Pre-proceedings of the International Workshop on

Recognising and Tracking Events on the Web and in Real Life

EVENTS@SETN 2010

Held as a satellite workshop of the
Sixth Hellenic Conference on Artificial Intelligence (SETN 2010)

Athens, 4 May 2010

Preface

Users and organizations collect data in various structured and unstructured digital formats, but they cannot fully utilize these data to support content and resource management process. It is evident that the analysis and interpretation of the available data needs to be automated, in order for large data volumes to be transformed into operational knowledge. Events are particularly important pieces of knowledge, as they represent activities of special significance both for users and organisations. Therefore, the recognition of events is of outmost importance. Consider, for example, automatic event (e.g. emergencies) and trend detection by analysing users contributions to social Web 2.0 applications, the recognition of attacks on nodes of a computer network given the exchanged TCP/IP messages, the recognition of suspicious trader behaviour given the transactions in a financial market, and the recognition of various types of cardiac arrhythmia given electrocardiographs. The current proceedings include papers focusing on various aspects of event recognition, including analysis of video, audio, text and other sensor data, as well as recognition on fused data sources and temporal reasoning systems.

Athens, May 2010

Thomas Winkler
Alexander Artikis
Yiannis Kompatsiaris
Phivos Milonas

Workshop Organisers

Thomas Winkler	Fraunhofer IAIS, Germany
Alexander Artikis	NCSR "Demokritos", Greece
Yiannis Kompatsiaris	CERTH-ITI, Greece
Phivos Mylonas	NTUA, Greece

Programme Committee

Christian Bauckhage	University of Bonn (B-IT), Germany
Sotiris Diplaris	CERTH-ITI, Greece
Fausto Giunchiglia	University of Trento, Italy
Vana Kalogeraki	Athens University of Economics and Business, Greece
Kostas Karpouzis	NTUA, Greece
Lyndon Kennedy	Yahoo! Research, USA
Jobst Löffler	Fraunhofer IAIS, Germany
David Luckham	Stanford University, USA
Vassilis Mezaris	CERTH-ITI, Greece
George Paliouras	NCSR "DEMOKRITOS", Greece
Themis Palpanas	University of Trento, Italy
Ansgar Scherp	University of Koblenz-Landau, Germany
Kostas Stathis	Royal Holloway, University of London, UK
Nenad Stojanovic	University of Karlsruhe, Germany
Athena Vakali	Aristotle University of Thessaloniki, Greece

Contents

Historical Event-based Access to Museum Collections

Chiel van den Akker[1], Lora Aroyo[2], Agata Cybulska[1], Marieke van Erp[2], Peter Gorgels[3], Laura Hollink[4], Cathy Jager[3], Susan Legêne[1], Lourens van der Meij[2], Johan Oomen[5], Jacco van Ossenbruggen[2,6], Guus Schreiber[2], Roxane Segers[2], Piek Vossen[1] and Bob Wielinga[2,7]

[1] Faculty of Arts/VU University Amsterdam
[2] Department of Computer Science/VU University Amsterdam
[3] Rijksmuseum Amsterdam
[4] Web Information Systems Group/Delft University of Technology
[5] Netherlands Institute for Sound and Vision
[6] Centrum Wiskunde & Informatica (CWI), Amsterdam
[7] Human-Computer Studies Laboratory/University of Amsterdam
{C.vandenAkker,AK.Cybulska,S.Legene,P.Vossen}@let.vu.nl
{L.M.Aroyo,Marieke,Lourens,Schreiber,RH.Segers}@cs.vu.nl
{P.Gorgels,C.Jager}@rijksmuseum.nl
L.Hollink@tudelft.nl
joomen@beeldengeluid.nl
Jacco.van.Ossenbruggen@cwi.nl
B.J.Wielinga@uva.nl.

Abstract. This paper presents research in the context of two multidisciplinary projects aimed at facilitating the history domain with an automatic approach for event extraction and modelling. To realise this, the Semantics of History project is providing a historical ontology and a lexicon to support the detection of historical events in textual data whilst the Agora project focusses on exploring the modelling aspects of historical events and employing the combined results in an event-driven browse and search approach. Furthermore, the historical events are used as a flexible model to identify semantically relevant relationships between objects in highly diverse museum collections, creating meaningful 'cause' and 'effect' links along the key event dimensions 'who', 'what', 'where' and 'when'. This should finally support the (re)interpretation process of history research, by allowing end-users to create their own personal narratives, leading to theoretical reflection on the meaning of digitally mediated public history in contemporary society. In this paper, we give a high-level overview of the research challenges in the realisation of a desired search and browse scenario. Finally, we outline the open issues and future research.

1 Introduction

There is a vast amount of historical knowledge locked in museum collections. This knowledge is often explicitly present in textual descriptions accompanying museum objects or implicitly present through the fact that an object belongs to a particular collection and was collected for a particular purpose. In this sense, objects from one museum collection only tell part of the story, as they present a view on the past from one only perspective, limited to their collection. Through combining objects from different collections, a more comprehensive view of a certain historical period can be given. When unlocked, this knowledge can help casual users understand the significance of museum objects

and historical events better and aid experts (curators, art historians and historians) in their search for objects relevant to a topic.

The Agora project started in October 2009 with the aim of facilitating context-driven browsing and search in heterogeneous museum collections. The context that unites these collections is provided by historical events that can be linked to the collection objects, as historical event-descriptions are comprised of causal language, locations, the actors involved and the time of the event. Agora is a four year project with a team made up of experts from the computer science and the history departments at the VU University in Amsterdam[8,9], as well as from the Netherlands Instititute for Sound and Vision (henceforth: S&V)[10] and Rijksmuseum Amsterdam (henceforth: RMA)[11]. The goal of Agora is threefold: (1) a historical event thesaurus linked to museum artefacts, (2) a semi-automatic event modelling approach that satisfies both the needs of experts and the general public, and (3) an online social platform in which both the general public and expert historians can explore various perspectives on events, build their own narratives, and contribute to the evolution of the event thesaurus.

The Semantics of History project started in August 2009 with the aim to model changes of historical reality over time and through different writer perspectives which are revealed in historical text archives. For this purpose, Semantics of History aims at developing a historical ontology and a lexicon which will facilitate detection of historical events in textual data. The resulting models and the event extraction from text will be implemented and tested in the context of the Agora project. Semantics of History is funded by the VUA Interfaculty Research Institute CAMeRA and is carried out in cooperation between two departments at the VU University in Amsterdam: the linguistics department and the computer science department.

The results of both Agora and Semantics of History will be deployed in a social cultural heritage platform that will allow different users (e.g., experts, interested laypersons and secondary school students) to have event-based access to the S&V and RMA collections. As Agora and Semantics of History are tightly interwoven, we will from this point on report on the combined results of these projects.

This paper is structured as follows. In Section 2, the motivation for this project is given by discussing the shortcomings of current collection access and modelling and by describing the needs from different user groups in two use-cases. In Section 3, the technical challenges of Agora and Semantics of History are presented, along with the approaches that we are investigating. To conclude, we will present open issues in Section 4. In Figure 1, the different parties, goals and domains that play a role in the two projects are summarised.

2 Motivation

In the humanities domain, there are different user groups with a great need for advanced cross-collection access to resources. In our case, we are dealing with the question of how museums can present the specific information that belongs to objects in their collection in a way that strengthens users' historical understanding and involvement in relevant historical debates. We are also asking ourselves how we can prevent users from ending up perusing the collection in a zapping-like,

[8] http://www.cs.vu.nl/
[9] http://www.let.vu.nl/
[10] http://www.beeldengeluid.nl
[11] http://www.rijksmuseum.nl/

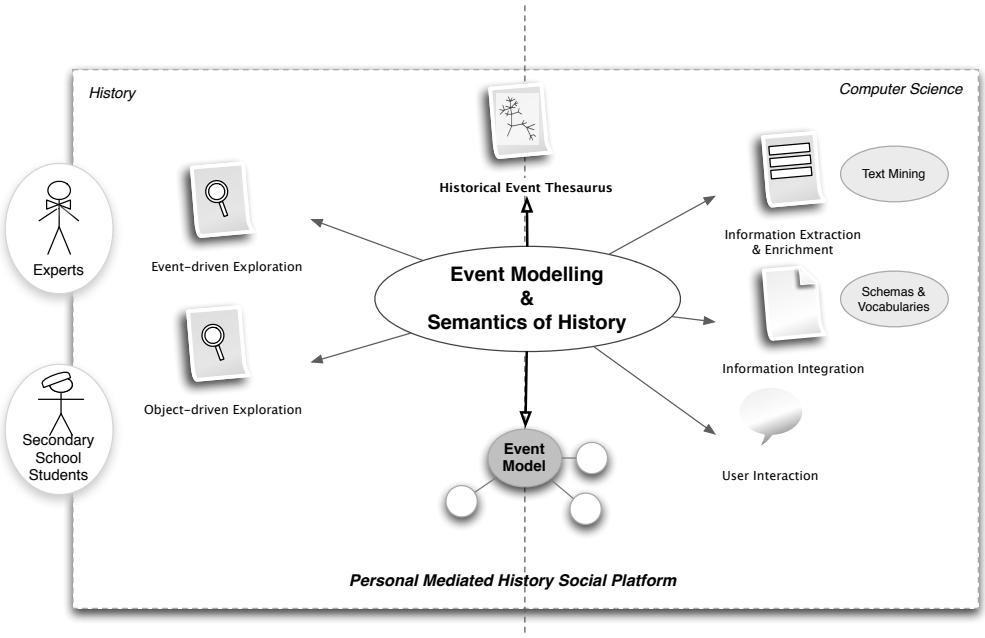

Fig. 1. Overview of domains and goals involved in Agora

incidence-based viewing that will only lead to a confirmation of his or her preconceived views and insights as no relations between objects are presented that are novel or surprising to the user. The answer of museums to this challenge so far has been to prepare thematic 'Web-specials': portals and other Web-presentations that present historical narratives as an extension of the regular exhibition and education practice. With an approach that is centred around historical events for collection access we aim to strengthen the role of museum collections in the public discourse about the past.

Although museum collections and their accompanying information are becoming available in digital forms, search is often limited to keyword search and browsing through predefined facets[1]. These access methods are not optimal; in keyword search, for example, it is not clear how the retrieved results are related to each other as they are simply presented in a list. Specifying a search query through facets resolves this problem, as it enables users to specify which relations they find relevant. However, facet browsing is often limited in that there is usually one set of facets that may not provide sufficiently finegrained access to all artefacts. Most museum collections, for example, are searchable through facets that describe meta-data that is available for every object such as *title*, *year*, *artist*, *technique*, *dimensions* and *object id*, but users may also want to search via locations that play a role in the object (for example because a location is depicted in the artwork, or because the birthplace of the artist might be relevant). Combined access to heterogeneous collections from different institutions only augments the shortcomings of keyword and facet access strategies as keyword search will often return more unorganised results when more collections are searched, and facets from different collections are often incompatible.

In addition to the general technical shortcomings of current access methods, there are shortcomings that stem from the cultural heritage domain. Users in this domain have a strong desire to explore collections through a personal narrative or from a personal perspective. Currently, cultural heritage collections lack event-based annotations that would provide the context to facilitate such explorations, but cultural heritage institutions have expressed the wish to have a formal definition of events to include in their annotation of collections.

We argue that collection access through events can remedy these bottlenecks as events provide the context that can link a variety of objects together, providing a more comprehensive overview than facets. To facilitate cross-collection access, we aim to develop an event thesaurus with which different collections can align their internal thesauri. Furthermore, our approach will combine both searching and browsing as this ensures maximum flexibility for the user to explore the collection whilst keeping track of relations between objects. We speculate that by providing a social platform for history, laypersons and experts will complement each other in the process of creating a digitally mediated public history. We believe in an open and social environment in which lay and expert users can together explore and contribute to the evolving collections of objects, events and thesaurus terms. In this way, we will research and develop ways to support a dynamic (community-based and event-centred) creation of narratives of digitally presented material objects as well as multimedia objects. Events are central to narrative and perspective. For a narrative is a sequence of events with a beginning, middle and end, and different sequences of events provide different perspectives on those events.

We envision two types of exploration: object-driven and event-driven exploration. Object-driven exploration involves a search or browsing activity where the user starts by selecting an object from the collections and subsequently finds new objects and events through the relations with the first object. In the event-driven exploration, the user starts by selecting an event and builds a sequence of related events and objects. As a user may hop between events and objects on his or her search through the collections the object-driven and event-driven explorations alternate. An example of a cross-collection exploration scenario in the Agora platform is presented in Figure 2. In this scenario, the object-driven and event-driven exploration is presented as an alternative to the typically currently used small, fixed set of relations imposed by the owner of the collection. It enables the user to wander through the RMA and S&V collections via event-based relations between collection objects that are most relevant to the user. In the example illustrated in Figure 2, the user starts by selecting the RMA print "Arrival of Van Spilbergen in Kandy". This object is related to an event that has *VOC* as actor and *Batavia* as place. In this way, the user can explore these facets and discover in the results another RMA painting "The Castle of Batavia, seen from West Kali Besar" depicting the Tradeport of Batavia. Via this object, the user can find another object from the RMA collection that depicts an event that takes place such as "A Tea Visit in Batavia". This object is related to a set of events, such as *acts of colonialism*, which can also have sub-events, e.g. *Police Actions*. The user can choose any of these events or sub-events to explore the collections further and for example arrive at one of the S&V videos that reports on the *Police Actions*. The user then continues the sequence along the facet of another sub-event *Indonesian War for Independence*, which offers the S&V video "Suriname and the Netherlands Antilles" annotated with the sub-event *Suriname's Independence*. Finally, the user is recommended (from an external resource) the sculpture "Slavery" located in the Oosterpark in Amsterdam, annotated with the same *Suriname*.

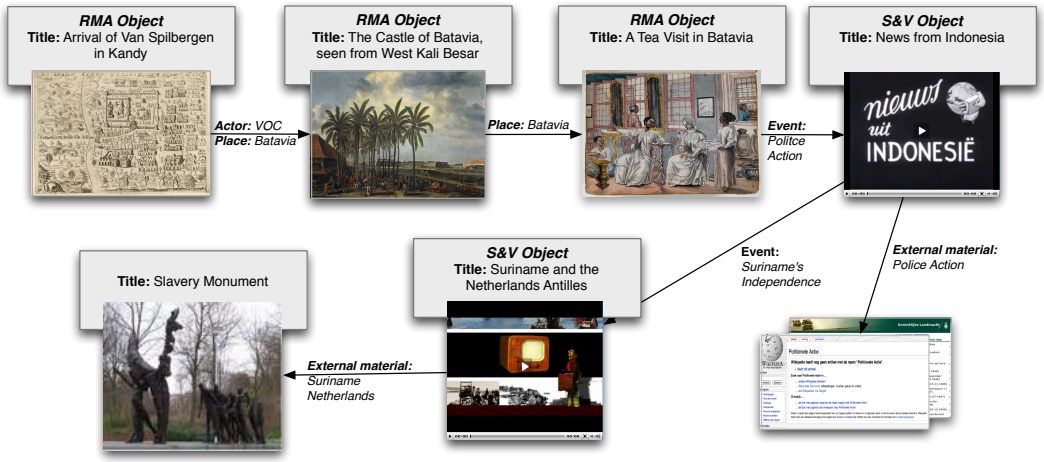

Fig. 2. Agora Exploration Scenario

2.1 Use Cases

We have defined two use cases that illustrate the need to present the same collections in different ways to different user groups. The exploration scenario presented above fits in our first use-case: assisting secondary school students for their Culture and Society end assignment. Since 1998, every student in the two higher tiers of secondary education in the Netherlands is required to write a piece on a particular topic. To facilitate the search for relevant objects and references for (art)historic topics, we will build a use-case that is focussing on event-driven browsing for different museum collections. By combining different objects, students are enabled to present their own view on the events they write about. Different relations between objects and events can result in different views on events.

Our second use case is aimed at facilitating experts in (art)historic research. For historical research, these experts want access to all objects that are related to certain events. To them, Agora platform may present an overview of objects and events a certain actor may be related to, and thus aiding the researcher in his or her information gathering task. We may also help curators to document new objects and add data to old ones, providing structured data by means of the event-model.

3 Challenges

Three domain challenges lie at the basis of our work: (1) sharing terminology, (2) understanding the meaning of events, and (3) building a historical event thesaurus. Each of these challenges is detailed below.

Sharing terminology
A clear definition of the shared terminology is a prerequisite for successful multidisciplinary collaboration. This need is particularly pressing within collaborative projects in the field of cultural

heritage and computer science such as ours as each field has different definitions and theories about shared concepts. Even our central concept, *event*, is treated differently by the parties involved in our project; in the history domain, the notion of an event has been defined as "what agents make happen or undergo"[2]. This has the implication that actions are a species of events. Furthermore, events are concrete particulars. They are unrepeatable entities with a location in space and time[3]. Within computational linguistics, the notion of event is often not defined, and if defined, the definition is mostly pragmatic and broad to ensure reusability across different domains. Another difference is that in computer science 'event extraction' often does not stretch beyond the literal task of identifying event labels, participants, locations and time stamps, whereas historians are interested in the interpretation of events. Computer science also considers events mostly as separate entities, whereas historians consider events in their connection with other events. The significance of an event depends on this connection and is usually expressed in the form of a narrative[4]. Through continual dialogue we are acknowledging the differences in each other's dealing with shared concepts and ensure that we have a stable middle-ground to work from.

Understanding the meaning of events
We are in the process of finding out what the notion of an event means for computer science and for history, and how to incorporate both views in an event model. Although there is no consensus on the definition of event in computer science, most event modelling approaches share the characteristic that they want to model: *Who does what, where and when?*. We take this as at least the minimal requirement an event modelling approach should be able to express. Once a minimal event definition has been developed, we can start to think about modelling additional aspects that play a role in our domain and are closely related to events, namely: granularity, interpretation, perspective, and causality. We are currently investigating the use of the simple event model (SEM) to model historical events[5]. SEM aims to provide the minimal set of classes to describe events, minimising possible clashes between different domain-specific event definitions. It is designed to use external type definitions and has mappings to other models such as DOLCE[12] and LODE[13]. We may also borrow from other models such as F[6] for modelling causation and interpretation to extend SEM. However, we do not make any commitments to a particular event model as we aim for a flexible approach in which we can switch to more specific or general models when the need arises.

Building a historical event thesaurus
In order to build a historical event thesaurus, it is important to investigate how historical events are referred to and how they relate to each other in the museum collections. This provides insights into how museum collections are annotated. Next, we will try to model these events in such a way that they can be used to provide better cross-collection access for diverse groups of users. We therefore want the model to be rich enough to capture the intricacies of historical events, but flexible enough to ensure its understandability. The first two steps should result in a thesaurus that can be used to better support users in searching and browsing rich and heterogeneous museum collections.

In addition, we identify three different types of technical challenges that come into play when one wants to disclose museum collections: (1) information extraction and enrichment, (2) information integration, and (3) user interface design. Fortunately, the information extraction, enrichment and integration will not have to start from scratch, as the Rijksmuseum Amsterdam and Netherlands Institute for Sound and Vision have linked their collections to existing domain-specific vocabularies

[12] `http://www.loa-cnr.it/DOLCE.html`
[13] `http://linkedevents.org/ontology/`

and thesauri such as AAT[14] and Iconclass[15] but also to general vocabularies, such as WordNet[7]. For the user interface, we can build on previous work from the MultimediaN E-Culture project[8]. In the following subsections, we will first discuss the two technical challenges and then our approach to dealing with them. The issues that the event modelling and object- and event-driven collection exploration requirements pose on the interface design depends on the outcomes of resolving the first two technical challenges and will be addressed in a later stage of the project.

