

The Nested Event Model for Multimedia Narratives

Ricardo Rios M. do Carmo
TeleMidia Labs
Pontifical Catholic University of Rio de Janeiro
Rio de Janeiro, RJ, Brazil
e-mail: rrios@telemidia.puc-rio.br

Luiz F. G. Soares and Marco Antonio Casanova
Informatics Department
Pontifical Catholic University of Rio de Janeiro
Rio de Janeiro, RJ, Brazil
e-mail: {lfgs, casanova}@inf.puc-rio.br

Abstract—The proliferation of multimedia narratives has contributed to what is known as the “crisis of choice”, which demands a much more active participation on the part of the user to consume multimedia content. To address this issue, a strategy is to offer users efficient search mechanisms, sometimes based on ontologies. However, one may argue that such mechanisms are often based on abstractions that do not adequately capture the essential aspects of multimedia narratives. This paper proposes a conceptual model to specify multimedia narratives that overcomes this limitation. The model is based on the notion of event and is therefore called Nested Event Model (NEMo). The paper also includes a complete example to illustrate the use of the model.

Keywords—component; semantic model; multimedia narrative; multimedia semantics; event model

I. INTRODUCTION

Narratives are considered one of the first forms of entertainment and one of the most important and ubiquitous aspects of human communication. Narrative is the specification of a logical ordering, possibly non-linear, of real or fictitious events to present a story [3]. Events are occurrences in time that have some essential associated properties. Event precise definition is deferred to Section III.A. A multimedia narrative is a narrative that can be presented through using multiple media types. Examples of multimedia narratives are interactive digital TV programs, electronic games, etc.

The growing number of distributed multimedia content has led users to devote more time to decide what will be consumed, and where, when, and how such content will be consumed. This situation, known as the “crisis of choice” [8], is particularly true for multimedia narratives. To reduce the cognitive overhead solutions point to the adoption of more sophisticated search mechanisms.

Following this approach, some proposals adopt ontology-based formalisms. However, some of them are domain-specific, or are based on models that consider only specific metadata for each type of media. In general, they only offer ways for describing each of their individual media content type, redundantly, without a formal semantics for representing the specific characteristics of multimedia narratives and their events as a whole, in a higher abstract level.

Some other proposals, despite offering mechanisms to annotate the narratives with (semi-)formal metadata, do not

consider events as first-class entity, which restricts the precise specification of multimedia narratives. They also do not consider the event semantics and the semantics embedded in causal, temporal, and spatial relationships among events.

The major contribution of this paper lies in defining a generic conceptual model for the specification of multimedia narratives based on their events, called Nested Event Model (NEMo). More precisely, contrasting with related work discussed in Section II, the proposed model considers events as first-class entities. It also allows for specifying the semantics associated with narrative’s events within the context in which these events occur, taking into account the relationships established among these events. By using NEMo, specialized mechanisms can be provided not only to locate, but also to adapt and offer alternative narratives that meet different profiles of presentation platforms and contexts with multiple users.

The rest of this paper is organized as follows. Section II discusses some related work. Section III raises the requirements that guided our proposal, firstly defining events and the advantages of using this concept as first class entity in modeling narratives. Section IV introduces NEMo. Section V presents a partial example that illustrates its use. Finally, Section VI presents our conclusions and directions for future work.

II. RELATED WORK

Hypermedia data models, such as AHM [13], MHEG [21], and NCM [27], use the concept of events and the temporal and spatial relationships among them to specify multimedia presentation. In these models, events are not first class entities and are always tied to media objects (presentation of media objects, selection of media object’s content, etc.). In addition, neither AHM, nor MHEG, nor NCM deals with the usual imprecision intrinsic to concepts media objects represent. Besides the structural, spatial and temporal aspects that may be associated to events, those models do not treat other aspects, like informational, causal, and experiential aspects.

The Video Event Representation Language (VERL) [10, 22], used in multimedia content analysis, defines formal methods to infer real world events from changes in the state of objects in a video stream. VERL uses an ontology for structuring the relationships between events recognized in a video stream. The main language focus is data indexing and

inference, based on metadata. VERL includes a vocabulary that allows for specifying the structural and causal aspects of events. The model does not consider the temporal, spatial, and informational aspects. To overcome this restriction, it is possible to import specialized ontologies and use them to model those aspects. The Video Event Markup Language (VEML) is the VERL markup language that allows for annotation on videos to capture their experiential aspect. However, the VEML metadata may lead to duplicated descriptions. Another drawback is that events are always defined as state changes in video stream objects, which precludes the description of other event types related with other types of media objects than stream video.

In [5], a model is introduced where events are the central information unit, identified by Universal Unique Identifiers (UUIDs). In the model, events describe the availability of multimedia content that: represents something important in a given domain; happens at or during a certain time; and occurs at a certain place. In other words, the temporal and spatial aspects of events together with their associated media objects form the backbone of the model. However, events are highly dependent on media objects. Events are treated as metadata of media objects, which leads to undesirable event specification duplications. For instance, an event represented by different images is repeatedly specified for each image. An alternative, adopted in the model proposed in this paper, would be treating events as first-class entities, independently from media objects that represent them. In this case, multiple media objects may be linked to a single event, offering different views of the event, each one suitable for a different presentation environment. Another drawback of the model proposed in [5] is that temporal and causal relationships between events are not explicit, which makes difficult the authoring of multimedia narratives. To overcome this limitation those relations could be specified through using RDF descriptions, but we would again face the problem of description duplications.

