springer, pp. 169–179.
- Papadias, D., Mamoulis, N. and Delis, V. 2001. Approximate spatio-temporal retrieval. *ACM Transactions on Information Systems (TOIS)* **19**(1):53–96.
- Patel-Schneider, P. F., Hayes, P., Horrocks, I. et al. 2004. Owl web ontology language semantics and abstract syntax, *W3C recommendation* **10**.
- Pongpaichet, S., Singh, V. K., Gao, M. and Jain, R. 2013. Eventshop: recognizing situations in web data streams, *Proceedings of the 22nd international conference on World Wide Web companion*, International World Wide Web Conferences Steering Committee, pp. 1359–1368.
- Scherp, A. and Mezaris, V. 2014. Survey on modeling and indexing events in multimedia, *Multimedia Tools and Applications* **70**(1):7–23.
- Shaw, R. and Larson, R. R. 2008. Event representation in temporal and geographic context, *Research and Advanced Technology for Digital Libraries*, Springer, pp. 415–418.
- Wattamwar, S. S. and Ghosh, H. 2008. Spatio-temporal query for multimedia databases, *Proceedings of the 2nd ACM workshop on Multimedia semantics*, ACM, pp. 48–55.
- Westermann, U. and Jain, R. 2006. E-a generic event model for event-centric multimedia data management in echronicle applications, *Data Engineering Workshops, 2006. Proceedings 22nd International Conference on*, IEEE, pp. x106–x106.



Westermann, U. and Jain, R. 2007. Toward a common event model for multimedia applications, *IEEE MultiMedia* **14**(1):19–29.

Wu, L. n.d. Representing and inferring events from deforestation observations.

## Biographies



**Nisha Pahal** received the B.E. degree in computer science and engineering from the Lingayas University, Faridabad, India, in 2005, and the M.Tech. degree in computer engineering from the YMCA University of Science and Technology, Faridabad, India in 2007. She received her Ph.D. degree from the Indian Institute of Technology Delhi, India in the year 2017 and has many publications to her credit in reputed international conferences. In due course of Ph.D. degree, she has been a part of the industrial project on Context-aware reasoning framework for Multi-user recommendations in Smart Home.

She is currently working as an Assistant Professor at Amity University Noida, India. Her current research interests include Multimedia Analysis, Ontology, Bayesian Network, NLP, and Machine learning.



**Brejesh Lall** (Member, IEEE) received the B.E. and the M.E. degrees in electronics and communication from Delhi College of Engineering, DU

Delhi, India, in 1991, and 1992, respectively. He completed Ph.D. degree, in 1997 from IIT Delhi in the area of multirate signal processing. During the Ph.D. he worked on “some studies on characterization and modeling of stochastic processes in the multiscale framework.”

He joined Hughes Software Systems, in 1997 and worked there for nearly eight years with the Signal Processing Group. He returned to his alma mater and joined IIT Delhi as a faculty member, in 2005. Since July 2005, he has been in the Electrical Engineering Department and has contributed to research and teaching in the general area of Signal Processing. He has successfully completed numerous sponsored projects and consultancies and is working on several others. He is the current head of Bharti School of Telcom Technology and Management, and the co-ordinator of two centers of excellence, viz. Airtel IIT Delhi Centre of Excellence in Telecommunications and Ericsson IIT Delhi 5G Center of Excellence.



**Santanu Chaudhury** received the B.Tech degree in electronics and electrical communication engineering, and the Ph.D. degree in computer science and engineering from the Indian Institute of Technology Kharagpur, Kharagpur, India, in 1984 and 1989, respectively.

He is a Professor with the Department of Electrical Engineering, Indian Institute of Technology Delhi, New Delhi, India, and is currently serving as the Director with the Indian Institute of Technology Jodhpur, Jodhpur, India. Recently, he completed his tenure as the Director with the Central Electronics Engineering Research Institute, Pilani, India. He has more than 300 research publications in peer reviewed journals and conference proceedings, 15 patents, and four authored/edited books to his credit. His research interests include image and video processing, computer vision, machine learning, and embedded systems.

Prof. Chaudhury was the recipient of the Distinguished Alumnus award from the Indian Institute of Technology Kharagpur. He is a Fellow of the Indian National Academy of Engineering, the National Academy of Sciences, and the International Association for Pattern Recognition. He was awarded the Indian National Science Academy Medal for Young Scientists in 1993. He was also the recipient of the Advanced Computing and Communications Society-Centre for Development and Advanced Computing (ACCS-CDAC) award for his research contributions in 2012.