3.1 Information Extraction and Enrichment

One of the biggest challenges is information extraction on historical events from different textual data. We will start by identifying the 'bigger' events that have been deemed important enough to receive a proper name (e.g., *French Revolution*, *Second World War*). As the behaviour of references to this type of events is similar to other named entities, we are recasting the identification of these event labels as an named entity recognition task[9]. In order to detect accompanying actors, locations and temporal information and the relations between these, as well as smaller events that do not have a proper name we will first employ state of the art named entity and term recognition techniques[10], followed by relation finding[11].

Once we are able to detect references to events, we will need to identify which references belong to distinct events, and which are a variation on the description of the same event. From manual extraction from a small number of newspaper texts, we observed that by identifying time, locations and participants of events and by defining their relations with those of other events, are able to relate historical events with each other. We are developing different matching functions to determine what descriptions refer to the same historical events and test these in larger and more heterogeneous corpora (consisting of e.g., collection catalogues, historical resources, and secondary literature). For example, every description that involves the same type of event, the same participants, the same location and the same time period is likely to refer to the same event. Another heuristic we are investigating involves descriptions that abstract from any of those elements but that we can identify as being semantically compatible which may indicate that these descriptions also refer to the same event. Typically, we see that different descriptions add or leave out details of what happened or group events into large happenings with bigger impact[12].

By means of the event model, relations between specific event descriptions and the more general ones can be represented; also event presentations influenced by writers' perspective towards historical events can be recognised and captured. Eventually, we want to be able to maximise the recall for finding events in text regardless of the description and secondly be able to infer added subjectivity and interpretation layers to the events.

3.2 Information Integration

The RMA and S&V collections represent different components of the Dutch cultural heritage. The Rijksmuseum Amsterdam focusses on art, crafts and history. A large part of its one million object collection consists of 17th century Dutch paintings. The Netherlands Institute for Sound and Vision aims at preserving the Dutch audiovisual heritage. Its collection contains about 700,000 hours of radio, television, movie and music material, of which most is less than fifty years old. Although the two collections are different in age and focus, there is a fair overlap in the topics they deal with

[14] `http://www.getty.edu/research/conducting_research/vocabularies/aat`
[15] `http://www.iconclass.nl/`

as the S&V collection contains, for example, documentaries about events that are also depicted or play a role in the RMA collection.

In order to access these two collections simultaneously we first need to align the collection metadata schemas with each other (e.g., *artist* in the RMA collection database may correspond to *creator* in the S&V collection database). From previous experience in the MultimediaN E-Culture project we have learnt that a good way to do this is to map both to an accepted metadata schema such as Visual Resource Association core categories (VRA)[16][8]. To consolidate the collection integration, we also aim to map the values of the fields in the collection databases to a shared vocabulary[13] or to other relevant external resources such as the Dutch Biography Portal[17].

The backbone of our event-driven and object-driven exploration method shall be based on ClioPatria which provides a basis for collection exploration that combines searching and browsing[14]

4 Open Issues

One of the most central open question in this work focusses on what are the recurring historical events that can be traced, in the RMA and S&V collections, based on documentations and attributions of meaning and provenance; and how can we interpret the historical time-lines and narratives that emerge from such a search, using semantics to derive and explain various views, biases, contradictions, opinions and emotional reactions? Currently, in various types of historic documents and collections, events are captured with a single interpretation or perspective. However, we aim at *allowing for multiple local, national, international, and personal perspectives* on historical events and their sequences. In this context we are in search for answers to the following questions:

- how can events be placed in historical sequences, in the context of various collections that address different past-relationships;
- how to include different perspectives in individual events and in event narratives;
- how to extract and model causal relations between events;
- how to involve and motivate the end users in the process of collaborative editing of historical event narratives

Critical for the success of this research is to step upon previous experiences and analyse the implicit historical event model that cultural institutions created by constructing their collections and collection description (i.e., what do events mean to them and how are they represented?). It is interesting to explore the past selection and interpretation processes in order to facilitate new access to (enriched) cultural heritage data and to ultimately investigate how this digitally mediated public history is related to current history writing.

Finally, to allow for effective deployment of the research results we need to specify the envisioned role of a social cultural heritage platform for both the cultural heritage professionals and for the well-informed or interested lay people; and to what extent this platform will be the main drive to maintain the dynamics both in the shared historical thesaurus and the historical events descriptions and their relationships.

[16] `http://www.vraweb.org/projects/vracore4/`
[17] `http://www.biografischportaal.nl`

Acknowledgements

Agora is funded by NWO in the CATCH programme and Semantics of History is funded by VU University of Amsterdam's Interfaculty research institute CAMeRA.

References

1. Cohen, D.J.: History and the second decade of the web. Rethinking History **8**(2) (2004) 293–301
2. Ricoeur, P.: Time and Narrative. Volume 1. Chicago and London (1984)
3. Davidson, D.: Essays on Action and Events. Oxford (1980)
4. Danto, A.: Analytical Philosophy of History. Cambridge (1968)
5. van Hage, W.R., Malaisé, V., de Vries, G., Schreiber, A.T., van Someren, M.: Combining ship trajectories and semantics with the simple event model (sem). In: Proceedings of 1st ACM International Workshop on Events in Multimedia (EiMM09), Bejing, China, ACM (October 23 2009)
6. Scherp, A., Franz, T., Saathoff, C., Staab, S.: F–a model of events based on the foundational ontology DOLCE+DnS ultralight. In: Proceedings The Fifth International Conference on Knowledge Capture (K-CAP 2009), Redondo Beach, CA, USA, ACM (2009)
7. Fellbaum, C., ed.: WordNet: An Electronic Lexical Database. The MIT Press (1998)
8. Schreiber, G., Amin, A., Aroyo, L., van Assem, M., de Boer, V., Hardman, L., Hildebrand, M., Omelayenko, B., van Osenbruggen, J., Tordai, A., Wielemaker, J., Wielinga, B.: Semantic annotation and search of cultural-heritage collections: The MultimediaN E-Culture demonstrator. Journal of Web Semantics **6**(4) (2008) 243–249
9. Sundheim, B.M.: Overview of results of the muc-6 evaluation. In: Proceedings of the 6th conference on Message understanding. (1993) 13–31
10. Ratinov, L., Roth, D.: Design challenges and misconceptions in named entity recognition. In: Proceedings of Thirteenth Conference on Computational Natural Language Learning (CoNLL 2009), Boulder, CO, USA (2009) 147–155
11. Suchanek, F.M., Ifrim, G., Weikum, G.: Combining linguistic and statistical analysis to extract relations from web documents. In: Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining, Philadelphia, PA, USA (2006) 712–717
12. Cybulska, A., Vossen, P.: Event models for historical perspectives: determining relations between high and low level events in text, based on the classification of time, location and participants. In: To Appear in: Proceedings of LREC 2010, Valletta, Malta (2010)
13. Tordai, A., Omelayenko, B., Schreiber, G.: Thesaurus and metadata alignment for a semantic e-culture application. In: Proceedings of the 4th international conference on Knowledge capture (K-CAP'07), Redondo Beach, CA, USA, ACM (2007) 199–200
14. Wielemaker, J., Hildebrand, M., van Ossenbruggen, J., Schreiber, G.: Thesaurus-based search in large heterogeneous collections. In: The Semantic Web - ISWC'08. Volume 5318 of LNCS., Tenerife, Spain, Springer-Verlag (May 2008) 695–708

A Representation Language for Describing and Managing Elementary/Complex Events in a Non-Fictional Narrative Context

Gian Piero Zarri[1]

[1] University Paris-Est, Créteil, Val de Marne (UPEC), LiSSi Laboratory, 120-122 rue Paul Armangot, 94400 Vitry-sur-Seine, France
gian-piero.zarri@u-pec.fr, zarri@noos.fr

Abstract. Making reference, mainly, to the "non-fictional narratives" domain, we will suggest, firstly, an operational definition for highly ambiguous terms like "elementary events" and "complex events". We will then raise the problem of how to represent these events/complex events in a computer-suitable form. We will introduce therefore a language, NKRL (Narrative Knowledge Representation Language), expressly specified and implemented for dealing with narratives and temporal information. Afterwards, we will show briefly how this language can be used for questioning and inferencing operations on knowledge bases of "events" formalised according to the NKRL approach.

Keywords: Elementary events, complex events, knowledge representation, querying and inferencing.

1 Introduction

"Event-based" tools and systems seem to be particularly popular today. They range from highly formalised systems like event algebras [1] and event ontologies [2], to practical applications like, e.g., narrative-based video annotation and editing [3], the design of event-driven software architectures integrating actuator and sensor networks [4], or the use of an event-centric approach for modelling policy decision-making in a business process context [5]. For the great majority of these tools and systems, however, the term "event" seems to denote simply a sort of 'primitive' or 'intuitively understood' notion that does not ask for any sort of definition. When this last is given, it is often limited to basic statements like "something that happens at a given place and time", "something that takes place", "perduring entities that unfold over time" or – in the glossary of terms of the Event Processing Technical society – "anything that happens, or is contemplated as happening".

In the framework of the NKRL project (NKRL = Narrative Knowledge Representation Language), see [6, 7, 8, 9], we have now found *satisfactory definitions, even if largely pragmatic and operational*, for basic essential notions like those of "*events*" and "*complex events*". These definitions have been derived from work concerning the proper NKRL domain, i.e., the representation and management of *non-fictional narrative information* (or *non-fictional narratives*) – even if, as we

will show below, extensions to other domains dealing with the notion of 'event' are certainly possible.

In the following, we will first explain informally, Section 2, what is denoted by terms like "elementary/complex events" in a narrative/NKRL context. We will then show, Section 3, how these informal notions can be translated into precise formal representation structures that constitute the gist of the NKRL language. Section 4 will supply some details about NKRL and will include two sub-sections, the first devoted to an outlook of the NKRL's knowledge representation techniques and the second to the querying/inferencing procedures. Section 5 will consists in a short "Conclusion".

2 Narratives, elementary and complex events

'Narrative' information concerns the account of some real-life or fictional story (a "narrative") involving concrete or imaginary 'characters'. In a NKRL context we are mainly concerned with "*non-fictional narratives*", like those typically embodied into corporate memory documents (memos, policy statements, reports, minutes etc.), news stories, normative and legal texts, medical records, many intelligence messages, surveillance videos, actuality photos for newspapers and magazines, material (text, image, video, sound…) for eLearning, Cultural Heritage material, etc. We can note that this choice is only due to very practical constraints – to profit, e.g., from the financial support of the European Commission – and, as it will appear clearly in the following, nothing (at least in principle) could prevent us from dealing with the whole "Gone with the wind" *fictional-narrative* novel according to an NKRL approach.

More precisely, we assume that (fictional or non-fictional) narratives correspond to the basic layer, the "*fabula* (a Latin word: fable, story, tale, play) *layer*", introduced by Mieke Bal [10] in her crucial work on the structures of narrative phenomena. Accordingly, an (NKRL) narrative consists then in a *series of logically and chronologically related events (a 'stream of elementary events') that describe the activities or the experiences of given characters*. From the above, we can immediately deduce that *a narrative coincides, in practice, with a "complex event"* – see also [8]. From this definition and other work in a "*narratology*" context – an introduction to this discipline can be found in [11] – we can infer some important characteristics of "narratives/complex events", see [9: 2-13] for a more detailed discussion – independently, once again, by any 'fictional/non fictional' consideration:

- One of the features defining the *connected* character of the elementary events that make up the stream concerns the fact that these are *chronologically related*, i.e., narratives/complex events *extend over time*. This diachronic aspect of narratives/complex events (a narrative normally has a *beginning*, an *end* and some *form of development*) represents indeed one of their most important characteristics.
- *Space* is also very important in the narrative/complex events domain, given that *the elementary events of the stream occur generally in well defined 'locations'*, real or imaginary ones. The connected events that make up a narrative/complex event are then both *temporally and spatially bounded*. Bakhtin [12] speaks in this context of "*chronotopes*" when drawing attention on the fact that time and space in narratives are strictly interrelated.

- A simple chronological successions of elementary events that take place in given locations cannot, however, be defined as a 'narrative' (a complex event) without some sort of '*semantic coherence*' and '*uniqueness of the theme*' that characterise the different elementary events of the stream. If this logical coherence is lacking, the elementary events pertain to different narratives: a narrative can also be represented by a single 'elementary event'.
- When the constitutive 'elementary events' of a narrative/complex event are verbalized in NL terms, their 'coherence' is normally expressed through syntactic constructions like causality, goal, indirect speech, co-ordination and subordination, etc. In this paper, we will systematically make use of the term '*connectivity phenomena*' to denote this sort of clues, i.e., to denote what, in a stream of elementary events, i) leads to a 'global meaning' that goes beyond the simple addition of the 'meanings' conveyed by a single elementary event; ii) defines the influence of the context in which a particular event is used on the meaning of this individual event, or part of it.
- Eventually, narratives/complex events concern the behaviour or the condition of some 'actors' (persons, characters, personages, figures etc.). They try to attain a specific result, experience particular situations, manipulate some (concrete or abstract) materials, send or receive messages, buy, sell, deliver, etc. In short, *they have a specific 'role' in the event* (in the stream of events representing the global narrative) – see, in a very peculiar 'narratology' context, the famous seven roles (the hero, the villain, the princess etc.) described by Vladimir Propp in its "Morphology of the Folktale" [13]. Note that these actors or characters are not necessarily human beings; we can have narratives concerning, e.g., the vicissitudes in the journey of a nuclear submarine (the 'actor', 'character' etc.) or the various avatars in the life of a commercial product.

Defining a narrative/complex event as a *'stream of elementary events'* would correspond, once again, to some sort of 'dull' definition without being able to specify what an "*elementary event*" can be. In an NKRL context, this point is also particularly important from a practical point of view given that, as we see later, *each elementary event is separately encoded making use of the NKRL knowledge representation tools*. According then to a well-known Jaegwon Kim's definition, see [14, 15], a "monadic" event – which can be considered as equivalent to an "elementary event" – is identified by a triple [*x*, *P*, *t*] where *x* is an object that exemplifies the *n*-ary property or relation *P* at time *t* (where *t* can also be an interval of time); "monadic" means then that the *n*-ary property *P* is exemplified by a single object *x* at a time. To make reference to one of the recurrent examples in the theoretical discussions about events, "Brutus stabs Caesar", the Kimian interpretation of this event corresponds then to the representation of an individual *x*, Brutus who, at time *t*, is characterised by the property *P* exemplified by his stabbing of Caesar. Without entering now in the theoretical controversies raised by this sort of definition, see [9: 8-13] for some information in this context, we can note that, from an NKRL point of view, a more 'practically useful' – more complete and structured – definition of elementary event is that introduced by Donald Davidson [16, 17], particularly popular in the linguistic domain. This last focuses the representation of an elementary event on the "*action verb*" characterising the *global conceptual category* of the event more than – as in the

Kimian approach – on the "*generalised properties*" of this event. In this way, the Davidsonian representation of "Brutus stabs Caesar" becomes: $\exists e.\mathrm{stab}(e, b, c)$, where e is an *event variable*. The global meaning of this formalism corresponds to: "There is an event e such that e was a stabbing of Caesar (c) by Brutus (b)". Moreover, as emphasised above when we have listed some important characteristics of narratives/complex events, *"roles" have a particular importance in a narrative*. Our preferred formalism for the representation of *elementary* events is then the so-called "*neo-Davidsonian*" approach: the neo-Davidsonians, see [18, 19, 20], assume in fact that the event argument e above must be *the only argument of the (verbal) predicate*: this implies then, necessarily, *the introduction of thematic (functional) "roles" for expressing the relations between events and their participants*. The formalization of "Brutus stabs Caesar" becomes then now:

$$\exists e[\mathrm{stab}(e) \,\&\, \mathrm{agent}(e) = b \,\&\, \mathrm{object}(e) = c]\,. \qquad (\mathbf{1})$$

Apart from the theoretical implications, what expounded above is particularly important because it supplies us with a *'pragmatically useful' and 'operational' criterion for recognizing and isolating – in some way, for 'defining' – "elementary events"*. The criterion consists then in the identification, *within the description in natural language (NL) terms of the global stream representing a narrative/complex event, of a specific 'generalized natural language predicate'*: this represents then the 'core' of a new elementary event. The predicate corresponds usually to a verb – to stick to the previous example, recognizing "stabs" as a verb in the NL chain "Brutus stabs Caesar" should be sufficient for signalling the presence of an elementary event – but, according to the neo-Davidsonian approach, this predicate can also, in case, correspond to a noun (…Jane's *amble* along the park…) or an adjective ("… *worth* several dollars…") when these last have a *predicative function*. Of course, a drawback of this criterion concerns the fact that its utility is limited to the recognition of the elementary components of narratives/complex events expressed in NL terms, while narratives are "multimedia" in essence – a photo representing President Obama addressing the Congress, or a short video showing three nice girls on a beach are surely narrative documents (the first including only an elementary event) but they are not, of course, NL documents. A classical way of getting around this problem consists in annotating multimedia narratives in natural language see, e.g., [21]. In [3], Lombardo and Damiano propose a method for annotating and editing video fragments (i.e., video narratives) with respect to their semantic content where the basic unit of description – corresponding then to our 'elementary events' – are "beats" defining "… the minimal units for story advancement that will exposed to the audience" [3: 707]. The criteria used to identify the "beats" in an unambiguous and repeatable way are not, however, completely defined.

3 Representing Elementary and Complex Events

Eq. 1 above – an *n*-ary form of representation – shows clearly that the now so popular W3C proposals like RDF(S), OWL or OWL 2 – see [22, 23, 24] – are, *at least in their standard format*, unable to supply a basis for representing elementary events on a

computer. All the W3C representations are, in fact, of the *binary* type, based on the classical '*attribute – value*' model, where a property/attribute can only be a *binary relationship* linking two individuals or an individual and a value. The inadequateness of this approach to take into account complex representational problems like those linked with narratives, spatio-temporal information, and any sort of events and complex events is now widely recognized see, e.g., [25, 26, 27, 28, 29].

Note that the argument often raised in a W3C context and stating that any representation making use of *n*-ary relations can be always converted to one making only use of binary relations *without any loss of expressiveness* is incorrect with respect to the last part of the sentence. It is true in fact that, from a pure formal point of view, any *n*-ary relationship with $n > 2$ can always be reduced, in a very simple way, to a set of binary relationships. This possibility is well illustrated, among other things, by the successful representation of the NKRL's core in the terms of a (W3C) binary language like RDF, see [30]. However, this fact does not change at all the *intrinsic, 'semantic'* n-*ary nature* of a simple statement like "Bill gives a book to Mary" that, to be understood, requires to be taken *in its entirety*. This means – see also Eq. 1 above – to make use of a semantic predicate of the GIVE type that introduces its *three* arguments, "Bill", "Mary" and "book" through *three* functional relationships (roles) like SUBJECT (or AGENT), BENEFICIARY and OBJECT, the whole *n*-ary construction being – this is the central point – *reified and necessarily managed as a coherent block at the same time*. Only in this way it will be possible to infer, in a querying/inferencing context that, e.g., the above elementary event is linked, in the framework of a wider narrative, to another elementary event relating Mary's birthday; for the formal details see, e.g., [25].