References [18] and [24] treat events as first-class entities and define models to respectively describe multimedia narratives and real world occurrences. The first reference specifies the ABC ontology, which was generated from four other ontologies: MPEG-7, MPEG-21, CIDOC/CRM¹, and another ontology coming from the biomedical domain. The second reference presents a formal model, the Event-Model-F (EMF), which is based on DOLCE+DnS Ultralite² (DUL) ontology.

As in [5], the ABC ontology allows for describing multimedia narratives using the concept of event. An event is defined as the transition between two situations [15, 18]. Its duration is obtained indirectly through relations with other entities and it is not an aspect, but a generalization of the concept of event. By contrast, in our model temporal aspect is treated as a contextual property of events. ABC treats the structural, causal, and experiential aspects similarly to our proposal. Finally, we observe that the ABC ontology was

based on museum and biomedical ontologies. It was not conceived as a generic approach to other domains.

The EMF model considers multimedia content as metadata of events. Although EMF does not restrict the type of the multimedia content, it does not specify how multimedia content can be used to document events. Another important point to highlight is that EMF is based on the DUL ontology, which makes it vulnerable to changes DUL may suffer. In this paper, we chose to define a model that is not dependent of any ontology, but only uses ontologies to enrich the specification and description of events and multimedia narratives.

The E model [28] also considers events as first-class entities. The class Activity groups events that satisfy the same description, while class Constellation is a set of events, not necessarily sharing the same description. This modeling approach provides a good characterization of the structural aspects and allows for the definition of composite event. It is also important to note that in E events have six aspects, discussed in Section III, similarly to what we propose. However, E defines events as instantaneous occurrences, which differs from the concept we adopted in our model. Another characteristic to highlight is that, in E, Activity and Constellation are not defined as events, but Occurrences. We argue that these are not sibling classes, but types of events, as is discussed in [11]. The E model is based on the DOLCE ontology, which makes it vulnerable to changes DOLCE may suffer.

The E* model [11] is a temporal data model for events. Like the E model, E* models events as first-class entities that has the six aspects. The main focus of E* is the temporal synchronization of real world events and it does not consider other types of relationships, such as informational ones (for example, the best moments of a soccer match, like goals, penalty kicks, fouls, ordered from the most impressive to the less one); causal ones (events ordered in a causal chain); experiential ones (events ordered according to the type of media objects that represent them; for example: first MP4 medias, then the JPG and HTML ones); or any other relationship pattern a multimedia narrative author, or a user profile, or a presentation environment may define. E* has the same problem of the E model, regarding its dependency of the DOLCE ontology. By contrast, we argued that events must be flexible enough to promote semantic interoperability, scalable assimilation of information from various sources, and harmonization of ontologies coming from various domains. The E*ML language expresses the constructs of E* and it is influenced by RDFS. Another feature that must be highlighted is that E* is a data model for storing information about events, and does not cover the presentation of events in multimedia environments.

DOLCE and DUL are upper ontologies, i.e., foundation ontologies that describes very general concepts that are common to all knowledge domains. They have the purpose of supporting broad semantic interoperability between those domains and are devoted to the treatment of social entities, such as organizations, collectives, plans, norms, and information objects (musical, text, word, picture, etc.), which has no further specializations, in order to meet the features of

¹ <http://cidoc-crm.org>

² <http://socop.oor.net/ontologies/1011>

each information object type. Both ontologies provide a classification of events. DUL classifies an event as being an action or a process. DOLCE offers a complex classification, specializing an event in three types (accomplishment, achievement, and cognitive-event), each one with its own specialization hierarchy. What is common in both classifications is that an event is either atomic or compound. For the purpose of this work, this common characteristic is used as the generic classification for events and it is left to authors of multimedia narratives the freedom of additionally classifying events with DOLCE, DUL, or any other description model.

In general, our model proposal addresses the limitations discussed in all other mentioned work considering the description of multimedia narratives. Contrasting with [12, 13, 27], our proposal considers events as first-class entities. This approach provides a better abstraction to specify multimedia narratives and uses aspects of events to describe their occurrences and contexts.

The proposed model covers real world events, as in [11] and [24], and additionally fictional events. Events have duration and they can be synchronized in circumstances not covered by [11]. Differently from [5, 12, 13, 27], we link media objects to events. This allows for associating events with media objects of several types and qualities, avoiding event description replications. Moreover, this allows for adapting multimedia narratives to users' profile and presentation environment. Unlike references [18] and [24], which use abstract entities defined in imported domain ontologies and multimedia description models, and, therefore, are affected by changes in such models, the model we propose does not depend on any other description model. This feature, however, does not prevent us to use external sources of information to enrich the description of multimedia narratives. By adopting an independent approach, the model we propose is not limited to any (external) event classification, as in the case of [11] and [28].