Efforts done, in a strict W3C context, to extend their binary languages using some n-*ary* features have not been very successful until now, see, for more details, [9: 17-21]. Another way of trying to adapt/extend the traditional 'binary' tools to take into account complex, dynamic situations consists *in reducing the notion of "role" from its normal status of 'functional relationship' to that of 'static' concept or class*, see [9: 138-149] for a discussion about this topic. These role-concepts can then be used within all sorts of complex *binary schemata or patterns* to represent causality, mereology, participation etc. A well-known example of this approach is the "Descriptions and Situations (DS)" model, see [31], implemented as a plug-in extension to the DOLCE system [32], an (OWL-based) 'upper ontology'. A recent variation on this approach is represented by Event-Model-F, see [33], based in turn on a reduced version of DOLCE, DOLCE+DnS Ultralite (DUL), see http://www.loa-cnr.it/ontologies/DUL.owl.

Several actual *n*-ary models that, among other things, can be used to represent in computer-usable ways elementary events have been described in the literature, see [9: 22-33] for a review. The *n*-ary model used in NKRL can be denoted as:

$$(L_i\ (P_j\ (R_1\ a_1)\ (R_2\ a_2)\ \ldots\ (R_n\ a_n)))\ , \qquad (2)$$

where L_i is the symbolic label identifying ('reifying') the particular *n*-ary structure (e.g., the global structure corresponding to the representation of previous "John gives a book to Mary" example), P_j is a conceptual predicate, R_k is a generic "functional role" and a_k the corresponding predicate argument (e.g., the individuals JOHN_, MARY_ etc.). Note that each of the $(R_i\ a_i)$ cells of Eq. 2, *taken individually*, represents

a *binary relationship* in the W3C (OWL, RDF…) languages style. The main point is here that, as already stated, *the whole conceptual structure represented by (2) must be considered globally.*

Similarities between neo-Davidsonian expressions for elementary events like that of Eq. 1 and the formal structure of Eq. 2 are evident. However, some important differences exist. To avoid both the typical ambiguities of natural language and possible 'combinatorial explosion' problems – see the discussion in [9: 56-61] – both the (unique) conceptual predicate of Eq. 2 and the associated functional roles are 'primitives'. Predicates P_j pertain in fact to the set {BEHAVE, EXIST, EXPERIENCE, MOVE, OWN, PRODUCE, RECEIVE}, and the functional roles R_k to the set {SUBJ(ect), OBJ(ect), SOURCE, BEN(e)F(iciary), MODAL(ity), TOPIC, CONTEXT}. Two special operators, date-1 and date-2 – that can be assimilated to functional roles – are used to introduce the temporal information associated with an elementary event: see, e.g., [9: 76-86, 194-201] for a detailed description of the *formal system* used in NKRL for the representation and management of temporal information. The NKRL representation of specific elementary events – that corresponds to the *concrete instantiations* (called "*predicative occurrences*") of general structures in the style of Eq. 2 – is then a sort of *canonical representation.*

Several predicative occurrences – denoted by their symbolic labels L_i and representing formally a (possibly structured) set of elementary events – can be associated within the scope of second order structures called "*binding occurrences*". These are, in practice, *labelled lists* formed of a *"binding operator Bn" with its arguments*. The operators are those used in NKRL to represent the "*connectivity phenomena*" that guarantee the *global coherence of narrative/complex events*, see Section 2 above. They are: ALTERN(ative), COORD(ination), ENUM(eration), CAUSE, REFER(ence) – the 'weak causality operator', introducing two arguments where the second is necessary but not sufficient to explain the first – GOAL, MOTIV(ation) – the 'weak intentionality operator', where the first argument is not necessary to realise the second, which is however sufficient to explain the first – COND(ition), see [9: 91-98]. The general expression of a binding occurrence is then:

$$(Bn_k \ \mathrm{arg}_1 \ \mathrm{arg}_2 \ \ldots \ \mathrm{arg}_n) \ , \qquad (\mathbf{3})$$

Eq. 3 is particularly important in an NKRL context because it supplies also the *formal expression – once again, a 'pragmatic/operational' form of definition – of the notion of "narrative/complex event"*. The arguments arg_i of Eq. 3 can, in fact, i) correspond directly to L_i labels – i.e., they can denote simply the presence of particular elementary events represented formally as predicative occurrences – or ii) *correspond recursively to new labelled lists in Eq. 3 format*. In the first case, the global narrative/complex event represents merely *a chronological stream of elementary events, temporally characterized, where all these events have the same logical/semantic weight* and the operator *Bn* corresponds to COORD (or ENUM/ALTERN). In the second case, we can suppose, e.g., that a given sequence of events – an Eq. 3 list of the COORD... type – represents the CAUSE of another sequence of events. The global representation of this narrative/complex event will then correspond to an Eq. 3 list labelled as CAUSE, having as arguments arg_1 the COORD... list including the elementary events at the origin of the complex event and as arg_2 the COORD... list including the elementary events that represent together the

consequence, see also the simple example at the end of Section 4.1 below. What expounded above is in agreement with the remarks expressed by several authors – see [34] for example – about the possibility of *visualizing under tree form the global, formal expression of a narrative/complex event made up of several elementary events.*

4 A Short Description of the NKRL system

After having introduced, in the previous Sections, the general theoretical framework underpinning the NKRL approach to the narrative/complex events problem, we will now illustrate briefly some points concerning its concrete implementation – see [9] for a complete description.

3.1 The Knowledge Representation Aspects

NKRL innovates with respect to the current *ontological paradigms* by adding to the usual '*ontologies of concepts*' an '*ontology of (elementary) events*', i.e., a new sort of hierarchical organization *where the nodes correspond to* n-*ary structures in the style of Eq. 2 above*. In the NKRL's jargon, these *n*-ary structures are called "*templates*" and the corresponding hierarchy – i.e., the ontology of elementary events – is called HTemp (*hierarchy of templates*). Templates can be conceived as the canonical, formal representation of *generic classes of elementary events* like "move a physical object", "be present in a place", "produce a service", "send/receive a message", etc.

Note that, in the NKRL environment, an '*ontology of concepts*' (according to the traditional meaning of these terms) not only exists, but it represents an essential component of this environment. The 'standard' ontology is called HClass (*hierarchy of classes*): structurally and functionally, HClass is *not fundamentally different* from one of the ontologies that can be built up by using tools in a 'traditional' Protégé style, see [35]. An (extremely reduced) representation of HClass is given in Figure 1 – HClass includes presently (April 2010) more than 7,000 concepts.

When a *specific elementary event* pertaining to one of the 'general classes' represented by templates must be encoded, the corresponding template is *instantiated* giving rise to a "*predicative occurrence*". To represent then a simple elementary event (*corresponding to the identification of the surface predicate "offer"*) like: "British Telecom will offer its customers a pay-as-you-go (payg) Internet service in autumn 1998", we must select firstly in the HTemp hierarchy the template corresponding to "supply a service to someone", represented in the upper part of Table 1. This template is a specialization of the particular MOVE template corresponding to 'transfer of resources to someone' – Figure 2 below reproduces a fragment of the 'external' organization of HTemp. In a template, the arguments of the predicate (the a_k terms in Eq. 2) are concretely represented *by variables with associated constraints*: these are expressed as HClass concepts or combinations of concepts, i.e., the two ontologies, HTemp and HClass, *are then strictly intermingled*.

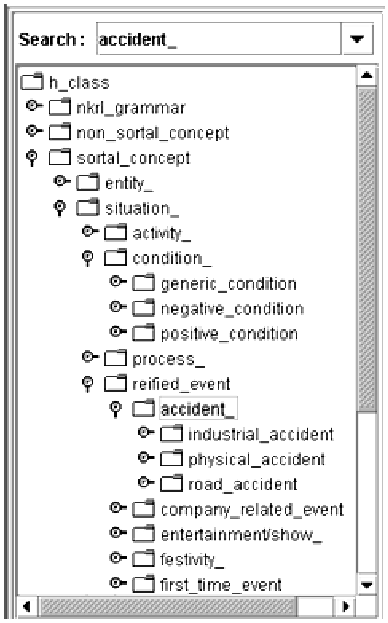

Fig. 1. Partial representation of HClass, the 'traditional' ontology of concepts.

Table 1. Deriving a predicative occurrence from a template**.**

name: Move:TransferOfServiceToSomeone
father: Move:TransferToSomeone
position: 4.11
natural language description: "Transfer or Supply a Service to Someone"

MOVE	SUBJ	*var1*: [*var2*]
	OBJ	*var3*
	[SOURCE	*var4*: [*var5*]]
	BENF	*var6*: [*var7*]
	[MODAL	*var8*]
	[TOPIC	*var9*]
	[CONTEXT	*var10*]
	{[modulators]}	

var1	=	human_being_or_social_body
var3	=	service_
var4	=	human_being_or_social_body
var6	=	human_being_or_social_body
var8	=	process_, sector_specific_activity
var9	=	sortal_concept
var10	=	situation_
var2, *var5*, *var7*	=	geographical_location

c1)	MOVE	SUBJ	BRITISH_TELECOM
		OBJ	payg_internet_service
		BENF	(SPECIF customer_ BRITISH_TELECOM)
		date-1:	after-1-september-1998
		date-2:	

When creating a *predicative occurrence (an instance of a template)* like c1 in the lower part of Table 1, the role fillers in this occurrence *must conform to the constraints of the father-template*. For example, in occurrence c1, BRITISH_TELECOM is an individual, instance of the HClass concept company_: this last is, in turn, a specialization of human_being_or_social_body. payg_internet_service is a specialization of service_, a specific term of social_activity, etc. The meaning of the expression "BENF (SPECIF customer_ BRITISH_TELECOM)" in c1 is self-evident: the beneficiaries (role BENF) of the service are the customers of – SPECIF(ication) – British Telecom. The 'attributive operator', SPECIF(ication), is one of the four operators used for the set up of the *structured arguments (expansions)* of conceptual predicates like MOVE, see [9: 68-70]. In the occurrences, the two operators date-1 and date-2 materialize *the temporal interval normally associated with an elementary event*, see again [9: 76-86, 194-201].

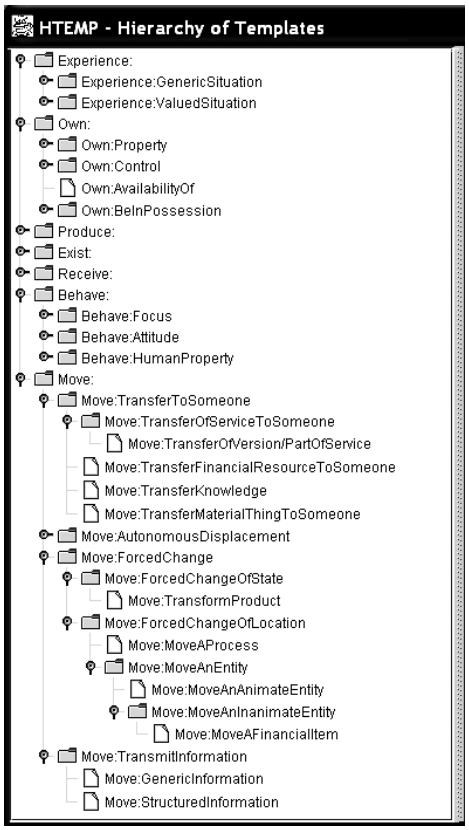

Fig. 2. 'MOVE' etc. branch of the HTemp hierarchy.

More than 150 templates are permanently inserted into HTemp; HTemp, *the NKRL ontology of events*, corresponds then to a sort of '*catalogue' of narrative formal structures*, which are very easy to extend and customize.

To supply now an at least intuitive idea of how a complete narrative/complex event is represented in NKRL, and returning to the Table 1 example, let us suppose we would now state that: "We can note that, on March 2008, British Telecom *plans to offer* to its customers, in autumn 1998, a pay-as-you-go (payg) Internet service…", where the specific elementary event corresponding to the offer is still represented by occurrence c1 in Table 1.

To encode correctly the new information, we must introduce first an *additional predicative occurrence* labelled as c2, see Table 2, meaning that: "at the specific date associated with c2 (March 1998), it can be noticed, modulator obs(erve), that British Telecom *is planning* to act in some way" – the presence of a *second surface predicate* in the NL expression of the complex event denotes the presence of a second elementary event. obs(erve) is a '*temporal modulator*', see [9: 71-72], used to identify a *particular timestamp* within the temporal interval of validity of an elementary event. We will then add a *binding occurrence* c3 labelled with a GOAL *Bn* operator, see the previous Section, to link together the conceptual labels c2 (the planning activity) and c1 (the intended result). The global meaning of c3 can be verbalized as: "The activity described in c2 is focalised towards (GOAL) the realization of c1". In agreement with the remarks at the end of the last Section, c3 – the representation of the global narrative/complex event – can also represented *under tree form*, having GOAL as top node, and two branches where the leaves are L_1 = c2 and L_2 = c1.

Table 2. Binding and predicative occurrences**.**

```
c2)  BEHAVE   SUBJ      BRITISH_TELECOM
              MODAL     planning_
              { obs }
              date1:    march-1998
              date2:

Behave:ActExplicitly (1.12)

*c1)   MOVE   SUBJ      BRITISH_TELECOM
       OBJ    payg_internet_service
              BENF      (SPECIF customer_ BRITISH_TELECOM)
              date-1:   after-1-september-1998
              date-2:

Move:TransferOfServiceToSomeone (4.11)

c3)  (GOAL c2 c1)
```

3.2 The Querying/Inferencing Aspects

Reasoning in NKRL ranges from the *direct questioning* of a knowledge base of narratives represented in NKRL format – by means of *search patterns* p_i (formal queries) that unify information in the base thanks to the use of a *Filtering Unification Module* (*Fum*), see [9: 183-201] – to *high-level inference procedures*. These last make use of the richness of the representation to establish 'interesting' relationships among the narrative items stored within the base; a detailed paper on this topic is [6].

The NKRL rules are characterised by the following general properties:

- All the NKRL high-level inference rules can be conceived as *implications* of the type:

$$X \text{ iff } Y_1 \text{ and } Y_2 \ \ldots \text{and } Y_n \ . \qquad (4)$$

- In Eq. 4, X corresponds either to a *predicative occurrence* c_j or to a *search pattern* p_i and $Y_1 \ldots Y_n$ – the NKRL translation of the '*reasoning steps*' that make up the rule – correspond to *partially instantiated templates*. They include then, see the upper part of Table 1 above, *explicit variables* of the form var_i.
- According to the usual conventions of *logic/rule programming*, see [33: 105-170] – *InferenceEngine understands each implication as a procedure*. This reduces '*problems*' of the form X to a *succession of* '*sub-problems*' Y_1 and $\ldots Y_n$.
- Each Y_i is interpreted in turn as a *procedure call* that tries to convert – using, in case, *backtracking procedures* – Y_i *into (at least) a successful search pattern* p_i. These last must then be able to unify (using the standard *Fum* module) one or several of the occurrences c_j of the NKRL knowledge base.
- The *success* of the unification operations of the pattern p_i derived from Y_i *means that the 'reasoning step' represented by* Y_i *has been validated*. *InferenceEngine* continues *trying then to validate the reasoning step corresponding to* Y_{i+1}.
- In line with the presence of the operator 'and' in Eq. 4, the implication represented by Eq. 4 *is fully validated iff all the reasoning steps* $Y_1, Y_2 \ldots Y_n$ *are validated.*

All the unification operations p_i/c_j required from the inference procedures make use only of the unification functions supplied by the Filtering Unification Module (*Fum*) introduced above. Apart from being used for the *direct questioning* operations, *Fum* constitutes as well, therefore, the '*inner core*' of the *InferenceEngine* modules.

From a practical point of view, the NKRL high-level inference procedures concern *mainly* two classes of rules, 'transformations' and 'hypotheses' see, e.g., [6].

Let us consider, e.g., the 'transformations'. These rules try to '*adapt*', from a *semantic* point of view, a search pattern p_i that '*failed*' (that was unable to find a unification within the knowledge base) to the *real contents* of this base making use of a sort of '*analogical reasoning*'. They attempt then to *automatically 'transform'* p_i into one or more *different* p_1, $p_2 \ldots p_n$ that *are not strictly 'equivalent' but only 'semantically close'* (analogical reasoning) to the original one. In a transformation context, the '*head*' X of Eq. 4 is then represented by a search pattern, p_i.

Operationally, a transformation rule can be conceived as made up of a *left-hand side*, the '*antecedent*' – i.e. the formulation, in search pattern format, of the 'query' to be transformed – and of one or more *right-hand sides*, the '*consequent(s)*' – the NKRL representation(s) of one or more queries (search patterns) to be substituted for the given one. Denoting with A the antecedent and with Cs all the possible consequents, the transformation rules can then be expressed as:

$$A(var_i) \ \Rightarrow \ Cs(var_j), \quad var_i \subseteq var_j \qquad (5)$$

With respect to Eq. 4 above, X coincides now with A – a *search pattern* – while the reasoning steps $Y_1, Y_2 \ldots Y_n$ are used to produce the *search pattern(s)* Cs *to be used in place of* A. The restriction $var_i \subseteq var_j$ – all the variables declared in the antecedent A

must also appear in *Cs* – assures the logical congruence of the rules. More formal details are given, e.g., in [9: 212-216].

Let us consider a concrete example, which concerns a recent NKRL application about the 'intelligent' management of 'storyboards' in the oil/gas industry, see also [37]. We want then ask whether, in a knowledge base where are stored all the possible *elementary and complex events* related to the activation of a gas turbine, we can retrieve the information that a given oil extractor is running. In the absence of a direct answer we can reply by supplying, thanks to a transformation rule like that (*t11*) of Table 3, other related events stored in the knowledge base, e.g., an information stating that the site leader has heard the working noise of the oil extractor, see Figure 3.

Table 3. An example of 'transformation' rule.

t11: *"working noise/condition" transformation*

antecedent:

OWN	SUBJ	*var1*
	OBJ	property_
	TOPIC	running_

var1 = consumer_electronics, hardware_, diagnostic_tool/system, surgical_tool, technical/industrial_tool, small_portable_equipment

first consequent schema (*conseq1*):

EXPERIENCE	SUBJ	*var2*
	OBJ	evidence_
	TOPIC	(SPECIF *var3 var1*)

var2 = individual_person

var3 = working_noise, working_condition

second consequent schema (*conseq2*):

BEHAVE	SUBJ	*var2*
	MODAL	industrial_site_operator

Being unable to demonstrate directly that an industrial apparatus is running, the fact that an operator can hear its working noise or note its operational aspect can represent a proof of its running status.

Expressed in natural language, this result can be paraphrased as: "The system cannot assert that the oil extractor is running, but it can certify that the site leader has heard the working noise of this extractor".

With respect now to the *hypothesis rules*, these allow us to build up automatically a sort of '*causal explanation*' for an elementary event (a predicative occurrence c_i) retrieved within a NKRL knowledge base. In a hypothesis context, the 'head' X of Eq. 4 then represented by a predicative occurrence, c_j. Accordingly, the 'reasoning steps' Y_i of Eq. 4 – called 'condition schemata' in a hypothesis context – *must all be satisfied* (for each of them, at least one of the corresponding search patterns p_i must find a successful unification with the predicative occurrences of the base) *in order that the set of* $c_1, c_2 \ldots c_n$ *predicative occurrences retrieved in this way can be interpreted as a context/causal explanation of the original occurrence* c_j.

For example, to mention a 'classic' NKRL example, see [6], let us suppose we have directly retrieved, in a querying-answering mode, information like: "Pharmacopeia, an USA biotechnology company, has received 64,000,000 dollars from the German company Schering in connection with an R&D activity" that

corresponds then to c_j. We can then be able to automatically construct, using a 'hypothesis' rule, a sort of 'causal explanation' of this event by retrieving in the knowledge base information like: i) "Pharmacopeia and Schering have signed an agreement concerning the production by Pharmacopeia of a new compound" (c_1) and ii) "in the framework of the agreement previously mentioned, Pharmacopeia has actually produced the new compound" (c_2).