III. MULTIMEDIA NARRATIVE MODEL REQUIREMENTS

A. Event

Several research efforts have adopted events as an abstraction to better manage media content [16, 27]. The major motivation is that event provides a natural and basic abstraction to model real world processes [9]. In these efforts, including in ours, events are occurrences that have a measurable duration [23]. In addition, events have other properties, such as location, media objects that represent it, and other associated concepts. A soccer match, for example, is an event, with related conditions such as begin and end times, a location, a set of concepts that represent its agents (players, referees, coaches), etc. However, note that the match may also be considered a narrative (a composite event), that is, a sequence of events, such as fouls, offsides, penalties, etc.

Media object is used, in general, to register events of a narrative. A media object is a container that stores media content and its metadata. Several different media objects can register the same event.

In multimedia computation, the concept of event is amply adopted, but sometimes with different semantics. In models of event-oriented programming languages, for example, events are treated as instantaneous occurrences. This concept is used in video representation language, such as the VERL. Instantaneous events are also adopted in multimedia content description standards, such as MPEG-7 [19], MPEG-21 [17] and TV-Anytime [20], which offer the possibility of annotating media content with events.

In some multimedia systems, models and formats, events may have duration. Nonetheless, events are treated as second-class entities: they do not exist by themselves and are tied to media objects. Some hypermedia conceptual models, such as NCM [27] and AHM [13], go further and introduce the concept of composition of events, defined as the presentation of any combination of media content of a composite node (that groups media objects and internal composite nodes – thus grouping events – together with their relationships) [1]. Multimedia presentation formats, such as MHEG [21], NCL 3.0 [26] and SMIL 3.0 [6], use this concept to specify the dynamics of multimedia presentations.

Models in which events are attached to media objects proved to be inappropriate to describe narratives. The main reason lies in that media content is just another source of information about events. Other reasons are: the difficulty in classifying media content using tags, references, etc.; the redundancy introduced when several media content are associated with the same event; the lack of a user interface to manage events beyond media content; and scalability problems [25, 28]. Models that treat events as first-class entities offer an answer to these concerns. In these models, several different media objects, or several combinations of media objects can represent the same event. For example, a video exhibition, an audio narration, a text report print, etc. can define a same event: let's say, a soccer game.

In our proposal events are first-class entities that are part of a narrative. They can be atomic or composed by other events. Events have a time and a location, they are represented by media objects, and they are bound to other concepts, as exemplified in the next paragraph. Similar events can share common characteristics.

Events are associated to some related conditions that enable them to occur. For example, a soccer match depends on a set of conditions such as: an agreement between two soccer clubs regarding to the match date, the soccer stadium, the time the match must occur; two teams with 11 players each, referees, some spherical balls; etc. A soccer match can also be seen as a compound event, which groups and relates (sub-)events involving the players, the ball, and the referees, to cite a few. An event of this composition, for example a penalty kick, also requires conditions such as: a foul committed by a defending player; the referee blowing the whistle and pointing to the penalty mark; a penalty kick taker in the penalty mark; and a goalkeeper preparing to make a save from the penalty kick, etc.

A set of related conditions that enables an event to occur can be seen as the context of an event. If at least one condition is not satisfied, an event is not characterized. For example, in the case of a penalty kick event, if a referee does

not blow the whistle, or if there is no soccer ball, or there is no soccer field, the penalty kick is not characterized. Another important characteristic of events is that they may only make sense in a narrative (compound event). For example, a penalty kick does not make sense in a narrative of a basketball match, since some necessary related conditions, such as a goal and a goalkeeper preparing to make a save, are not satisfied.

The related conditions that form the context of an event can be seen as contextual information and they usually fit in six elementary aspects: informational, which is partially context-dependent; and temporal, spatial, causal, experiential, and structural, which are fully context-dependent.

There are two types of informational aspects. One type is about the information related to the description of media objects associated with an event. The media object description is technical metadata that record the media object technical attributes, such as production or creation information, capturing process, file formats, resolutions, color profiles, etc. The other type of informational aspect is related to the semantic description of the event (atomic or compound) itself. Both types of informational aspects are useful for search engines to retrieve narrative documents.

An event has at least two important context-independent description attributes (a name and universal identifier) and a set of context-dependent description attributes (concepts), which are based on domain-specific models. For example, a soccer match has a name and a universal identification, and a set of soccer-domain specific information, such as the number of players involved, referees, the rules of the game, etc. In the case of domain specific information, specialized knowledge models, like ontologies, can be used to structure the relations between the concepts of the domain. This flexibility raises several difficulties since domain-specific models may involve mutually inconsistent concepts, or concepts at different levels of abstraction [30]. For example, if mutually inconsistent concepts, each one from two or more independent soccer ontologies, are used to describe a penalty kick, no trivial inference about those concepts is feasible.

The temporal aspect allows for specifying temporal synchronization among events. Synchronization can be established referring to an absolute point in time, or by causal or constraint relationships between events. In the soccer example, temporal synchronizations are defined by the start and end times of the match.

Likewise, the spatial aspect permits specifying spatial synchronization, absolutely or relatively, between events. This aspect has a broader scope than that of SMIL 3.0 and NCL 3.0. For example, such languages just specify the location where media must be exhibited, while in the proposed model it allows for specifying the place where a real world event occurs, besides the location where each media object that represents the event can be exhibited. The stadium geographical location, the dimensions of the rectangular field, and the location of the players in the field are some examples of spatial aspects supported by NEMO model.

In addition to temporal and spatial relationships, the causal aspect covers other causal relationships between events. For example, a foul committed by a defending player may cause the loss of the championship.

The experiential aspect permits representing events using alternative media objects of different types. This allows for defining different representations for the same event, each one adequate to a specific presentation environment and/or user profiles. The experiential aspect also helps to deal with the semantic gap [14] binding events and media objects together.

Lastly, the structural aspect allows for grouping events and relationships among them into a logic structure. This aspect is similar to the concept of composition in NCM. It allows for specifying an event as a composition of other related child (sub-)events, as is the case of the relationship between a soccer match (event) and a penalty kick (sub-event), aforementioned.

Composition of events is also useful for defining interpretations of a given narrative. A narrative interpretation is an alternative narrative that have adapted some part of the original narrative in order to satisfy a user profile, a presentation environment profile, or other restrictions. Returning to our soccer-match example, such alternative narratives may be specified for a user who wants to watch first the best moments of a soccer match (events of goals, fouls, penalties, etc.). In this case, only a subset of the match's events will be presented. In addition to the possible adaptation of events (and their media objects), relationships between those events may be altered. Alternative narratives can be defined during specification or consumption phases.

In summary, in the proposed model, an event E is a contextualized first-class entity that occurs during certain moment in time. Its occurrence is determined by some cause, information, or by spatial-temporal relations. It may be associated with one or more media objects M_1, \dots, M_k , and not vice-versa. Therefore, besides helping dealing with the semantic gap problem, redundancy is reduced since the description of E need not be represented in each media object M_i . In addition, in order to help dealing with the 'crisis of choice' problem, the model aims to facilitate the task of describing events and narratives at various degrees of abstraction. Furthermore, the model allows for defining alternative narratives for the same set of events.

B. Raised Requirements

The definitions in Section III.A may be rephrased as a set of requirements that a model for the description and specification of multimedia narratives must meet:

- Multimedia narrative specifications must be based on the notion of event. Events and their relationships specify a narrative, a compound event.
- Events are first-class entities; they are independent from media objects. Media objects are second-class entities not required during the specification phase.
- One or more media or composite objects can be associated to an event (or a composite event). At least one media or one composite object must be associated to an event so it can be presented.

- Event can have context-dependent elementary aspects that must be captured: structural, causal, temporal, spatial, and experiential; and partially context-independent informational aspects. At least one aspect of an event must be characterized. The informational one seems to be the most appropriate due to its partially context-independent characteristic.
- Events can be associated to one or more abstract entities of different domains. This requirement describes the concepts involved in the event. Knowledge bases can be used to enhance the description of concepts.

Besides the requirements coming from the event definition, the following requirements are added:

- The model must be generic so as to cover a wide spectrum of domains. There are only two basic types of event: atomic and compound. All other events must be specializations of them. The coverage of a wide spectrum of domains, with their particular requirements on event types, must be achieved through the composition of events, or by the use of a URI classifier, which helps classifying events into a semantic knowledge domain.
- The model must be formally defined and equipped with a sound and complete inference procedure. This requirement is essential for narrative generation tools to automatically generate alternative narratives and alternative multimedia presentations.
- Each entity in the model must be univocally identified so that they can be unambiguously referenced and to facilitate entity management.
- An event can be associated to provenance metadata, i.e., metadata that refers to the crucial information about its sources, especially those expressing uncertainty, imprecision of the data sources, and who/what created them. This is important to verify whether the event comes from a trusted source, to give credits to event creators, etc.

IV. NEMO – NESTED EVENT MODEL

Considering the requirements of Section III.B, we can introduce the Nested Event Model (NEMo) to support the specification, description, personalization, and retrieval of multimedia narratives. Although the formal specification of the model is a requirement stated in Section III.B, only a semi-formal specification is presented in this paper.

Fig. 1 shows the NEMo classes. As usual, instead of “an instance of class C ”, we use the abbreviated form “a c ”.

The Entity root-class is represented by the expression $\square = [id, n, u, c, cdt]$, in which id (*identifier*), n (*name*), u (*uncertainty*), c (*creator*) and cdt (*creationDateTime*) are context-independent properties. Properties id and $name$ associate a unique universal identifier (UUDI) and name to an entity, respectively. Properties u , c , and cdt are model-specific. The *creator* property indicates the author of the entity (which can be a human, software, or a hardware agent), identified by a URI (Uniform Resource Identifier).

The *creationDateTime* property indicates the time the entity was created, which can be before, during or after the occurrence of the event. The *uncertainty* property registers the imprecision of the information extracted from multiple sources, taken values from the interval $[0, 1]$. For entities of the *Event* class, this property indicates if it is a real world (zero) or fictional (one) event. For entities of the *Concept* class, this property indicates the distinct levels of abstraction or divergent concepts that database schemas and ontologies may have.