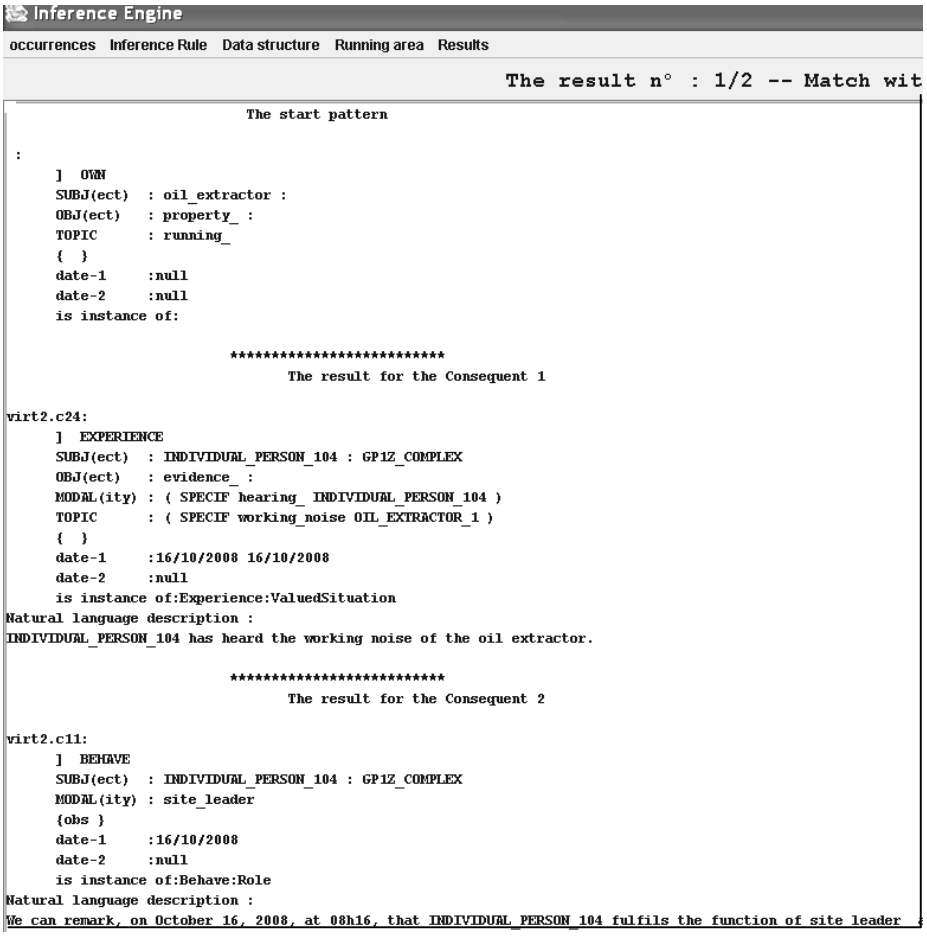

Fig. 3. Using the NKRL *InferenceEngine* in a 'transformation' context.

A recent development of NKRL concerns the possibility of using the two above modalities of inference *in an 'integrated' way*, see [9: 216-234]. More exactly, *it is possible to make use of 'transformations' when InferenceEngine is working in the 'hypothesis' environment*. This means that, whenever a search pattern p_i is derived from the *'condition schema' of a hypothesis* to implement a step of the reasoning process, we can use it '*as it is*' – i.e., in conformity with its 'father' condition schema – but also in *a 'transformed' form if the appropriate transformation rules exist*. The advantages are essentially that i) a hypothesis deemed to fail *can now continue* if a

transformed p_i is able to find an unification within the knowledge base, getting then new values for the hypothesis variables; ii) this strategy allows us to explore in a systematic ways all the possible *implicit* relationships among the data in the base.

4 Conclusion

NKRL deals with the representation and management of 'elementary' and 'complex' events by making use of *n*-ary and second order knowledge representation structures. One of its main characteristics concerns the addition of an *ontology of events* to the usual *ontology of concepts*. Its inference solutions employ advanced causal- and analogical-based reasoning techniques to deal with the events and their relationships.

NKRL is also a fully operational environment, implemented in two versions (file-oriented and Oracle-based) and developed thanks to several European projects. Many successful applications in many different domains (from 'terrorism' to the 'corporate', 'cultural heritage' and 'legal' domains, to the management of 'storyboards/historians' for the gas/oil industry…) have proved the practical utility of this tool.

References

1. Kowalski, R.A., Sergot, M.J.: A Logic-Based Calculus of Events. New Generation Computing 4, pp. 67--95 (1986)
2. Kaneiwa, K., Iwazume, M., Fukuda, K.: An Upper Ontology for Event Classifications and Relations. In: Proceedings of the Twentieth Australian Joint Conference on Artificial Intelligence. LNCS, vol. 4830, pp. 394--403. Springer, Heidelberg (2007)
3. Lombardo, V., Damiano, R.: An Intelligent Tool for Narrative-Based Video Annotation and Editing. In: Proceedings of the 2010 International Conference on Complex, Intelligent and Software Intensive Systems, pp. 706--711. IEEE Computer Society Press, Los Alamitos, CA (2010)
4. Zarzhitsky, D., Schlegel, M., Decker, A., Pack, D.: An Event-Driven Software Architecture for Multiple Unmanned Aerial Vehicles to Cooperatively Locate Mobile Targets. In: Optimization and Cooperative Control Strategies. LNCIS, vol. 381, pp. 399--318. Springer, Heidelberg (2009)
5. El Kharbili, M., Stojanovic, N.: Semantic Event-Based Decision Management in Compliance Management for Business Processes. In: Intelligent Complex Event Processing, Papers from the AAAI Spring Symposium, pp. 35--40. AAAI Press, Menlo Park, CA (2009)
6. Zarri, G.P.: Integrating the Two Main Inference Modes of NKRL, Transformations and Hypotheses. Journal on Data Semantics (JoDS) 4, pp. 304--340 (2005)
7. Zarri, G.P.: Modeling and Advanced Exploitation of eChronicle 'Narrative' Information. In: Proceedings of the Workshops of the 22nd International Conference on Data Engineering, ICDEW'06 Workshop on eChronicles, pp. 110--119. IEEE Computer Society Press, Los Alamitos, CA (2006)
8. Zarri, G.P.: Representation and Processing of Complex Events. In: Intelligent Complex Event Processing, Papers from the AAAI Spring Symposium, pp. 101--106. AAAI Press, Menlo Park, CA (2009)
9. Zarri, G.P.: Representation and Management of Narrative Information, Theoretical Principles and Implementation. Springer, London (2009)
10. Bal, M.: Narratology: Introduction to the Theory of Narrative, 2d ed. University of Toronto Press, Toronto (1997)
11. Jahn, M.: Narratology: A Guide to the Theory of Narrative (version 1.8). English Department of the Cologne University, http://www.uni-koeln.de/~ame02/pppn.htm (2005).
12. Bakhtin, M.M.: Forms of Time and of Chronotope in the Novel. In: The Dialogic Imagination: Four Essays, translated from Russian, pp. 84--258. University of Texas Press, Austin, TX (1982)
13. Propp, V.: Morphology of the Folktale, translated from the Russian by Scott, L., 2nd ed. University of Texas Press, Austin, TX (1968)

14. Kim, J.: Supervenience and Mind: Selected Philosophical Essays. University Press, Cambridge (1993)
15. Kim, J.: Events as Property Exemplifications. In: Events (International Research Library of Philosophy, 15), pp. 117--135. Dartmouth Publishing, Aldershot (1996)
16. Davidson, D.: Causal Relations. The Journal of Philosophy 64, pp. 691--703 (1967)
17. Davidson, D.: The Logical Form of Action Sentences. In: The Logic of Decision and Action, pp. 81--95. University Press, Pittsburgh, PA (1967)
18. Higginbotham, J.: On Semantics. Linguistic Inquiry 16, pp. 547--593 (1985)
19. Higginbotham, J.: On Events in Linguistic Semantics. In: Speaking of Events, pp. 49--79. University Press, Oxford (2000)
20. Parson, T.: Events in the Semantics of English. The MIT Press, Cambridge, MA (1990)
21. Saggion, H., Cunningham, H., Bontcheva, K., Maynard, D., Hamza, O, Wilks, Y.: Multimedia Indexing Through Multi-Source and Multi-Language Information Extraction: The MUMIS Project. Data & Knowledge Engineering 48, pp. pp. 247--264 (2004)
22. Manola, F., and Miller, E.: RDF Primer, W3C Recommendation 10 February 2004. W3C, http://www.w3.org/TR/rdf-primer/ (2004)
23. Bechhofer, S., van Harmelen, F., Hendler, J., Horrocks, I., McGuinness, D.L., Patel-Schneider, P.F., and Stein, L.A., eds.: OWL Web Ontology Language Reference, W3C Recommendation 10 February 2004. W3C, http://www.w3.org/TR/owl-ref/ (2004)
24. Hitzler, P., Krötzsch, M., Parsia, B., Patel-Schneider, P.F., and Rudolph, S., eds.: OWL 2 Web Ontology Language Primer, W3C Recommendation 27 October 2009. W3C, http://www.w3.org/TR/owl2-primer/ (2009)
25. Zarri, G.P.: An *n*-ary Language for Representing Narrative Information on the Web. In: SWAP 2005, Semantic Web Applications and Perspectives, Proceedings of the 2nd Italian Semantic Web Workshop. CEUR, http://sunsite.informatik.rwth-aachen.de/Publications/CEUR-WS/Vol-166/63.pdf (2005)
26. Hoekstra, R., Liem, J., Bredeweg, B., and Breuker, J.: Requirements for Representing Situations. In: Proceedings of the OWLED*06 Workshop on OWL: Experiences and Directions. CEUR, http://SunSITE.Informatik.RWTH-Aachen.DE/Publications/CEUR-WS/Vol-216/submission_4.pdf (2006)
27. Mizoguchi R., Sunagawa E., Kozaki K., and Kitamura Y.: A Model of Roles within an Ontology Development Tool: Hozo. Journal of Applied Ontology 2, pp. 159--179 (2007)
28. Salguero, A.G., Delgado, C., Araque, F. : Easing the Definition of N-Ary Relations for Supporting Spatio-Temporal Models in OWL. In: Computer Aided Systems Theory, 12th International Conference, EUROCAST 2009. LNCS, vol. 5717, pp. 271--278. Springer, Heidelberg (2009)
29. Liu, W., Liu, Z., Fu, J., Hu, R., Zhong, Z.: Extending OWL for Modeling Event-oriented Ontology. In: Proceedings of the 2010 International Conference on Complex, Intelligent and Software Intensive Systems, pp. 581--586. IEEE Computer Society Press, Los Alamitos, CA (2010)
30. Zarri, G.P.: A Conceptual Model for Capturing and Reusing Knowledge in Business-Oriented Domains. In: Industrial Knowledge Management: A Micro-Level Approach, pp. 37—53. Springer, London (2000)
31. Gangemi, A., and Mika, P.: Understanding the Semantic Web through Descriptions and Situations. In: Proceedings of the CoopIS, DOA, and ODBASE03 OTM Confederated International Conferences. LNCS, vol. 2888, pp. 689--706. Springer, Heidelberg (2003)
32. Gangemi, A., Guarino, N., Masolo, C., Oltramari, A., and Schneider, L.: Sweetening Ontologies with DOLCE. In: Knowledge Engineering and Knowledge Management, Ontologies and the Semantic Web – Proceedings of EKAW'2002. LNCS, vol. 2473, pp. 166--191. Springer, Heidelberg (2002)
33. Scherp, A., Franz, T., Saathoff, C., and Staab, S.: F – A Model of Events Based on the Foundational Ontology DOLCE+DnS Ultralite. In: Proceedings of the Fifth International Conference on Knowledge Capture, K-CAP '09, pp. 137--144. ACM Press, New York (2009)
34. Mani, I., and Pustejovsky, J.: Temporal Discourse Models for Narrative Structure. In: Proceedings of the ACL Workshop on Discourse Annotation, pp. 57--64. Association for Computational Linguistics, East Stroudsburg, PA (2004)
35. Noy, F.N., Fergerson, R.W., and Musen, M.A.: The Knowledge Model of Protégé-2000: Combining Interoperability and Flexibility. In: Knowledge Acquisition, Modeling, and Management, Proceedings of EKAW'2000, pp. 17--32. Springer, Heidelberg (2000)
36. Bertino, E., Catania, B., and Zarri, G.P.: Intelligent Database Systems. Addison-Wesley and ACM Press, London (2001)
37. Zarri, G.P.: Creation and Management of a Conceptual Knowledge Base in an Industrial Domain. In: Proceedings of the 2009 International Conference on Knowledge Engineering, KEOD'09, pp. 214--219. Escola Superior de Tecnologia (INSTICC), Setubal (2009)

Media Aggregation via Events

Fausto Giunchiglia, Pierre Andrews, Gaia Trecarichi, and Ronald Chenu-Abente

University of Trento, Italy {fausto,andrews,gtrecari,chenu}@disi.unitn.it

Abstract. Events have been recognised as important metadata to fill the semantic gap between our experience of the world represented in media and its conceptualization. In this paper, we argue that, once event metadata can be extracted, there remains a gap between different users conceptualizations. We then show how a compositional event model can mitigate such a social semantic gap through higher level descriptions of events where an agreement can be reached. In turn, this enables semantic services which improve event-centric search and navigation of shared media.

1 Introduction

With the increase of information and media streams available to us, everyday's tasks such as searching and relating relevant data have become intractable. One of the recognised causes for this issue is the semantic gap existing between our *conceptualizations* of the world, usually expressed using language or other high level abstractions, and our *experience of the world*, whose most direct physical representation is kept in media. In fact, even if automatic image processing algorithms can help by extracting high level concepts from media (e.g., who is present in a photo), they still miss the general semantics of the experience memorized by media such as the context in which a photo was taken or what it means for the user (e.g., the feelings and impressions of what happened when a photo was taken). Such experiential aspect cannot yet be extracted automatically and thus a number of solutions and services are being proposed to tame the incoming streams of data. To this purpose, event models have been proposed to help in the extraction and indexing of event information within data streams.

However, once a high level representation of an event is extracted or manually provided in the local context of a single user, there is no guarantee that this conceptualization will be understood by other users. We believe that there is a second gap, the *social semantic gap* between a local user experience conceptualization and what other users might understand of this conceptualization.

In this paper, we propose a general event model that we believe helpful in aligning different local event representations and show how it can be applied to the issue of media management. In Section 2, we provide a motivating example for our event model. In Section 3, we introduce in more details the issue of social semantic gap, while in Sections 4, 5 and 6 we describe in more detail this new model. Section 7 shows how it can be applied to model experiences and media to

help the sharing of such media with better metadata. Finally, Section 8 relates our work with state-of-the-art event models.

2 Motivating Example

Danda has just returned from a tour in the Italian region of Trentino with her friend[1]. She collected lots of material (e.g., digital photos, diaries, videos) and now wants to organise it digitally to revisit it later on and share her trip memories with her friends. With the current Web 2.0 technologies, she can rely on blogs, video sharing websites (e.g., YouTube) and online photo management software (e.g., Flickr, Virgilio Foto Album), to store and share the material. Since she likes writing, she opts to communicate the experience through her blog and thus dedicates some blog posts to describe the three days spent in Val di Non, the locality visited. In the first blog post titled "My journey diary - 11 Aug 2008", she describes the things that happened during the first day of the trip: the journey from Rimini to Trento by train, the one from Trento to Cles by the local railway Trento-Malé, the nice chats she had with the owner of the B&B during the journey from Cles to the B&B located near Revó, a small village close to the Santa Giustina lake, and so forth. The second day is described in a second blog post providing detailed descriptions of the breakfast she had in the B&B and of the itinerary followed to go to the Tovel lake. Finally, in two other blog posts, she describes the visits made to the Novella river park and the Tret waterfall and the concert of the Ramadas band attended during the third day. In all posts, images illustrate snippets of text to enhance the visual impact of the blog. However, the full gallery of pictures is maintained in a separate online photo management system which is pointed to by a link included in the blog post. An excerpt of the first blog post follows:

> "At Trento we wait for another train, this time on a local railway, the Trento-Malé. Our last stop is Cles from where buses depart for various villages, including Revó; however, the kind owner of the B&B waits for us there and gives us a lift by car..."

The way Danda discloses her trip experience allows her to fix her memory of the journey and to make friends and interested bloggers aware of it. It is important to note that Danda is ultimately interested in describing and sharing the events happened to her rather than just sharing a bunch of photos, these last being rather supporting material to give evidence of her experience and embellish the story. She describes the three-days trip by splitting it into days – or part of days – each corresponding to few significant events, which, in this case, are mostly visits to natural locations. The events range from small-scale ones (e.g., the breakfast, the move from Trento to Cles) to more large-scale, composite ones (e.g., the whole trip, the visit to a natural location, the concert) and span several types

[1] Our running example is based on a real "blog" story found at: `http://dandaworld.blogspot.com/2008/09/appunti-di-viaggio-my-journey-diary-11.html`.

(e.g., a visit, a move from one place to another, breakfast, walk, conversation, concert). Also, all their descriptions mention different entities such as locations (e.g., Trento, Cles, Tovel, Novella river park, Tret waterfall), time periods (e.g., 28th August 2008), people or group of people (e.g., trip companion, B&B owner, Ramadas band), and others (e.g., "Trento-Malé").

If an user-oriented personal media management system existed that allowed for an easy way to describe complex events, Danda would be supported in reliving the "trip event" by recalling salient events at the desired level of details, the persons met, and the places she visited in a more active, experiential manner. Moreover, she would experience pre- or post-trip visits by knowing more about the locations (e.g., facts and media), some co-located events and related stories. The framework we envision supports the well-understood need for event-centric media management systems.

In addition, by having a structured description of the media and event metadata, the events described by Diana can be matched with other users' events and with global definitions of events. We believe, this will reduce the social semantic gap between local descriptions and global understanding of events.

3 The Social Semantic Gap

There has been a long stream of research in media processing and media management to fill the semantic gap between what can be seen in a media (the person in a photo for instance) and what it actually represent as an experience for the person that created the media. This can be partly solved by adding event metadata to the media to help understand the context in which they were created. However, in many cases, this event information is created locally, either semi-automatically or manually, at a user level. The particular descriptions of the events at a local level, even if abstracting the content of the media, can still be different from their shared global conceptualization. As shown in Fig. 1, two different users might have a different high level description of a media, and thus there is still a gap between the different personal users' representations of their experiences.

In professional applications, like news media management for instance, there is a top-down agreement on the existing events (e.g., football championship, political conflicts) and thus the gap between the meaning of the events represented in the media and the ones described in the final products (e.g., newspaper articles) composed with these media is kept under control. For instance, everyone shares the understanding of what the "New Year Celebration" is and what experience the media used in the news for that event might represent. However, in the personal application sphere, any of the user's personal experiences can be transformed in an event and there is thus no widespread agreement about events and what they might represent for each single user. A user can create an event about her "Family Holiday" and understand the experiences that are represented in the media for this event, but, outside of her local context, such experiences and events are not meaningful anymore.