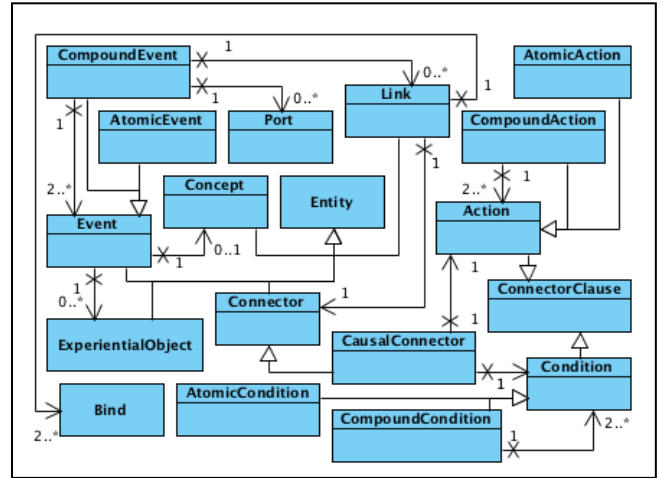


Figure 1. NEMo overview.

The classes *Event* (E), *Concept* (C), *ExperientialObject* ($ExpO$), *Connector* (Ct), and *Link* (L) are specializations of the *Entity* class (Fig. 1):

Event, represented by the expression $E = [loc, dur, expo_s, cpt, desc]$, is an entity that has five additional properties: location (loc), duration (dur), a set of experiential objects ($expo_s$), a concept (cpt), and an additional description ($desc$). The loc and dur properties indicate the location and the beginning and ending times of the event, respectively. It is important to note that values of these properties are references to RDF or RDFS/OWL documents that describe the location and time of the event through the use of a formal semantic model. The property $expo_s$ is a set of experiential objects that represents the event. This property is defined by the expression $ExpO_s = [(ExpO_1, \dots, ExpO_n)]$, where $ExpO_n$ is an instance of the class *ExperientialObject*. The cpt indicates the concept that defines the event type, for example, if it is a goal or penalty event. The $desc$ is an optional property that, like loc and dur , uses URIs to refer to RDF or RDFS/OWL documents for enhancing the description of an event.

Contrasting with the classification presented in [28] and [24], since NEMo is generic, the *Event* class is specialized into just two subclasses, *AtomicEvent* (E_A), and *CompoundEvent* (E_C).

AtomicEvent is an event that cannot be decomposed into sub-events. An atomic event is characterized by the degree of detail an application requires. For instance, a penalty kick, i.e., a severe foul in a soccer match, may be considered an atomic event at some level of abstraction or it may be

decomposed into sub-events (kick preparation, kick, goalkeeper defense or goal) in another.

CompoundEvent is an event with three additional properties: a set of events (E_S), a set of ports (P_S), and a set of links (L_S). It is represented by the expression $E_C = [loc, dur, cpt, expo_S, desc, L_S, E_S, P_S]$. E_S is represented by the expression $E_S = [e_1, \dots, e_n]$, where e_n is an atomic or compound event; P_S is an optional property represented by the expression $P_S = [p_1, \dots, p_n]$, where p_n represents a port instance; and L_S is represented by $L_S = [l_1, \dots, l_n]$, where l_n represents a link instance that binds child events.

A compound event must have at least two events. A soccer match, for example, is a compound event that groups two sub-events (first and second periods), which in turn are compound events that group other sub-events (fouls, penalties, goals, kicks, etc.).

A *Port* (P) is a second-class entity represented by the expression $P = [n, e]$, where n is the name of the port and e indicates a sub-event of a compound event. A port allows defining mappings between a compound event and its sub-events. As a result, a sub-event may be made visible for external references.

Inside a compound event, child events can be related to other events through a set of relationships (links) represented by L_S . A link specifies a directional relationship between two or more events [27] with the help of two additional classes: *Connector* and *Bind*. A link refers to a *Connector* and defines a set of binds (instances of the class *Bind*), which associate each link endpoint to a clause of the used connector. *Link* is represented by the expression $L = [Ct, B_S]$, where B_S is a set containing at least two binds and Ct represents a *Connector*.

A bind connects an event, or a port of a compound, to one clause of the referenced connector [27]. A set of binds, represented by the expression $B_S = [b_1, \dots, b_n]$, where b_n is a bind, defines the connected parts. A *Bind* is represented by the expression $B = [e, r, p]$, where e defines the NEMO event that participates in a relationship, r indicates the role the event assumes in the relationship by referencing a connector role, and p is an optional property used for referencing a port defined in a compound event.

A *Connector* is a NEMO entity that has similar behavior to the hypermedia connectors defined in [27]. The semantics of a connector is defined through a set of connector clauses, which define roles played by the participants of a relationship. A role must be unique in the connector clause set and it is used by links to associate its participants to a connector clause. The role values are references (URIs) for descriptions, in RDF or RDFS/OWL, of entities of a knowledge domain. A connector is represented by the expression $Ct = [cc_1, \dots, cc_n]$, where cc_n is an instance of the class *ConnectorClause*. In the current version, NEMO defines one type of connector, a *CausalConnector*, and two types of connector clauses: *Condition* and *Action*. A *CausalConnector* is represented by the expression $CCt = [cond, act]$, where $cond$ is an instance of the class *Condition* and act is an instance of the class *Action*.