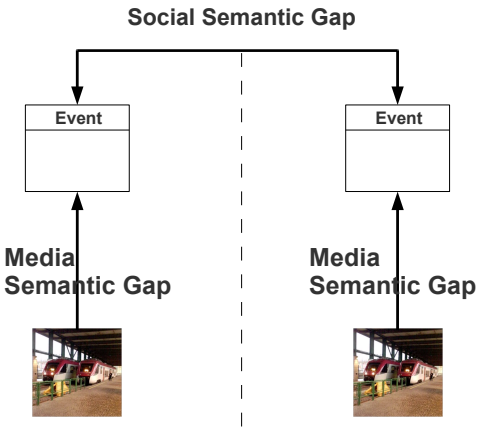

Fig. 1. Media and Social Semantic Gap

There can still be a social agreement between users that sit between the Global and Local contexts when the event is shared by a small group of participants. For instance, all the people attending a "Friends' night-out" event will understand the experiences in the media attached to this event, but this understanding will be lost to anyone outside of this group. Thus the events that can be extracted – or provided manually by the user – for the personal user's media are built bottom-up and are difficult to share: there is still a large *social semantic gap* between the local descriptions and their perception by the community.

We believe that the main cause of such a social semantic gap is due to the lack of aggregable metadata for these media. This makes it difficult to align single event descriptions from different users. We propose to solve this issue by modeling the events structural aspect [1] and provide a compositional event model where fine grained events can be composed into higher level events in order to provide a middle ground between top-down event agreement and bottom-up event creation. In fact, if the media are stored in a rooted structure of events, it will be easier to match between diverse local events into a global consensus. For instance, if a set of media is associated to a personal "Toast" event, it is impossible to know if this is a toast during a normal dinner, during a graduation ceremony or at a wedding reception. If this event is provided in the context of subsuming events, as illustrated in Fig. 2, it is then easier to know its semantics.

4 Entities

In our example, we can first notice that Danda refers to a number of "objects" in the real world, such as the places she visited or the people she met. There is a need to represent all these entities and their metadata into a model for managing the media relevant to the events she is describing. [2] proposes a unique entity space to store such resources and we follow a similar entity centric model to

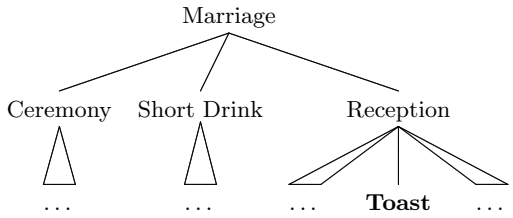

Fig. 2. Wedding Event Organisation

provide a uniform representation of objects in the real and virtual world. In our model, an entity En is described by its metadata and associated services:

$$En =< id, type, \boldsymbol{Attr}, \boldsymbol{Rel}, \boldsymbol{S} >$$

Where: 1. *id*, is a unique identifier (e.g., an URI); 2. *type*, is the type of entity, that is, the category to which it belongs to (e.g., the entity "Danda" is of type `Person`); 3. ***Attr***, is a set of *attributes* composed of pairs $attr =< attr\_name, attr\_value >$ describing the properties (e.g., "date of birth") of that particular entity; 4. ***Rel***, is a set of *relational attributes* composed of pairs $rel =< rel\_name, rel\_value >$ describing the entity's relations (e.g., "friendOf") with other entities; 5. $\boldsymbol{S}$, is a set of *services* that can be leveraged on that specific entity; for example, a service "send email" can be enabled on the `Person` entity type (etype).

An important aspect of our model is that both attributes and relationships can be further defined by *meta-attributes*. For example, attributes like "job position" or relations such as "friendOf" are provided with metadata of their own, for example to describe the time period when these are valid or the circumstances (i.e., the events) that made them true.

Another interesting aspect is the *lattice* of etypes that is encoded in the *type* property of the entity. The specific type (e.g., `Person`, `Location`, `Event`) to which the entity belongs to is used to infer its possible attributes and services. Moreover, the hierarchy defined by the lattice allows to easily define new derived etypes by just inheriting the metadata and available services of parent etypes. For example, the new etype `Author` inherits both attributes (e.g., "name", "date of birth") and services (e.g., "send email") from the etype `Person` but extends them with more specific ones (e.g., "affiliation", "get h-index").

In our running example, we can identify a number of entities in Danda's recollection that can be represented in our model and used for future retrieval and reasoning: Danda participates in a "concert", which can be represented by an `Event` entity; "the owner of the B&B" is a `Person` entity; and specific `Location` entities are described, such as the "Trento" or the "Novella river park".

The following sections describe the less trivial entities used to construct a workable model of the events in Danda's blog post. The data structures discussed hereafter can be modeled as a set of attributes, relations and meta-attributes

as formalised in the general Entity model but we provide a higher-level view for clearer reading.

5 Events

Events, unlike facts, are closely linked to their spatio-temporal collocation and, also, to the things constituting their subject (e.g., a sparrow in the event "a sparrow falls"). In addition, unlike objects, they have clear temporal boundaries but fuzzy spatial boundaries and they have a time-span [3]. Moreover, they are usually provided by descriptions and they may be composed of sub-events which are temporally, spatially and causally connected [4].

In our model, we assume that "local" events are created by users and their structuring and descriptions are thus subjective. An event entity Ev is modeled as follows:

$$Ev =< evid, t, \boldsymbol{LEv}, Cx > \tag{1}$$

Where the elements of the tuple are:

- *evid*, the unique identifier of an event;
- t, the temporal collocation of an event, i.e., the time interval described by the event; it can be either a specific time interval marked off by the initial and final instants – e.g., "2009:01:14:10:00" to represent the 14th January 2009 at 10am – or a generic period of time where the temporal delimitations are not specified; also, the information on the date can be incomplete (e.g, "2009:01:::"), or relative (e.g., "the day before Christmas"), and the time interval may not be continuous as, for instance, it would happen for a "Champions League" event.
- $\boldsymbol{LEv}$ represents a set of linked events and is described in more details in the following section.
- Cx, the *event context*, this being regarded as a distinguishing feature of event entities. As demonstrated in [5, 6], this context is useful for localised reasoning. The event context is represented as:

$$Cx =< l, type, \boldsymbol{Pc} > \tag{2}$$

Where:

- l, defines the spatial collocation of an event. It identifies a "geographical entity" such as a geopolitical entity, a natural body (e.g., mountain, river, lake), or a man-made infrastructure (e.g., building, stretch of road). For example, consider our running example where Danda moved around the north of Italy, the location of this "transfer" event can be modeled by defining two geographical (point) locations (Rimini and Trento) that form the stretch of train track between Rimini and Trento.
  Note that, although the spatial collocation could be objectively defined, the participant's perception of it, that is represented in the context, is itself subjective, e.g, in terms of the actual extension of the location itself ("Trento" vs. "Trentino Region").

- *type*, is the type of event (e.g., conference, trip, visit, concert);
- ***Pc***, the event's participants is a set of relations to other entities. The corresponding entity values could belong to `Person`, `Organization` as well as to non-agentive etypes. Each relation to these participating entities is annotated with *meta-attributes* describing aspects of each entity which are only relevant in the current event's context. This includes the role of the entity in such an event, i.e., the modality of its participation in the event: for instance, a person can be a professor in a graduation event but is then a mother in the event describing the birthday of her daughter. Our vision of role is in line with the one given in [7], where one of the key features of a role is that of being linked to the notion of context – the event's context in our model.
  In addition, an entity participating to a given event could be described by properties valid only within the event context: for example, a temporal attribute such as "jacketColour" is only attached to the relation between a particular event and the entity. Note that, since we regard events as subjectively perceived entities, the above mentioned properties are meant to be objective (e.g., "jacket colour") as well as subjective (e.g., "personality"). As for the attributes, relations defining the metadata of an event's participants can also be relevant only to the event's context; for instance, relations such as "girlfriendOf" or "near".

6 Event Compositionality

Danda's trip has several events that can be identified and captured through the previously given event definition. For example, the transfer from Rimini to Trento can be represented as:

$$Transfer(\text{Rimini}, \text{Trento}) =< evid01, \text{“11/08/2008 on the early morning”}, < \text{“Railroad between Rimini and Trento”}, \text{“train transfer”}, \{\text{Danda}, \text{Danda's friend}\} >, \emptyset >$$

while the transfer from Trento to Revó can be represented as:

$$Transfer(\text{Trento}, \text{Revó}) =< evid02, \text{“11/08/2008 on the late morning”}, < \text{“Railroad between Trento and Revó”}, \text{“train transfer”}, \{\text{Danda}, \text{Danda's friend}\} >, \emptyset >$$

These two events can be aggregated to define a more general *Departure* event representing the journey that would include both *Transfer(Rimini,Trento)* and *Transfer(Trento,Revó)*:

$$Departure =< evid03, \text{“11/08/2008 morning”}, < \text{“Railroad between Rimini and Revó”}, \text{“train transfer”}, \{\text{Danda}, \text{Danda's friend}\} >, \{evid01, evid02\} >$$

The ***LEv*** from the event *Departure* contains references to the events it is aggregating. The running example, as a whole, can be represented as a single complex entity *Trip* as illustrated in Fig. 3.

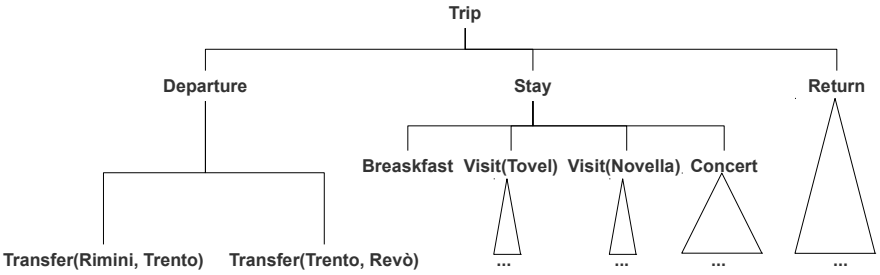

Fig. 3. Event Structure for the Danda's trip

Note that, in Fig. 3, the event that represents the whole *Trip*, is subdivided in *Departure*, *Stay* and *Return* (which refer to the going journey, the stay period and the return journey respectively). Furthermore, each of these sub-events is, in turn, subdivided in other sub-events. This tree-like compositional structure is enabled by the use of the ***LEv*** component from the event definition in (1).

As explained before, ***LEv*** represents a set of linked events that are parts of the event to which that ***LEv*** belongs. However, to keep ***LEv*** as useful as possible for its complex event modeling purposes, the following restrictions are applied to it:

Restrictions on time t, the time duration defined for a complex entity CEv, must subsume the time duration of all of its sub-events pointed by ***LEv***. That is, if we have a function $\texttt{time}(Ev) \rightarrow t$:

$$\text{Given } CEv =< evid, t, Cx, \boldsymbol{LEv} >, \forall e \in \boldsymbol{LEv}, \texttt{time}(e) \sqsubseteq t$$

By enforcing the previous restriction, all the individual time periods involved in the children events, are guaranteed to be subsumed in the time period of the parent complex event. This enables the representation of a complex event, like the one from our running example, in a timeline as shown in Fig. 4.
Note how in Fig. 4 all events comply with this rule. For example, the event *Stay* spans from the first to the third day and is within this period that its children events (*Breakfast*, *Visit(Tovel)*, *Visit(Novella)* and *Concert*) take place. Furthermore, note that the whole time period of *Stay* is not entirely covered by sub-events (e.g., a small period of time exists between the end of the event *Visit(Novella)* and the start of the event *Concert*). These blanks correspond to unspecified events in Danda's trip such as, for example, the transfer between visits where she had no memorable experiences.

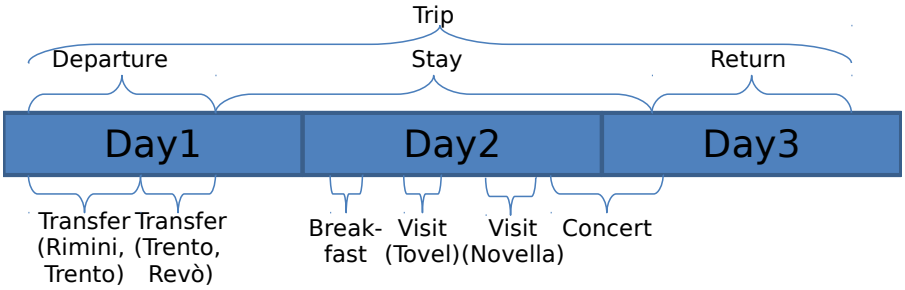

Fig. 4. Running example as a timeline.

Restriction on context the context Cx defined for the complex entity CEv must subsume the context metadata of all of its sub-events pointed by $\boldsymbol{LEv}$. That is, given (1) and (2), if we have the functions:

$$\texttt{location}(Cs) \rightarrow l$$
$$\texttt{location}(Ev) \rightarrow \texttt{location}(Cx)$$
$$\texttt{participants}(Cx) \rightarrow \boldsymbol{Pc}$$
$$\texttt{participants}(Ev) \rightarrow \texttt{participants}(Cx)$$

Given $CEv =< evid, t, \boldsymbol{LEv} >, Cx$

$$\forall e \in \boldsymbol{LEv}, \texttt{location}(e) \sqsubseteq \texttt{location}(Cx)$$
$$\forall e \in \boldsymbol{LEv}, \texttt{participants}(e) \subset \texttt{participants}(Cx)$$

By enforcing the previous restriction, all the geographic locations from the children events are guaranteed to be subsumed by the parent's location and all the participants from the sub-events are guaranteed to be included in their parent's set of participants.
From our previous examples, it is clear that the location for the event *Transfer(Rimini,Trento)* is the railroad between the cities Rimini and Trento; furthermore, the location for the event *Transfer(Trento,Revó)* is the railroad between the cities Trento and Revó. Applying the location subsumption restriction, the location for the parent event *Departure* would be the train road between Rimini and Revó or more generally the entity for North of Italy (both of which subsume the locations from the sub-events). Likewise, the participants from *Breakfast* and *Visit(Tovel)* would also be included in the set of participants of its parent event *Stay*.

Restriction on linked events an event cannot be included in $\boldsymbol{LEv}$ if doing so would cause the creation of a loop in the events structure. Let $\texttt{connection}(Start, End)$ be a function that returns a sequence of events that, through their $\boldsymbol{LEv}$ components, define a directed path from the *Start* event to the *End* event (or $\emptyset$ if no such sequence exists). Then, the restriction can be

expressed as:

$$CEv =< evid, t, \boldsymbol{LEv}, Cx >,\ \texttt{connection}(CEv, CEv) = \emptyset$$

This restriction is introduced to avoid the conceptual problems that would arise from an event being its own predecessor, directly or through other intermediate events.

The composition of events presented in this section has the following advantages:

- *Avoid repetition of information*: if there is a particular information that applies to all of the children of a complex event, instead of repeating the content on each of the sub-events, this information can be included directly at the event that is aggregating them. For example, thanks to this, it is not necessary to describe the weather in both the *Transfer(Rimini,Trento)* and the *Transfer(Trento,Revó)* events. If the weather did not change between these two events, the details of the weather can be included in their aggregating event *Departure* and, through compositionality, this information will apply to all of its children.
- *Capture information emerging from the aggregation*: there may exist information that emerges from the composition itself and is not part of any of its individual sub-components. For example, Danda could describe the *Concert* event as being "long and tiring" but each of its sub-events may not have these properties individually. The "long and tiring" description would then only apply to the aggregation of these individual events into the *Concert* event.
- *Capture information from unspecified sub-events*: as seen in Fig. 4, there may exist some blanks between events at high granularity levels. A lower granularity or parent event can then be used to capture information belonging to these blanks: for example, suppose Danda wanted to add a photo she took right after having her breakfast on the 2nd Day (*Breakfast* event) but before her visit to Tovel (*Visit(Tovel)* event); instead of adding a new event only for that photo, she could just include it in the *Stay* event.

In the following section we show an example of how this structured metadata of the events and the media that can be attached to them can be used to fill the semantic gap.

7 Event-Centric Media Management

In the previous sections, we presented a general event model where an event is independent from media. In fact, an event can exist totally independently from media in many applications. However, in this paper we are interested in how such an event model can be used to move from a media-centric management to a different metaphor where events are of importance for organizing media,

which is the goal of a number of projects such as GLOCAL[2], PRONTO[3] and EventMedia[4]. This is supported by [1] that prompts for a common event model for media management. Our approach is to separate clearly the event metadata and the experiential aspect of this event, which depends on the user describing the event and the intended audience. Thus, we introduce a new entity in our model to store a particular description of an experience as the relation between an event and the media describing it:

$$Exp =< Ev, \boldsymbol{M} >$$

where Ev is a relation to the event in question and $\boldsymbol{M}$ is a set of relations to media describing the user's experience of this event.

In this context, the creation and structuring into complex events proves useful when a "story" about that event's experience has to be told. For example, in our example, Danda goes into the higher granularity events for the part of the story that her audience will favor. Conversely, she also chose to stay at a lower level of details when describing events that might be of less importance for that particular audience. However, for her own use when searching for media or when telling stories to other audiences about her experience, she is still interested in keeping as most details as possible in the events structure and metadata.

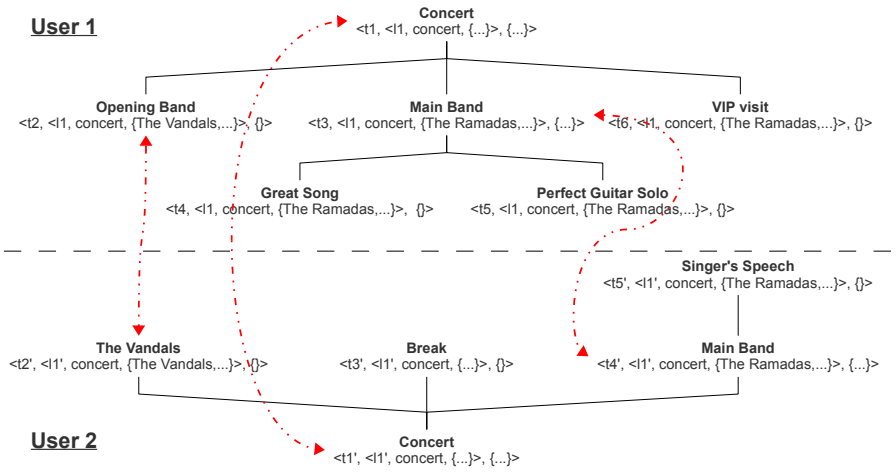

Fig. 5. Two Local Structures for the Same Concert Event

A second application of this granular description of events in media management is the support provided when sharing an experience with a community

[2] http://www.glocal-project.eu/
[3] http://www.ict-pronto.org/
[4] http://eventmedia.cwi.nl/

(i.e., going from a local context to a shared or global context as discussed in Section 3). If we consider the "concert" event in our running example, more than one user will have attended the concert and annotated media of this concert with subjective event information; we could thus have the two event sub-structures shown in Fig. 5. The two users have a different experience of the concert and have represented it as a set of different sub-events. If only the lower level of granularity was available, it would be difficult to say that the media describing the "Singer's Speech" and the "Great Song" are related. However, by providing information about the "concert" higher level event, general metadata is made available to help match between users local events as the subjective description of the time ($t1$ and $t1'$) and location ($l1$ and $l1'$) are more likely to overlap. The media of both users can be related by matching as much events as possible in a top-down – general to specific – fashion to reduce the social semantic gap. In particular, the sub-structure of events is slightly different, but by comparing the metadata of each event, the events "Opening Band" and "The Vandals", as well as "Main Band" and "The Ramadas", can be matched for instance with an ontology matching algorithm [8]. In addition, once the matching is done, the management system can propose to the *User 1* to add the "Break" sub-event to her own event structure and thus organise her local media in a more refined manner. After that, the media can be reclassified automatically within the sub-event, for instance with the get-specific algorithm [9]. Thus, the users will be able to share, search and navigate new media of the events they have experienced more easily.

8 Related Work

Research fields spanning from Semantic Web [10, 2] to Information Extraction [11, 12] to Digital Libraries [13] have recognised the key role played by the entities and their linking relationships.