Condition defines an expression that, when satisfied, triggers an action expression (*Action*). A condition may be

atomic (*AtomicCondition*) or compound (*CompoundCondition*). An atomic condition is a connector clause represented by the expression $Cond_A = [r, min, max, qualifier]$, where r is the role played by a participant of a relationship. Properties *min* and *max* specify the minimal and maximal number of participants that may play the role, respectively. If *max* value is not specified, the 0 value may be used for specifying that the number of participants is unbounded. A *CompoundCondition*, represented by the expression $Cond_C = [(cond_1, \dots, cond_n), qualifier]$, where $cond_n$ is an instance of the class *Condition*, groups two or more conditions (atomic and compound), each condition with a different role. The property *qualifier* of the class *Condition* informs the logical relationship among the participants of a relationship, i.e., among the several binds that play the same role (*AtomicCondition*) or different roles (*CompoundCondition*). The possible values for the *qualifier* property are: "OR" and "AND". If the qualifier establishes an "OR" logical operator, the link action will be triggered whenever any condition occurs. If the qualifier establishes an "AND" logical operator, the link action will be triggered whenever all the conditions occur.

Action defines an expression that is triggered when a condition is satisfied. An action may be atomic (*AtomicAction*) or compound (*CompoundAction*). An atomic action is a connector clause represented by the expression $Act_A = [r, min, max, qualifier]$, where r is the role played by a participant of a relationship. The *min* and *max* properties specify the minimal and maximal number of participants that may play the role, respectively. If *max* value is not specified, the 0 value may be used for specifying that the number of participants is unbounded. A *CompoundAction* groups two or more actions (atomic and compound). A *CompoundAction* is represented by the expression $Act_C = [(act_1, \dots, act_n), qualifier]$, where act_n is an instance of the class *Action*. The property *qualifier* informs the logical relationship among the participants of the relationship, i.e., among the several binds that play the same role (*AtomicCondition*) or different roles (*CompoundCondition*). The possible values for the *qualifier* property are: "SEQ" or "PAR". Parallel ("PAR") and sequential ("SEQ") compound actions specify that the triggering of actions shall be performed in any order or in a specific order, respectively.

As an example, a link may be used for defining the relationship between three events: a penalty kick ($e_A: PenaltyKick01$), and two of advertisements about soccer shoes ($e_A: Adv03$ and $e_A: Adv04$). Two atomic connector clauses of a causal connector $CCt: Conn01$ may be defined: one with properties $min = 1$, $max = 1$, and $role = http://localhost/soccer/ontology/PenaltyKickTeam01$; and another one with properties $min = 1$, $max = 2$, $role = http://localhost/adv/ontology/SoccerShoes02$, and $qualifier = SEQ$ (see Fig. 2). The role values are references to RDF or RDFS/OWL descriptions of entities of a soccer knowledge domain. The $e_A: PenaltyKick01$ role describes a penalty kick any player performs, and the $e_A: Adv03$ and $e_A: Adv04$ roles describe shoes used in soccer games. NEMO does not restrict that both roles must come from the same knowledge domain. As the descriptions may be in RDF or RDFS/OWL, it is also

possible to use complex role types. The property *qualifier* defines that the advertisements shall occur in a sequence.

Concepts and experiential objects help dealing with the semantic gap. NEMo allows events to be conceptually described by an entity instance of the class *Concept*. An event e , for instance, may have the concept ‘Penalty Kick’, giving e the semantics of an event occurring in the soccer domain. This behavior is achieved using the property *cpt* of e (Figure 1). Examples of concepts in the soccer domain include penalty kicks, fouls, and etc. A concept is represented by the expression $C = [ecid, sch]$, where *ecid* (*externalConceptID*) is the id of a domain-specific abstract entity in the information source and *sch* (*schema*) is the information source schema. A concept uses RDF or RDFS/OWL entities as values for the property *ecid*. The schema property, which is referenced through URIs, allows for using schema processors to manipulate the information behind the concept. This feature also adds flexibility to the model allowing distinct schema processors for each concept in use. As in [28], the schema property may be used by schema processors to search for concept type declarations and to establish a namespace mechanism.

Instances of the class *ExperientialObject* are used to record the occurrence of events. For instance, JPEG, MPEG, GIF, etc. media objects can be used to register a visual occurrence of an event. NEMo restricts neither the number, nor the type of media object that can be used to register an event, but requires a specific and different experiential object to represent each media object. Experiential objects of an event are defined by the property *expos* of the class *Event*. An experiential object is represented by the expression $ExpO = [t, desc, loc, techDesc]$, where t is the type of the experiential object (MIME type), *desc* is the reference to a semantic description (in RDF or RDFS/OWL) of the media object, *loc* is the location where the media object can be found, and *techDesc* is any technical description about the object (production or creation information, capturing process, file formats, resolutions, color profiles, etc.).

V. NESTED EVENT LANGUAGE

Nested Event Language (NEL) is an RDFS/OWL-based language created from NEMo and for the specification of multimedia narratives. For brevity, we only illustrate how to specify an atomic event of a penalty kick; how to connect this event to other events through a link; and how to define the semantics of a link using a connector. For a complete example and more detailed explanation, including how experiential objects, concepts and other elements of a soccer match narrative can be related, refer to <http://www.telemidia.puc-rio.br/~rrios/ISM/2013/>.

According to the soccer rules, a penalty kick is a punishment given by the referee to a team that commits a foul inside its penal area. Three events must be described: the foul, the yellow card given by the referee, and the penalty kick. The occurrence of these proposed events may be seen as a partial view of the soccer match. A partial NEL code of the event penalty kick can be seen in Fig. 2. It is also shown the link that defines the relationship between the events and the connector that defines the semantics of the link.

```

01 <NEMo:AtomicEvent rdf:ID="PenaltyKick01">
02   <NEMo:uuid rdf:datatype="&xsd:anyURI">
03     urn:uuid:9A19EA6C-B57A-4411-818D-AE606C70538E
04   </NEMo:uuid>
05   <NEMo:concept rdf:resource="#PenaltyKick"/>
06   <NEMo:experientialObject rdf:resource="#Photo02"/>
07   <NEMo:experientialObject rdf:resource="#Audio02"/>
08 </NEMo:AtomicEvent>

09 <NEMo:Link rdf:ID="Link01">
10   <NEMo:bind rdf:resource="#Bind0101"/>
11   <NEMo:bind rdf:resource="#Bind0102"/>
12   <NEMo:bind rdf:resource="#Bind0103"/>
13   <NEMo:connector rdf:resource="#CausalConnector01"/>
14 </NEMo:Link>

15 <NEMo:CausalConnector rdf:ID="CausalConnector01">
16   <NEMo:condition rdf:resource="#AtomicCondition0101"/>
17   <NEMo:action rdf:resource="#CompoundAction01"/>
18 </NEMo:CausalConnector>

```

Figure 2. NEL sample code.

PenaltyKick01 is an event that has a concept (line 05) and experiential objects (lines 06-07). *Link01* (line 09-14) defines a causal link between the three events through properties *NEMo:bind*, which defines the binds (lines 10-12), and *NEMo:connector*, which defines the causal connector (line 13). Each bind defines the events that participate in the relation. *CausalConnector01* (lines 15-18) is the connector that defines the semantics of *Link01*. Property *NEMo:condition* defines the condition of the causal connector and property *NEMo:action* the action.

VI. CONCLUDING REMARKS

In this paper, we have defined the NEMo model, which features events as first-class entities to describe narratives. In particular, the concepts of *Link* and *Connector* express reference and synchronization relationships between events. A *CompoundEvent* expresses structural relationship and allows defining a narrative as a composition of events. We also presented a partial overview of NEL, the language used to describe NEMo narratives.

As discussed in the paper, by shifting the focus of the model from media objects to events, we move from (low-level) media streams to (high-level) narratives as descriptions of the real-world processes. This shift leads to a number of benefits, of which we highlight four. First, we reduce redundancy: media objects need not carry the same (redundant) metadata; the semantics is moved entirely to the events. Second, we gain in clarity since a user interface based on NEMo will reflect higher-level concepts, closer to the user’s experience (of the real-world processes). Third, the various criteria for event composition allow defining alternative narratives over the same set of events, thereby supporting narrative personalization. The personalization, not treated in the paper, is achieved through the use of profiles that specify the entities supported by users and by environment presentation. Lastly, the model is domain independent and expressive enough to capture the various aspects of events at different levels of abstraction; yet, it is extensible to facilitate meeting the needs of applications from different domains.

Ongoing work focuses on the development of the NEMO Authoring Tool (NEATool), based on a framework whose fixed-points are the NEMO classes and whose flexibilization points allow the specialization of the basic NEMO classes – *Event*, *Concept* and *ExperientialObject* – to account for the specificities of an application domain. The tool features a user interface that facilitates the generation of a narrative by annotating video with events; each event has the properties and structure defined in NEMO, and a set of new, domain-dependent properties. Other important features are: a converter that transforms a NEMO narrative into a synchronization language, like SMIL, NCL, etc.; a textual editor allowing an author to take a greater control of the narrative; and an inference machine to validate narratives.

Soccer matches and TV series (or soap operas) are the first target application domains. As a complete definition of a narration is too long to fit in the paper, Section V briefly illustrates a partial definition of a soccer match narrative using NEMO constructs. The specialization of the tool just reflects the major events in a soccer match: kicks, penalties, fouls, etc. The video annotator reviews the video and indicates the major events he wants to call attention to. The tool will then fill the event properties as automatically as possible and asks the annotator the values of the domain-specific properties (what player committed the foul, for example).

The specialization to TV series (or soap operas) is different in so far as the episodes have a script, which the annotator may resort to link events to the video shots. If he does not have access to the episodes' scripts, he will have to define the major events he wants to call attention to, just as for soccer matches.

NEATool is complemented with a Video Retrieval Tool (ViRTool), which offers a query language based on a fragment of Temporal Logic. The language is able to express queries such as “Get all videos where player X committed a foul on player Y in the first half of the match followed by a goal of player Z”, or “Get all scenes where actor X quarreled with actor Y and returned to actor Z”.

Future work will cover a complete formalization of the model and a query language based on a more expressive fragment of Temporal Logic.