In [10], entities are seen as atomic objects of reference and reasoning for Semantic Web applications which are described by a general conceptual model and categorized into types derived from a user-study. This work is part of the OKKAM project[5], which proposes a framework where entities are assigned with a globally unique identifier to ease data integration and the development of innovative "entity-centric" applications [2]. Categories for entities are also defined in [11, 12], together with guidelines for the accomplishment of named entity recognition and relation extraction tasks. The issue of linking together different kinds of entities in the Digital Library domain is addressed by Buckland [13], who advocates the need for a metadata infrastructure able to interconnect place-name gazetteers, biographical dictionaries, time period and subject indexes.

Furthermore, the concept of the "Web of Things[6]" (see [14]) has recently emerged and efforts such as the Linking Open Data Initiative[7] entered the scene

[5] http://fp7.okkam.org

[6] http://ercim-news.ercim.eu/content/view/343/536/

[7] http://linkeddata.org/

to address its principles. The role played by linked data for supporting users in tasks beyond the simple fact-finding and question answering (e.g., finding connections among people, places and events) is analyzed by looking at specific prototypes in [15].

In regards to the conceptual representation of events, research work exists on generic models as well as models tailored to domain-specific events (e.g., journalistic, historical, cultural-heritage, multimedia events): the generic E-model [16] is extended to enable "event-centric" rather than "media-centric" media management systems [17]; on the same line, the Eventory media repository [18], the Medi*AE*ther system [19] and a multimodal event browsing tool [20] are proposed. The F Event Model [21] proposes a formal model that, together with standard properties, supports mereological, causal, correlation relationships and interpretations of arbitrary events. The Event Ontology[8], developed as part of a music ontology framework, supports music events (e.g., compositions, recordings, performances) but is not tied to such domain. With a journalistic perspective, EventML[9] provides an XML schema for exchanging news events among news agencies. The CIDOC ontology [22] and the CultureSampo approach (see [23]) aims to solve interoperability issues between metadata standards for heritage resources. In [24], most of the above-mentioned models are compared and an event ontology is proposed to fulfill the vision of a Linked Data Event Model[10]. The model is purposely kept minimal to capture the well-understood "when", "where", "what" and "who" facets of historical non-composite events; aspects defined as more interpretative dimensions are excluded. In this respect, our model differs from others since user-driven contextual metadata make events as always being subjective entities. Furthermore, by means of the *experience* entity we define relationships between events and media. This is also allowed in Eventory [18] where such relationships are made explicit through ad-hoc interfaces for events, media and connection creation; however, we foresee a system where these types of relationships are deduced from the way the user describes her experience rather than being explicitly created.

For what specifically concerns the event's structural relationships – i.e., their compositionality – Rafatirad et al. [25] design "subevent-of" relationships forming directed acyclic graphs over events and implement composition operators allowing to aggregate the spatial and temporal attributes of composite events from atomic events for which these attributes are known; Singh et al. [26] regard the information needed to model compound events as semi-structured (dynamic) data and hence use XML to manage them: for example, a "group meeting" event is divided into compound events (e.g., "Introduction", "Presentation", "Discussion") and simple events (e.g., "ask_question", "answer_question"). The XML schema used to describe a compound event includes an "how" element which model its process, i.e., the simple events composing it. A hierarchical model to represent events at different granularities has been recently proposed in the domain of

[8] http://motools.sf.net/event

[9] http://iptc.org/

[10] http://linkedevents.org/ontology/

multimedia observation systems [27]; here, "transient" events are detected from sensors datastreams and clustered into "atomic" events, these last being in turn composed into "compound" event. The model is however application-specific and aimed at reducing the semantic gap rather than what we defined as social semantic gap. On the same line, van Hage et al. [28] present the Simple Event Model (SEM), applied in a Maritime Safety and Security use case to deduce simple behaviour events from sensor data. The model provides a minimal set of classes describing all the event's aspects, and add the notion of roles and the possibility to associate types to these classes thus to maintain the compatibility with external resources.

9 Conclusion

In this paper we present a general Event model to store the key metadata of an event. In particular, our model allows for the storage of complex subjective information relevant to the event's context.

We then show how such general model can be applied to the media management issue by introducing the *experience* entity that links an event with the media representing the user's experience of such an event. We believe that by clearly separating the event model from the experiential metadata of the event, this metadata is easier to use in heterogeneous applications.

In addition, by proposing a compositional model to represent the structural aspect of events, we allow for an easier alignment between users' personal descriptions of events and thus bridge the "social semantic gap" between different local representations of shared experiences.

Acknowledgement

The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement n248984 GLOCAL: Event-based Retrieval of Networked Media.

References

1. Westermann, U., Jain, R.: Toward a common event model for multimedia applications. IEEE MultiMedia **14** (2007) 19–29
2. Bouquet, P., Stoermer, H., Niederee, C., Mana, A.: Entity name system: The backbone of an open and scalable web of data. In: In Proc. of IEEE-ICSC. (2008) 554–561
3. Casati, R., Varzi, A.: Events. Stanford Encyclopedia of Philosophy. (2006)
4. Scheffler, U.: Events as shadowy entities. Konstanzer Berichte: Logik und Wissenschaftstheorie **28** (1992)
5. Giunchiglia, F.: Contextual reasoning. Epistemologia **16**(I Linguaggi e le Macchine) (1993) 345–364

6. Ghidini, C., Giunchiglia, F.: Local Models Semantics, or Contextual Reasoning = Locality + Compatibility. Artificial Intelligence **127** (2001) 221–259
7. Masolo, C., Vieu, L., Bottazzi, E., Catenacci, C., Ferrario, R., Gangemi, A., Guarino, N.: Social roles and their descriptions. In: Proc. of Knowledge Representation Workshop, AAAI Press (2004) 267–277
8. Giunchiglia, F., Yatskevich, M., Shvaiko, P.: Semantic matching: algorithms and implementation. In: JoDS, IX. (2007)
9. Giunchiglia, F., Zaihrayeu, I., Kharkevich, U.: Formalizing the get-specific document classification algorithm. In: ECDL. (2007) 26–37
10. Bazzanella, B., Chaudhry, J., Palpanas, T.: Towards a general entity representation model. SWAP (2008)
11. : Automatic content extraction english annotation guidelines for entities, version 6.6 2008.06.13. Technical report, Linguistic Data Consortium (2008)
12. : Automatic content extraction english annotation guidelines for relations, version 6.2 - 2008.04.28. Technical report, Linguistic Data Consortium (2008)
13. Buckland, M.K.: Description and search: metadata as infrastructure. Brazilian Journal of Information Science (2006)
14. Bizer, C., Heath, T., Berners-Lee, T.: Linked data - the story so far. IJSWIS, Special Issue on Linked Data (2009)
15. Hardman, L., Ossenbruggen, J.v., Troncy, R., Amin, A., Hildebrand, H.: Interactive information access on the web of data. In: WebSci'09: Society On-Line. (2009)
16. Westermann, U., Jain, R.: E - a generic event model for event-centric multimedia data management in echronicle applications. In: ICDEW, IEEE Computer Society (2006) 106
17. Scherp, A., Agaram, S., Jain, R.: Event-centric media management. In: SPIE. (2008)
18. Wang, X., Mamadgi, S., Thekdi, A., Kelliher, A., Sundaram, H.: Eventory - an event based media repository. In Semantic Computing. IEEE (2007)
19. Boll, S., Westermann, U.: Medither: an event space for context-aware multimedia experiences. In: ACM SIGMM workshop on Experiential telepresence. (2003) 21–30
20. Appan, P., Sundaram, H.: Networked multimedia event exploration. In: ACM Multimedia. (2004) 40–47
21. Scherp, A., Franz, T., Saathoff, C., Staab, S.: F–a model of events based on the foundational ontology dolce+dns ultralight. In: K-CAP'09. (2009) 137–144
22. Doerr, M., Ore, C., Stead, S.: The cidoc conceptual reference model - a new standard for knowledge sharing. In Conceptual modeling (2007) 51–56
23. Ruotsalo, T., Hyvnen, E.: An event-based approach for semantic metadata interoperability. In: ISWC. (2007) 407–420
24. Shaw, R., Troncy, R., Hardman, L.: Lode: Linking open descriptions of events. In: ASWC. (2009)
25. Rafatirad, S., Gupta, A., Jain, R.: Event composition operators: Eco. In: EiMM'09, ACM (2009) 65–72
26. Singh, R., Li, Z., Kim, P., Pack, D., Jain, R.: Event-based modeling and processing of digital media. In: CVDB'04. (2004)
27. Atrey, P.K.: A hierarchical model for representation of events in multimedia observation systems. In: EiMM'09, ACM (2009) 57–64
28. van Hage, W.R., Malaisé, V., de Vries, G., Schreiber, G., van Someren, M.: Combining ship trajectories and semantics with the simple event model (sem). In: EiMM'09. (2009)

What's on this evening? Designing User Support for Event-based Annotation and Exploration of Media

André Fialho[1], Raphaël Troncy[2], Lynda Hardman[1], Carsten Saathoff[3] and Ansgar Scherp[3]

[1] CWI, Amsterdam, The Netherlands, <firstname.lastname@cwi.nl>
[2] EURECOM, Sophia Antipolis, France, <raphael.troncy@eurecom.fr>
[3] WeST Institute, Koblenz, Germany, <lastname@uni-koblenz.de>

Abstract. We present an event-based approach for users to explore, annotate and share media. We are constructing a web-based environment that allows users to explore and select events, including discovering meaningful, surprising or entertaining connections among them. We build a knowledge base of events from event directories that will be linked to the Linked Open Data (LOD) cloud, in conjunction with event and media ontologies. The approach is user-driven and, having carried out initial user inquiries, we are designing interfaces that support user-identified tasks while exploring the connections between users, multimedia content and events.

1 Introduction

As with all developing technologies, it is difficult to identify novel user needs that can be satisfied with emerging semantic web technologies. At the same time, it is difficult to develop the technology in specific directions without knowing what users are likely to want to do with the technology. In previous work we have identified comparison search tasks that can be supported using a combination of thesaurus-based linked data search and a modular user interface design [1], and also historical print annotation tasks [5] that can be supported using a combination of existing RDF data sets, semantic search functionality and task-oriented user interface.

In the context of the Petamedia[4] Network of Excellence, we are exploring a similar method for designing an application that takes into account the "triple synergy" of users and their social networks, user-created content and metadata attached to this content in an application for supporting users in interacting with events. Events are a natural way for referring to any observable occurrence grouping persons, places, times and activities that can be described [10, 9]. Events are also observable experiences that are often documented by people

[4] http://www.petamedia.eu

through different media (e.g. videos and photos). We explore this intrinsic connection between media and experiences so that people can search and browse through content using a familiar event perspective.

While wishing to support such functionality, we are aware that websites already exist that provide interfaces to such functionality, e.g. eventful.com, upcoming.org, last.fm/events, and facebook.com/events to name a few. These services have sometimes overlap in terms of coverage of upcoming events and provide social networks features to support users in sharing and deciding upon attending events. However, the information about the events, the social connections and the representative media are all spread and locked in amongst these services providing limited event coverage and no interoperability of the description. Our goal is to aggregate these heterogeneous sources of information using linked data, so that we can explore the information with the flexibility and depth afforded by semantic web technologies. Furthermore, we will investigate the underlying connections between events to allow users to discover meaningful, entertaining or surprising relationships amongst them. We also use these connections as means of providing information and illustrations about future events, thus enhancing decision support.

The work reported here uses an explorative user-centered design approach, where users are asked about real-world tasks they would like to carry out, and then asked for their opinions on specific technologies that they are familiar with and how these might be used to support the tasks. This approach ensures making design decisions that contribute towards an efficient, effective and satisfying user experience. Section 2 describes the method and the results of this user study, and presents the requirements for an event-based system for discovering and sharing media. Section 3 describes the event and media ontologies developed to support the semantic description of events extracted from event directories. Section 4 explains the design rationale of the interfaces, and gives some interface mockups to illustrate the types of task support and expected functionality we will be providing in the coming months. Finally, we give our conclusions and outline future work in Section 5.

2 User Need Assessment

We follow a user-centered design process consisting of an assessment of user needs and insights, identified through interaction with potential users at different stages of development. Our research starts by identifying who the users are, their interests, their goals and which tasks need to be supported in order to achieve these goals. We collect this information to define a first set of requirements and identify prospective scenarios that illustrate the environment task scope and a first design concept. The steps that follow consist of iterative cycles of re-design and evaluation until a satisfactory design is reached.

2.1 Method

The first step of our research was done in order to collect potential end-user experiences, opinions and interests while discovering, attending and sharing events, and user insights about potential web-based technologies that support these activities. We collected this initial input through an exploratory study with 28 participants (11 females). Participants were mostly students and researchers with ages varying from 23 to 47 years old. The study was done through an on-line survey with 8 questions divided into 2 sections. The same topics were then presented in discussion sessions with two groups of master students totalizing 35 additional participants: One discussion was done with students (n=10) from an Interactive Multimedia Systems course and the other with students (n=25) from a Human-computer interaction for the Web course. The results from these discussions were used to validate the survey responses and to extend it with other collected insights.

The first half of the survey aimed at identifying participants' personal experiences and behaviors. It invited them to recall memorable previously attended events (e.g. festivals, conferences, concerts, art galleries, exhibitions, gatherings) and to share their opinions and experiences regarding: (1) how events are discovered; (2) characteristics that support deciding rather or not to attend to an event; (3) how the event experiences are registered and shared; and (4) meaningful, surprising or entertaining relationships amongst events.

On the second part of the survey, participants' were asked to share opinions regarding existing web technologies in the context of the aforementioned activities. To better address the triple synergy paradigm (Section 1), we explored the concept of merging event directories, media directories and social networks. With that in mind, we asked participants to share their opinions regarding: (1) the perceived benefits and drawbacks of event directories (e.g. Eventful); (2) enabled possibilities, benefits and drawbacks of combining media sharing websites (e.g. YouTube and Flickr) with event directories; (3) enabled possibilities, benefits and drawbacks of combining social networks (e.g. Facebook and Twitter) with event directories; and (4) suggestions regarding desired and useful features.

Answers obtained from the survey were analyzed through affinity diagramming. The process consists of iterative clustering cycles which allow organizing the collected ideas into common themes, thus allowing to identify the most common opinions for each raised subject. The results from this first exploratory study are described in the following section.

2.2 Results

In this section, we present a summary of the results of our user study. The summary contains main reported experiences, interests and opinions around event related activities.

Past experiences. Concerning participants' experiences when discovering events, the vast majority reported to find out about events through invitations and recommendations from friends and colleagues. Traditional media such as posters,

flyers, news articles and television ads seem to play a major role when discovering events. Social networks were also reported to be used, with specific reference to event posting and invitation features. More seldom participants use event directories (e.g. Livenation, local city event directories, Ticketmaster, last.fm) or participate in mailing lists, newsletters and forums to obtain updates. The use of search engines was reported, specifically when they knew what to look for. Moreover, participants also rely on previously attended events or venues as reference for finding new events. During the group discussions, participants seemed to rely more heavily on social networks in comparison to the survey responses.

When deciding whether or not to attend to an event, participants seem to prioritize background information. Location was often referred to for orientation and because of distance constraints. Price was commonly mentioned to allow identifying cost-benefit ratios and due to budget constraints. Time of the event was a main decision factor, followed by information about who else would be joining the event. and more specifically, which friends will attend. The content of the event itself (e.g. type, performer, topic) and subjective factors such as fun, relevance, interest, atmosphere, target audience and reputation were also mentioned. Students from the group discussions, preferred the event attendance ("who's joining?") and price constraints over all other characteristics.

Regarding how participants register the experience, they often take pictures for sharing after the event. Less commonly, participants record short videos. As for the how they share the information, they most commonly talk to others, describing their experience. Participants share the collected media directly (e.g. file transferring, showing on the mobile) or use media directories and social networks such as Facebook, Flickr or YouTube.

Concerning relationships between events, the most referred to characteristics that motivate participants to look into related events were the event categories (e.g. type of event, topic, genre). Another important factor was the event attendees, to find other events they would attend. This could refer to groups of people (i.e. target audience, users with similar interests), but most importantly, individuals in their social networks. Other main event characteristics also mentioned were: location, performers, organizers and time/duration. Lastly, future events from repeated events was also seen as a strong relationship.

Existing technologies. Existing event directories (e.g. eventful) perceived benefit was clearly to be a single access point providing an overview of event information. Another reported benefit is that it supports opportunistic event discovery and facilitates exploration based on different contexts (e.g. location, popularity, categories). Other positive features include: social features (e.g. commenting, sharing events), notification of upcoming events, and shortcuts (e.g. ticket purchase). As for the drawbacks, the main reference was about the unreliability (i.e. unknown source) and incompleteness of information. In particular, low coverage of events and insufficient information for decision support (e.g. lack of location map, videos) have been mentioned. In contrast, the information overload was also seen as a potential drawback making it difficult to find specific events.

When presented the possibility of combining media and event directories, participants recognize benefits due to information enrichment. They claim it would help illustrate events with videos and pictures of past related events (i.e. past performances), other people's experiences, promotional (marketing) material, and so on. The main recognized value would be to give a better idea about the event's environment/atmosphere and provide visual information to support decision making. Participants said it would also support remembering and sharing past experiences. Drawbacks from the merger concern information overload and privacy issues while sharing personal media.

Regarding the possibilities afforded by merging social networks and event directories, some participants think that the main benefits are communication between users, and the sharing of more information (e.g. invitations, opinions, pictures). It was also said to facilitate viewing event attendance, identifying event popularity, and even provide an overview about friends' whereabouts. Live event information (e.g. real-time tweets and comments, live pictures/streams) updates were also seen as a positive afforded feature. Despite the benefits, some participants think the amount of information could clutter the service. Others pointed out that services such as Facebook already provide enough event sharing features. Suggestions included making use of existing social network profiles and/or extending these services.

Other features that users would appreciate having when dealing with events were broadly described with little overlap. Some of these features are: recommendations (based on past attendance, preferences, and from people with similar interests); better visualizations for exploring and searching events (e.g. map integration); the potential to combine categories and attributes while browsing; obtain more information about events and users (e.g. opinions, price and availability).

Conclusions. The opinions gathered seem to support the development of an environment that merges event directories, social networks and media sharing platforms. Moreover, this information enrichment is thought to provide better means of supporting the decision making process. This assumption is based on the possibility of allowing users to better experience an event by viewing associated media. On the other hand, social information obtained implicitly (behaviors) and explicitly (comments, reviews and ratings) provide better judgments of events in terms of attendance, shared interests and reputation. A common concern about information overload suggests that the interface should avoid cluttering and provide only necessary information. Furthermore, there is a need to support different visualizations and better browsing possibilities depending on user interests and constraints. Lack of event coverage and information completeness is another important identified issue that can be addressed using and combining multiple information sources. These issues, along with other identified user interests were translated into a set of requirements in order to guide the following steps of the environment design and development. The following section describes these requirements in more detail.

2.3 User Requirements

Based on this user study, we define a first set of requirements, translating user needs into functionalities that the system should support (Table 2.3). It is important to note that the requirements presented here are representative of users who participated in the previous studies. They should be complemented with other non-functional and functional requirements as described in existing design patterns and interface guidelines [4, 7].

Discovering
Provide a comprehensive coverage of past and upcoming events
Allow searching events based on tags (e.g. performer name, genre, title)
Allow opportunistic discovery by filtering and combining properties (e.g. categories, location, time, price)
Inspecting
Show complete background information about events (e.g. title, location, description, venue, performers, time, category, genre, availability, size)
Allow identifying subjective aspects of events (e.g. popularity, fun, atmosphere, reputation)
Show media associated to events for reliving experiences and for decision support
Show who is joining or joined the event (attendance)
Allow identifying related and repeated events
Visualizing
Rely on traditional media information display (e.g. posters, flyers, ads)
Show only the necessary information in a simple way
Allow different visualizations and browsing contexts (e.g. time, location, people)
Enriching
Allow creating events
Allow associating pictures and videos with existing events
Allow associating comments and opinions with existing events
Sharing
Make use of existing social networks (e.g. Facebook, Twitter)
Allow inviting and recommending events using existing services
Recommending & Preferences
Allow receiving recommendation about events based on personal interests and behaviors
Allow receiving recommendations based on other people's preferences and behaviors (collaborative filtering)
Identify interests and preferences based on past event attendance

Table 1. Requirements

2.4 Scenarios

Scenarios are informal narrative descriptions that allow exploration and discussion of context, user needs and requirements [2].For the purposes of our research, we created a number of scenarios, each covering a range of the aforementioned

requirements and illustrating prospective goals and tasks supported by the system. To better emphasize the context and allow better interpretation and inference of user needs we created four personas. The personas were inspired by the different participants in the previous exploratory study and describe attributes and background information about the actors involved in the scenario. Characteristics are representative of different age groups, professions, preferences, and commonly used event, media and social network sources.

We provide below four different scenarios, each from one different persona.
Scenario 1: *Johnny was invited to a party by a friend and receives a link providing information about this event. He wants to know when and where this event will be and who else was invited. More importantly, he wants to know whether his closest friends confirmed to attend the event or not.*
Scenario 2: *Julie would like to go to a play on her favorite theater. She wants to see a comedy, hopefully playing the upcoming week. She has only been to a few comedies, but she remembers one she specifically enjoyed. Julie would like to see if there is something similar playing and read what other people say about it.*
Scenario 3: *Jack recorded a video with his mobile phone camera while he was attending the Haiti Relief concert from Radiohead given on 24 January 2010 in Los Angeles. He thinks it was a really nice experience and wants to share it on-line. He would also like to see what other pictures and videos were captured during the concert and see how other people experienced the show.*
Scenario 4: *Jessica is going to Paris on her honeymoon and she would like to see what will be happening there during her stay. She wants to do many different things, but cannot decide yet, so she wants to put these things on a "maybe" list in order to decide later. If possible, she would like to see videos of these events to make sure it has a cozy and romantic atmosphere.*

3 Event and Media Ontologies

In this section, we present the ontologies used for representing events, media and users metadata. We use the scenario 3 described above to illustrate these models.

3.1 The LODE ontology

The LODE ontology[5] is a minimal model that encapsulates the most useful properties for describing events [9]. The goal of this ontology is to enable interoperable modeling of the "factual" aspects of events, where these can be characterized in terms of the *four Ws*: *What* happened, *Where* did it happen, *When* did it happen, and *Who* was involved. "Factual" relations within and among events are intended to represent intersubjective "consensus reality" and thus are not necessarily associated with a particular perspective or interpretation. This model thus allows us to express characteristics about which a stable consensus has been reached, whether these are considered to be empirically given or

[5] `http://linkedevents.org/ontology/`

rhetorically produced will depend on one's epistemological stance. We exclude, at this stage, properties for categorizing events or for relating them to other events through parthood or causal relations. We will see in the next section how these aspects, that belong to an interpretive dimension, can be handled through the Descriptions and Situations approach of the Event-Model-F [8].

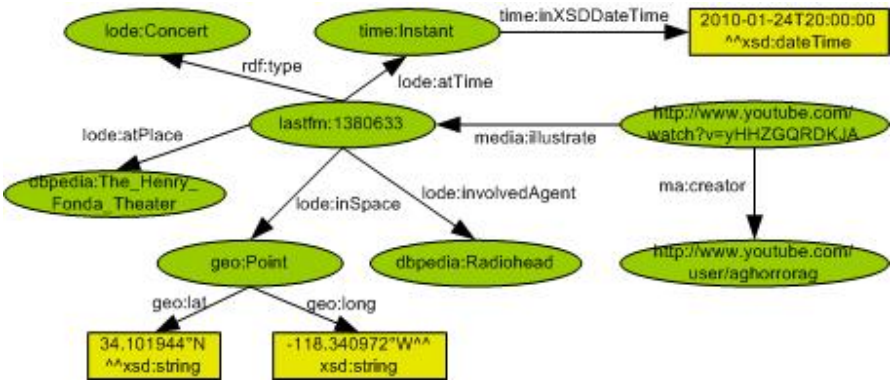

Fig. 1. The *Radiohead Haiti Relief Concert* described with LODE

The Figure 1 depicts the metadata attached to the event identified by `1380633` on last.fm according to the LODE ontology. More precisely, it indicates that an event of type `Concert` has been given on the `24th of January 2010 at 20:00 PM` in the `Henry Fonda Theater` featuring the `Radiohead` rock band.

LODE is not yet another "event" ontology *per se*. It has been designed as an *interlingua* model that solves an interoperability problem by providing a set of axioms expressing mappings between existing event ontologies. Therefore, an OWL-aware agent would infer that the resource identified by `dbpedia:Radiohead` is a `dul:Agent` as described in the Dolce Ultra Lite ontology.

3.2 The Media Ontology

The Ontology for Media Resource currently developed by W3C is a core vocabulary which covers basic metadata properties to describe media resources[6]. It also contains a formal set of axioms defining mapping between different metadata formats for multimedia. In the Figure 1, we see that the video hosted on YouTube has for `ma:creator` the user `aghorrorag`.

The Ontology for Media Resource can then be used to attach different types of metadata to the media, such as the duration, the target audience, the copyright, the genre, the rating. Media Fragments can also be defined in order to have a smaller granularity and attach keywords or formal annotations to parts of

[6] `http://www.w3.org/TR/mediaont-10/`

the video. The link between the media and the event is realized through the `illustrate` property, while more information about the user could be attached to his URI using for example the FOAF ontology.

3.3 The M3O and F Ontologies

The pattern-based ontologies M3O [6] and Event-Model-F [8] can also be used for associating a media with an event description, or for describing relationships between events. This results in a more complex description but brings more expressiveness for representing the context of an annotation such as stating its provenance. These ontologies are based on the foundational ontology DOLCE (Descriptive Ontology for Linguistic and Cognitive Engineering) [3].

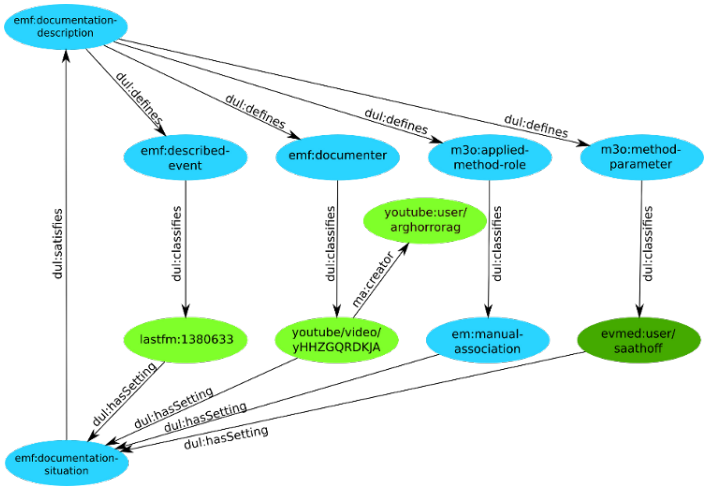

Fig. 2. Associating an event with a YouTube video and provenance information

The Figure 2 depicts the combination of the Event-Model-F Documentation Pattern and the M3O Provenance Pattern. While Media Ontology allows to specify the creator of the video, the patterns in Figure 2 express who actually created the association between the event and the video. We consider the Radiohead concert event to be a described-event and the YouTube video is classified as a documenter, expressing a reified illustrates relation. On the right hand side, we further detail this relation by adding a em:manual-association which is classified by an applied-method-role and the author of this manual association by adding a user evmed:user/saathoff, who is classified by the method-parameter. In summary, using the patterns of the Event-Model-F and the M3O, we can extend the LODE and Media ontologies with provenance information, making the distinction between the creator of some media or event and the creator of

the association between events and media, and even between the participants of this event.

3.4 Data Scraping and Semantization

We are populating these ontologies by scraping and semantifying data from event directories. As a first experiment, we explore the overlap in metadata between two popular web sites, namely Flickr as a hosting web site for photos and videos and Last.fm as a documentation of past and upcoming events. Explicit relationships between these two datasets exist using the `lastfm:event=XXX` machine tag. Hence, more than 1.5 millions photos are indexed with this tag yielding ten of thousands of events.

We use the Last.fm API to convert the event description into the LODE ontology (Section 3.1) and the Flickr API to convert the media description into the Ontology for Media Resources (Section 3.2). The result of this operation is a minimal description of the events where all values are strings (literals). Therefore, we perform an additional step in order to update this description into a truly linked data one. We invoke semantic web lookup services such as the dbpedia one in order to transform these strings into URI identifying unambiguously resources in the web of data. Hence, the "Radiohead" string is transformed into a dbpedia URI[7] which provides additional information about the band such as its complete discography. This URI is declared to be `owl:sameAs` another identifier from the New York Times[8] which provides information about the 38 associated articles from this newspaper to this band. The venue has also been converted into a dbpedia URI[9] but has been augmented with geo-coordinates thus increasing the amount of information available in the LOD cloud for the benefit of all semantic web applications.

The linked data journey can be rich and long. One of the challenges we want to address is how to visualize these enriched interconnected datasets while still supporting the user tasks identified in the Section 2.3.

4 User Interface

In this section, we illustrate initial interface possibilities derived from the requirements and tasks presented in the Section 2.3. The interfaces are represented through low-fidelity prototypes. The prototypes allow exploring, refining and validating prospective concepts along with interface and interaction aspects through small studies with potential end-users and usability experts. Unsurprisingly, the sketches below correspond to the basic properties defined in the LODE ontology.

[7] `http://dbpedia.org/page/Radiohead`
[8] `http://data.nytimes.com/N12964944623934882292`
[9] `http://dbpedia.org/page/The_Henry_Fonda_Theater`

4.1 Views and Perspectives

What - One prospective view is media centered and allows to quickly illustrate the event through associated media. In this view we display events through a representative images and convey different event characteristics (e.g. relevance, rating, popularity, etc) with one and/or more of the image properties, i.e., size and transparency. This approach has been used in other applications[10] to represent clustered result sets or convey sorting by size on different contexts (Figure 3a).

When - Ordering can also be used to represent chronological event occurrence. In fact, the time centric view can be interpreted as the sorting of events chronologically (Figure 3b).

Where - A location centric view can be used to represent where the events occur geographically to orient the user and convey distance. The use of maps is commonly used to visualize such information (Figure 3c).

Who - Events are intrinsically bound to a social component. Users want to know who will be attending to an event when deciding to attend to it. In this context, a people centric view would be relevant to explore the relationships between users and events. Alternatively we can combine attendance information to other views such as location, allowing users to browse for friends on a map and identify their attended events. It could also be used to provide means of visualizing event popularity ,e.g. identify the cities hot-spots on a map, indicate visual cues of popularity according to number of attendees.

In order to allow users to relive experiences from events attended in the past, follow future confirmed events, and keep track of authored events, it is necessary to display events in the context of the users own attendance and ownership. For this reason we will support a "my events" feature with overall browsing possibilities. If several views are to be supported one challenge that can arise concerns transitions between these different views. This is specifically important for facet browsing, due to sudden disappearance of items during navigation [7]. Animated transitions could be used in order to allow the users to maintain orientation during such navigational changes.

4.2 Search Interface

When discovering events we believe users will also rely on browsing, which allow them to analyze large sets of event sets, and narrow them according to their interests and constraints. Overall, we believe users will have different information or browsing/search needs as follows:

- Navigational - when the intention is to reach a particular known event;
- Contextual Browsing - discover one or more events given a specific context (e.g. by location, performer, type, time);
- Entertainment Browsing - serendipitous and opportunistic discovery of events;

[10] See for example `http://www.jinni.com` or `http://www.ted.com/talks`.

Since it is often easier to recognize a word or name than it is to think up that term, it is useful to prompt users with information related to their needs. Based on that principle, we will explore the use of hierarchical faceted metadata which will allow users to browse through multiple categories, each corresponding to different dimensions of the collection [4]. As a general guideline and given users' request, we will avoid empty results during search. Faceted browsing can avoid empty results by restricting the available filtering options in the given focus to only those which lead to non-empty results (poka-yoke principle) [7]. Consequently, the user is visually guided through an interactive query refinement process, while visualizing the number of results in different categories. Additionally, we will explore the information afforded by linked data to display results which are closely related to the user interests. For example, if during a search, no Jazz concerts are available in Amsterdam, we will show other events from nearby cities, time period or even other type of events closely related to Jazz music.

Fig. 3. Interface views illustrating a set of events under: (A) media centric perspective; (B) a chronological perspective; and (C) location centric perspective

While trying to reach a specific event, traditional keyword search can be done through entry forms. Dynamic term suggestions or auto-completion can be used to provide rapid and effective user feedback by suggesting a list matching terms

as the user types the message. Semantic auto-completion extends this method by providing means of clustering the terms according to different categories or facets [5]. Keyword search can also be integrated to faceted browsing and extend the defined classification options. In this context, it is important to indicate if the search will act as a keyword filter or if it will match the classification terms [7]. In regards to event attributes at initial search constraint definitions, time, place and event type seem to be the core indispensable inputs. A potential solution is to always display these attributes during the whole searching/browsing process to enable zooming-in and out from a search result set at any point. Since time period is a range variable input, a common solution is to use a timeline slider control input [7].

4.3 Event Representation

When representing an event instance, we show all information needed to support the decision making process (e.g. Figure 4). Since experiences are centered around media content, we wish to explore different media that better illustrate the event to end-users. Some information that can support decision making are the following.

- background information (e.g. performers, topic, genre, price, attendance list, etc)
- subjective or computed attributes (e.g. reputation, fun, atmosphere, audience)
- user opinions, comments and ratings (strangers and friends)
- representative media (ads, media from past related events, media from the audience, etc.)

Apart from the inspection of the event instance, other conceptual classes (e.g. users, venues, performers, media) should also have accessible views, so that the user can obtain more information about these instances and explore events related to them. In future work we will also identify what are the relevant associated information and how to represent navigation from and to these nodes.

4.4 Enriching Information

Regarding event content enrichment, interfaces that allow users to add/upload information and assign such information to events will be investigated and explored in future studies.

One of the required enrichment features refers to assigning user attendance and keeping track of the users' previously attended events. This information can be used so that the user can easily access past experiences. Moreover, attended events may be used to identify user interests for recommendation and personalization of the facet-pears during search [7]. In order to keep track of events, we will give options to allow users to say if they were in a past event (e.g. *I was there*) or if they are attending to an upcoming event (e.g. *I will go*). Another

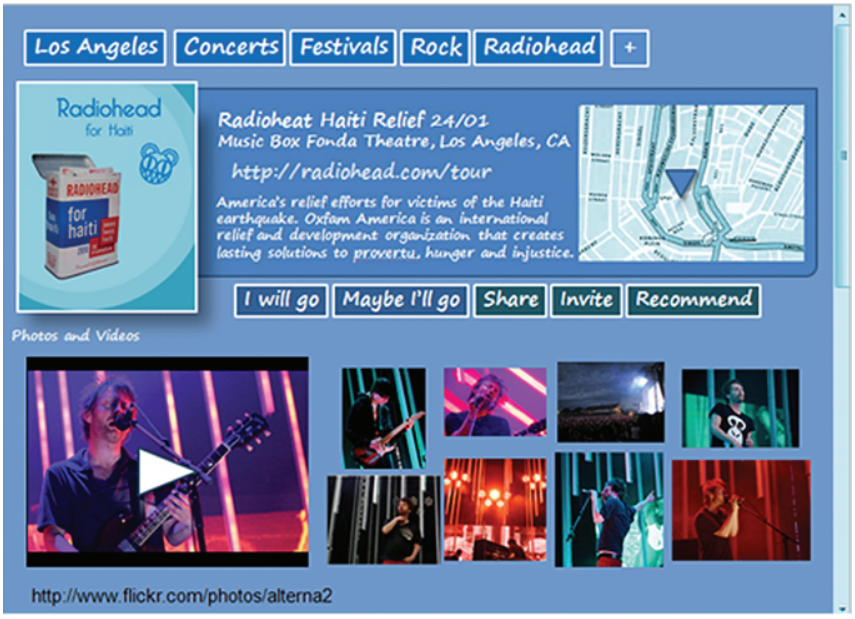

Fig. 4. Interface illustrating an event instance view for a Radiohead concert

prospective option is to allow the user to select events that he is unsure if he will attend (e.g. *I might go*). This will allow adding multiple events to a "maybe" list for future decision or even comparison.

Finally, since users are likely to revisit information they have viewed in the past [4], we will also support simple history mechanisms, by saving a list of recently viewed events. History mechanisms can also be incorporated into the facet search to allow users to undo query filtering and return to a specific query set.

5 Conclusion and Future Work

In this paper, we have described an event-based approach for users to explore, annotate and share media. We first conducted a user study where users were asked about real-world tasks they would like to carry out. We have then extracted requirements and described some scenarios for an event-based system for discovering and sharing media. We advocate the use of linked data technologies for integrating information contained in event directories and we described how event and media ontologies can be used. Finally, we present some sketches of user interfaces that we will develop in the coming months.

In following studies, we intend to use the scenarios we have written to understand how end-users interpret and fulfill associated goals. This will allow us

to identify patterns of interaction, information seeking strategies and information sources required to complete the described tasks. We will continuously increase our coverage of event directories by scraping more data sources and hence demonstrating how interoperability problems can be addressed using semantic web technologies.

6 Acknowledgments

The research leading to this paper was supported by the European Commission under contract FP7-216444, Petamedia Peer-to-peer Tagged Media, and contract FP7-215453, WeKnowIt. The authors would also like to thank Hyowon Lee from Dublin City University (DCU) for fruitful discussions on the design of the EventMedia interfaces.

References

1. A. K. Amin, M. Hildebrand, J. van Ossenbruggen, and L. Hardman. Designing A Thesaurus-Based Comparison Search Interface For Linked Cultural Heritage Sources. In *15th International Conference on Intelligent User Interfaces (IUI'10)*, pages 249–258, Hong Kong, China, 2010.
2. J. M. Carroll. *Making use: Scenario-based design of human-computer interactions.* MIT Press, 2000.
3. Aldo Gangemi and Peter Mika. Understanding the Semantic Web through Descriptions and Situations. In *2nd International Conference on Ontologies, Databases and Applications of SEmantics (ODBASE'03)*, pages 689–706, Catania, Italy, 2003.
4. Marti Hearst. *Search User Interfaces.* Cambridge University Press, 2009.
5. M. Hildebrand, J. van Ossenbruggen, L. Hardman, and G. Jacobs. Supporting Subject Matter Annotation Using Heterogeneous Thesauri, A User Study In Web Data Reuse. *International Journal of Human-Computer Studies*, 67(10):888–903, 2009.
6. Carsten Saathoff and Ansgar Scherp. Unlocking the Semantics of Multimedia Presentations in the Web with the Multimedia Metadata Ontology. In *19th World Wide Web Conference (WWW'10)*, Raleigh, USA, 2010.
7. Giovanni Maria Sacco and Yannis Tzitzikas. *Dynamic Taxonomies and Faceted Search: Theory, Practice, and Experience*, volume 25 of *The Information Retrieval Series.* Springer, 2009.
8. A. Scherp, T. Franz, C. Saathoff, and S. Staab. F—A Model of Events based on the Foundational Ontology DOLCE+ Ultra Light. In *5th International Conference on Knowledge Capture (K-CAP'09)*, Redondo Beach, California, USA, 2009.
9. R. Shaw, R. Troncy, and L. Hardman. LODE: Linking Open Descriptions Of Events. In *4th Asian Semantic Web Conference (ASWC'09)*, 2009.
10. Utz Westermann and Ramesh Jain. Toward a Common Event Model for Multimedia Applications. *IEEE MultiMedia*, 14(1):19–29, 2007.