ACKNOWLEDGMENT

This work was partly supported by CTIC/RNP, by CAPES, by CNPq (under grants 301497/2006-0 and 557128/2009-9), by FAPERJ (under grant E-26/103.070/2011), and by FAPEAM/SECT/Governo do Amazonas (under grant 020/2009).

REFERENCES

- [1] ABNT 2012. Digital Terrestrial TV – Data Coding and transmission specification for digital broadcasting. Technical Report #ABNTNBR:15606-2.
- [2] Aroyo, L. et al. 2009. NoTube – Making TV a Medium for Personalized Interaction. *Euro iTv 2009*. (Apr. 2009), 22–26.
- [3] Bal, M. 2009. *Narratology*. University of Toronto Press Incorporated.
- [4] Bellekens, P. et al. 2009. iFanz: A ubiquitous approach towards a personalized EPG. (Leuven, Belgium, 2009), 130–131.
- [5] Boll, S. and Westermann, U. 2003. MediEther — an Event Space for Context-Aware Multimedia Experiences. (Berkeley, California, 2003), 21–30.
- [6] Bulterman, D.C.A. et al. 2008. Synchronized Multimedia Integration Language (SMIL 3.0).
- [7] Cattelan, R.G. et al. 2008. Watch-and-comment as a paradigm toward ubiquitous interactive video editing. *ACM Transactions on Multimedia Computing, Communications and Applications*. 4, 4 (Nov. 2008), 28:1–28:24.
- [8] Donoso, V. et al. eds. 2009. Networked Television Adjunct proceedings of EuroITV 2009.
- [9] Events: 2002. <http://plato.stanford.edu/entries/events/>. Accessed: 2012-04-09.
- [10] Francois, A.R.J. et al. 2005. VERL: An Ontology Framework for Representing and Annotating Video Events. *IEEE Multimedia*. 12, 4 (Oct. 2005).
- [11] Gupta, A. and Jain, R. 2011. *Managing Event Information*. Morgan & Claypool Publishers.
- [12] Gutta, S. et al. 2000. TV Content Recommender System. (2000), 1–2.
- [13] Hardman, L. et al. 1993. The Amsterdam Hypermedia Model: extending hypertext to support real multimedia. *Hypermedia*. 5, (Jan. 1993), 47–69.
- [14] Hare, J.S. et al. 2006. Bridging the Semantic Gap in Multimedia Information Retrieval. (Budva, Montenegro, Jun. 2006), 1–15.
- [15] Hunter, J. 2003. Enhancing the Semantic Interoperability of Multimedia Through a Core Ontology. *IEEE Transactions on Circuits and Systems for Video Technology*. 13, (Jan. 2003), 49–58.
- [16] Jain, R. 2008. EventWeb: Developing a Human-Centered Computing System. *Computer*. 41, 2 (Feb. 2008), 42–50.
- [17] Karpouzis, K. et al. 2005. MPEG-21 digital items to support integration of heterogeneous multimedia content. *Science Direct*. (2005).
- [18] Lagoze, C. and Hunter, J. 2001. The ABC Ontology and Model. (Tokyo, Japan, 2001), 160–176.
- [19] Martínez, J.M. 2002. MPEG-7. *IEEE Multimedia*. 9, 3 (Jul. 2002), 83–93.
- [20] McParland, A. 2002. *TV-Anytime - using all that extra data*. Technical Report #WHP 050. British Broadcasting Corporation.
- [21] Meyer-Boudnik, T. and Effelsberg, W. 1995. MHEG Explained. *Multimedia*. (Mar. 1995), 1–13.
- [22] Nevatia, R. et al. 2004. An Ontology for Video Event Representation. (Washington, DC, USA, 2004), 119–129.
- [23] Pérez-Luque, M.J. and Little, T.D.C. 1996. A Temporal Reference Framework for Multimedia Synchronization. *IEEE Journal on Selected Areas in Communications*. 14, 1 (Jan. 1996), 36–51.
- [24] Scherp, A. et al. 2010. A core ontology on events for representing occurrences in the real world. *Multimed Tools Appl (2012)*. 58, 2 (Dec. 2010), 293–331.
- [25] Scherp, A. et al. 2008. Event-centric Media Management. (2008).
- [26] Soares, L.F.G. and Rodrigues, R.F. 2005. *Nested Context Model 3.0*. Technical Report #26/05. Pontificia Universidade Católica do Rio de Janeiro.
- [27] Soares, L.F.G. and Rodrigues, R.F. 2005. *Nested Context Model 3.0*. Technical Report #18/05. Pontificia Universidade Católica do Rio de Janeiro.
- [28] Westermann, U. and Jain, R. 2006. E - A Generic Event Model for Event-Centric Multimedia Data Management in eChronicle Applications. (2006), x106–x116.
- [29] Westermann, U. and Jain, R. 2007. Toward a Common Event Model for Multimedia Applications. *IEEE Multimedia*. 13, 1 (Jan. 2007), 19–29.
- [30] Shvaiko, P., Euzenat, J.: Ontology Matching: State of the Art and Future Challenges. *IEEE Transactions on Knowledge and Data Engineering*. 25, 1, 158–176 (2013).