GLOCAL: Pro-am collaboration in the news production

Denis Teyssou
Medialab R&D manager

Agence France-Presse (AFP)
{denis.teyssou}@afp.com
http://www.afp.com

Abstract. This paper presents the approach of the Glocal European funded project towards the co-production of news on World events, on an information marketplace involving both amateurs and professionals. It discuss the rise of *user-generated content* amongst worldwide media and how event modelling and technology usage may help to foster this pro-am collaboration. Glocal: Event-based Retrieval of Networked Media [1] is a large scale integrating and collaborative project which started in December 2009 for three years and involves ten partners.

Keywords: event, user-generated content, participatory journalism, citizen journalism, information marketplace, pro-am collaboration

1 Introduction

"There's a plane in the Hudson. I'm on the ferry going to pick up the people. Crazy". This sentence posted with an iPhone on Twitpic [2] is by no means a breaking news according to journalistic standards but the picture which goes with it was definitively a stunning news alert.

It has even become an iconic photograph of user-generated content (UGC). It is now a symbol of how eyewitnesses, amateurs with a camera or a smartphone in their pocket can play a major role in the co-production of news, providing, like professional journalists, that they are in the right place at the right moment, to witness and report on a breaking news event.

This picture of the Hudson miracle sent on Twitpic on January 15th 2009 when the US Airways flight 1549 plane "landed" into the Hudson River in New York City (reproduced here below by courtesy of its author Janis Krums) is a kind of paradigm of this UGC uprising, an event with an happy end in a world of earthquakes, tsunamis, wars, crisis and diseases: a kind of exception to the assertion of Marshall McLuhan who once wrote that "news are bad news" [3].

From the London bombings in July 2005 to the Iranian post-election protests in June 2009, through the Sichuan earthquake in August 2008 in China and the Mumbai attacks in India in November of the same year, participatory journalism has provided raw material, valuable firsthand information and eyewitness reports to mainstream medias professional journalists around the World.

©Janis Krums

Fig. 1. Rescue of passengers of a US Airways flight 1549 in the Hudson River on the 15th of January 2009

Months after months, events after events, news content produced by amateurs continue to flow on Twitter, YouTube, Flickr, and other websites, including media. The public is not anymore a passive reader or listener and participates in the news process, most of the time by commenting, but also as producers.

And this trend is likely to last. According to a recent study from the American Pew Research Center, "Understanding the participatory news consumer" [4], up to 9% of American "internet users have contributed their own article, opinion piece, picture, or video to an online news site".

A figure which seems to be in line with the empirical 90-9-1 principle stating, according to Jakob Nielsen, that "in most online communities, 90% of users are lurkers who never contribute, 9% of users contribute a little, and 1% of users account for almost all the action" [5]. This user participation is also rapidly spreading with the use of smartphones: according to the Pew Research Center study, 40% of on-the-go news consumers on mobile device are willing to participate in online news creation.

"What is important here: Citizen comment and writing have upended the relationship between the professionals, journalists being paid by the Old Media, and the amateurs, the rest of us (...) It's a Pro-Am World (...) Amateur content is expandly rapidly", writes Ken Doctor, a renown media analyst from Outsell research in his recent book Newsonomics [6].

In the next section, we will review the rise of *participatory journalism* on news events reporting and in section 3 we will detail the Glocal approach.

2 Participatory journalism

While citizen journalism websites are often struggling to challenge professional media, as the State of the Media special report of 2010 on community journalism [7] shows, the pro-am collaboration in participatory journalism is changing the way professionals are reporting about the events shaping our World.

We understand here the notion of "participatory journalism" not only as "the initiatives undertaken by mainstream media to enhance the integration of all kinds of user contributions in the making of news" [8] but also as a way to build a community around news production, bridging the gap between amateurs and professionals in the field of journalism.

Many mainstream media, from CNN to Fox, through MSNBC, ABC, NPR and Broadcast Interactive Media, have established channels to collect UGC. In Europe, TVs like the BBC UGC hub, France 24 Observers, free dailies such as 20 minutes and Metro are brands using UGC as well as Bild in Germany, Le Parisien in France, The Guardian in England, just to name a few of them.

CNN's iReport, launched in 2006, have more than 100,000 citizen journalists registered and has received more than 300,000 reports, with more than 10,000 per month, 1000 of them being displayed on TV [6].

Some startups have also emerged around UGC as information marketplace to collect that type of content and to offer it to mainstream media. American Demotix and French company Citizenside.com [9] are such companies which have created a two sided-market platform gathering the content and managing communities of several thousands of independant and amateur photojournalists and offering brokered transactions to a global market of media, at a price shared with the copyright owner. Citizenside.com is involved through AFP in the Glocal project.

If dealing with UGC and interacting with the audience are increasingly part of the professional work, sociologists of media have noticed the concerns of the journalists about newsworthiness, trust, legal rights, personal tone, subjective biais, independence and accountability of UGC.

High workload, lack of time and ressources, lack of procedures and tools to verify the reliability and accuracy of user-generated content make professional journalists more cautious, if not reluctant, to give credit to UGC, especially when some hoaxers manage to slip past the gatekeepers by producing fake photos or videos, not to say an entire false event like the Bluewater fake suicide bombing in September 2009 [10].

But these hoaxes do not prevent that amateur pictures like Mr Krums' Hudson River plane are stunning breaking news, while the *networked public sphere* of activists, amateurs, students and security experts, who did the news gathering, analysis and distribution on public concerns about the Diebold electronic voting machines in the US, outplayed with their investigation the mainstream media [11].

In newsrooms, "there is indeed a strong belief that the primary role of journalism lies in the selection stage of the news making process. Their gatekeeping

skills are among the major traits through which professionals distinguish themselves from amateur journalists", point out Social scientists Steve Paulussen and Pieter Ugille.

How to assess the quality of the content and increase confidence in the users news production to avoid what researchers Alfred Hermida and Neil Thurman called "a clash of cultures"? [12] How to engage a conversation with the amateurs producers to get more information on the context of their work and take into account the different views of an event? How to avoid hoaxes and fakes which hurt mainstream media reputation and discredit their newsroom as well as UGC?

So far, the answers to these questions rely only on the newsrooms facts checking policies and on the investigation skills, especially online, of their staff. Nevertheless, this is where both social media and innovative computer sciences techniques may help to assess the reliability of UGC and to increase confidence and collaboration between amateurs and professionals on events coverage.

The notion of event is still rather limited in the news industry where it has been mostly used to prepare agendas of planned events, using a common interchange format, instead of indexing unplanned breaking news and to propose a user experience entirely based on events retrieval.

Furthermore, prevention of hoaxes often relies on merely checking the coherence of the Exif/IPTC/XMP metadata with what the picture is about and to eventually detect visible photo retouching.

3 The Glocal approach

We precisely intend to go beyond those limitations in Glocal, a European Commission funded project, standing as a boundary object in the pro-am co-production of news. A boundary object is a concept borrowed from the sociology of Science describing the intersection of different social worlds. It was introduced in 1989 by Social scientists Susan Leigh Star and James R. Griesemer [13] in an ethnographical study of the coordination mechanisms of scientific work between amateurs and professionals, revisiting the Callon-Latour Actor-Network-Theory (ANT) [14].

Science and Technology are propitious to mass collaboration between amateurs and professionals. Astronomy with discoveries of exoplanets, asteroids or supernovas done by amateurs and open source software developpement, like Linux distributions relying on thousands of voluntary programmers, are well known examples of this so-called Pro-Am revolution [15].

With an innovative user interface fostering the pro-am collaboration and the sharing of different focus on the news, with event modelling based on the International Press Telecommunications Council (IPTC) standard EventsML-G2 [16] to improve the content metadata in a more professional way and to enhance the user search experience, with visual search by similarity, facts checking tools, linked data to other related events on news websites, Glocal aims at building a new kind of news information marketplace.

A study on the top queries performed on AFP's ImageForum photographic database (10 million pictures) [17] over 18 months showed that professional users are primarily querying for locations where a news event takes place (54%), then for persons names (22%), sports (12,5%) and other type of events such as disasters, entertainement (the Oscars Ceremony for instance) or political events (G8, elections...).[18]

This keyword-based queries gives a glimpse of the path to improvement that we can foresee by modeling events and by displaying in a user interface facets clustering the events related named entities (locations, persons, organisations, works) and the relationships between them, as well as relating those events to different thematics such as the IPTC newscodes taxonomy.

We consider Events as a kind of metaclass of an ontology describing *what happens*. If we take the analogy from the five W's of journalism (What, Where, When, Who, Why), we can consider that each newsworthy event describes *what happens* (the action) and its classes: *where it happens* (location), *when it happens* (time range), and *who is involved* (persons, companies, organisations, ...).

This factualization of events is very much at the heart of the recently released EventsML-G2 framework whose syntax is based on XML and also complies with RDF for exchanging event information in the news industry environment.

Why and How the event happens are beyond the scope but we plan to pay special attention to the focus, the point of view of who (actor, eyewitness, victim, direct or undirect source, media) is reporting about the event, a metadata which is the infosource element in EventsML-G2.

Coming back to our example of the Hudson miracle, the event behind the sentence "there's a plane in the Hudson. I'm on the ferry going to pick up the people. Crazy." was actually reported by AFP as the following news alert:

US-air-accident
A US Airways plane carrying commuters crashed on Thursday in New York City's Hudson River and was floating in the water, according to US media.

With an event modelling based on the main 4 W's of journalism (What, Who, When, Where) and proper sourcing, Glocal will provide to the amateurs an interactive framework to describe and categorize their event experience bridging the gap between their local representation and the global one, in an attempt to construct common definitions and meanings and define representativities like in the translation process of ANT.

Furthermore, Mr Krums' photograph as posted on Twitpic does not show any useful Exif metadata such as the camera used to take the picture or the GPS coordinates of the place where the picture was taken. If present, both elements would have increased the confidence in the reliability of the content.

In Glocal, special attention will be given to preserve as much metadata as possible and to check the coherence of this metadata with the photo or video object structure, taking also into account the user profile, uploads history or IP address, as well as the user description of the context of their contribution.

Location-awareness, with the use of geocodes eventually stored in the Exif metadata and/or geotags annotations added by the users or declared in their

registered profile, will help to display events on maps and whenever necessary, Glocal will help the user to disambiguate location tagging with the use of an external knowledge base such as GeoNames [19], minimising interactions between users and the system, by showing a map view of the event at an optimal zoom level.[20]

Although fact checking is the key to ensure the reliability of the content and deliver a guarantee to media willing to use this material, attention will also be paid to protect the user's privacy, especially when the content comes from countries where the freedom of speech is not respected.

Techniques like visual signatures, similarity search allowing to check if a picture or a video presented as a news event has already been published somewhere else on the web will help to increase confidence in the reliability of the content while deeper analysis of the image structure may help to detect if the digital images have been tampered with.

Social media tools will be used to manage, engage and reward the community of users while similarity and event search will also allow for large scale event matching on the web, linking Glocal news to other linked data sources of information.

Glocal will mainly focus on image material, photo and video. "While much of the public feels deficient in writing skills, anyone can take a picture or use a camcorder", remarks Ken Doctor, while in a comparative study on UGC use in The Sun (UK) and Aftonbladet (Sweden), Dr. Henrik Ornebring noted that "the only reader material that is given similar status to material produced by the news organisation is reader photos of breaking news events" [21] like the plane in the Hudson River.

4 Conclusion

In this paper, we presented the goals of the Glocal project, aiming at bridging the gap between the global (media) representation of events and the local (user) focus on those events, and at fostering collaboration between professionals and amateurs in the news production.

By analogy with the pro-am collaboration in Science and Technology, we showed that the rise of user-generated content is currently setting up a new deal in the media industry. Despite concerns about reliability and trust, media are dealing with UGC and encouraging participation and dialogue with the amateurs.

By using events as the primary means for organizing and indexing media, even in the case of unplanned breaking news, Glocal aims to go beyond the state of the art of existing systems in the news industry.

As a pioneer of the concept of the *information marketplace*, Michael Dertouzos, the late Director of the MIT Computer Science Laboratory, once wrote: "the Information Marketplace will inevitably cause us to bring together our humanistic and technical sides, which have been artificially split for centuries" [22].

In Glocal, we will try to bring together social media, news Pro-am co-production and computer science techniques with the goal to participate in the future of the news and of the networked public sphere.

References

1. http://www.glocal-project.eu/
2. Twitpic is a web platform allowing to share photos on Twitter http://twitpic.com/135xa
3. M. McLuhan. Understanding Media. Routledge (1964)
4. http://www.pewinternet.org/Reports/2010/Online-News.aspx
5. http://www.useit.com/alertbox/participation_inequality.html
6. K. Doctor. Newsonomics. St. Martins Press, New York (2010)
7. http://www.stateofthemedia.org/2010/specialreports_community_journalism.php
8. S. Paulussen, P. Ugille. User-Generated Content in the Newsroom: Professional and Organisational Constraints on Participatory Journalism in Westminster Papers in Communication and Culture (2008)
9. Agence France-Presse (AFP) holds a stake of 34% in Citizenside.com
10. http://www.wired.com/threatlevel/2009/09/bluewater/
11. Y. Benkler. Wealth of Networks. Yale University Press (2006)
12. A. Hermida, N. Thurman. A clash of cultures: The integration of user-generated content within professional journalistic frameworks at British newspaper websites. In Journalism Practice, Vol. 2, No. 3, pp. 343-356, 2008
13. S. Leigh Star, J. R. Griesemer. Institutional Ecology, 'Translations' and Boundary Objects: Amateurs and Professionals in Berkeley's Museum of Vertebrate Zoology, 1907-39; Social Studies of Science, Vol. 19, No. 3 (Aug., 1989)
14. http://carbon.ucdenver.edu/~mryder/itc_data/ant_dff.html
15. C. Leadbeater, P. Miller. The Pro-Am revolution. Demos (2004)
16. International Press Telecommunications Council: http://iptc.org/cms/site/index.html?channel=CH0112
17. ImageForum: http://www.imageforum2.afp.com/ and ImageForum diffusion: http://www.imageforum-diffusion.afp.com/
18. AFP internal study performed by the author in 2008 and later revised in 2009.
19. http://www.geonames.org/
20. P. Serdyukov, V. Murdock, and R. van Zwol. Placing flickr photos on a map. In SIGIR '09: Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval, pages 484–491, New York, NY, USA, 2009. ACM.
21. H. Ornebring. The consumer as producer - of what? User-generated tabloid content in The Sun (UK) and Aftonbladet (Sweden). In Journalism Studies 9 (5) 771-785 (2008).
22. M. Dertouzos. What will be: How the new World of Information will change our lives. HarperEdge (1997)

Search and retrieval of audiovisual content by integrating non-verbal multimodal, affective, and social descriptors

Antonio Camurri

Casa Paganini – InfoMus Intl Research Centre
DIST- University of Genova
Piazza Santa Maria in Passione 34, 16123 Genova, Italy

antonio.camurri@unige.it

Abstract. One of the research challenges for future search engines concerns the integration of multimodal and cross-modal, nonverbal, full-body, affective, social, and enactive interaction in the process of search and retrieval of audiovisual content. The paper gives a short presentation of the three-year EU project I-SEARCH (EU 7FP ICT STREP), aiming at creating a novel unified framework for multimodal and cross-modal content indexing, sharing, search and retrieval of audiovisual content. A couple of scenarios developing multimodal paradigms of search and retirieval of audiovisual content are introduced and briefly discussed to explain in concrete terms some of the main research challenges that are addressed in I-SEARCH. Finally, the paper presents preliminary results on a specific research challenge: analysis of non-verbal expressive and social behaviour to extract useful information from users for the retrieval of audiovisual content.

Keywords: non-verbal full-body multimodal interfaces; cross-modal descriptors; emotion; social signals; sound and music computing.

1 Introduction

Internet is quickly evolving towards providing richer and immersive experiences, in which the user interact seamlessly and transparently with digital and physical artefacts. Due to the widespread availability of digital recording devices, improved modelling tools, advanced scanning mechanisms as well as display and rendering devices, even on mobile environments, users are more and more empowered to have a more immersive and interactive experience. Digital media are moving to "User Centric Media" [2,3], enabling adaptive and active experiences of audiovisual content (see for example the EU ICT SAME Project www.sameproject.org). Users become "prosumers", and are more and more participating in the updating process of the information and in improving the resolution and richness of the media repositories. The emergence of embodied and social interaction with content, enabled by the dramatic advances of multimodal/intelligent/natural interfaces, enriches this scenario, providing users with further degrees of freedom and channels to access the content, in terms of full-body, non-verbal, expressive [7], social [5] interaction with content.

It is therefore now possible for users to rapidly move from a mainly textual-based to a media-based "embodied" Internet, where rich audiovisual content (images, graphics, sound, videos, 3D models, etc.), 3D representations (avatars), virtual and mirror worlds, serious games, lifelogging applications, multimodal yet affective utterances (gestures, facial expressions, eye movements,…) etc. become a reality. See for example [1] for an extensive survey on content-based multimedia information retrieval, and the "white papers" from EU on Future internet and on User-Centric Media [2,3] for an in depth analysis of future internet and emerging user-centric media.

Traditional search of music archives usually include descriptive methods (mainly textual, e.g., author, title, etc.). Recent search engines also include alternative querying modalities such as audio (Shazam, Google China Music) or image (Google goggles, Google similar images). However, these search engines are suited for query-by-content or query-by-example, where the research objective is clearly defined or where a specific information is targeted.

We aim at developing alternative, yet complementary, querying modalities, e.g., integrating expressive gesture and affective, emotional cues, that facilitate more explorative and creative search.

This paper presents some insights and preliminary research results on the problem of search and retrieval in cases where textual information is either missing or it is not sufficient or adequate, and therefore the need for the integration of non-verbal multimodal, cross-modal, full-body, affective, social descriptors emerges.

Use case scenarios on music content search and retrieval are adopted in this paper to discuss main research challenges and approaches.

2 The I-SEARCH EU Project

The EU 7FP ICT STREP I-SEARCH project aims to create a novel unified framework for multimedia and multimodal content indexing, sharing, search and retrieval. I-SEARCH is coordinated by Dimitrios Tsovaras by ITI-CERTH (Centre for Research and Technology Hellas – Informatics and Telematics Institute), and partners include JCP-Consult, INRIA, Athens Technology Center, Engineering Ingegneria Informatica S.p.A., Google, University of Genoa, Exalead, Erfurt University of Applied Sciences, Accademia Nazionale di Santa Cecilia, EasternGraphics. The project started in January 2010, with a duration of three years.

"I-SEARCH aims to create a novel unified framework for multimedia and multimodal content indexing, sharing, search and retrieval. I-SEARCH aims to be the first search engine able to handle specific types of multimedia (text, 2D image, sketch, video, 3D objects, audio and combination of the above) and multimodal content (gestures, face expressions, eye movements) along with real world information (GPS, temperature, time, weather sensors, RFID objects,), which can be used as queries and retrieve any available relevant content of any of the aforementioned types and from any end-user access device. Towards this aim, I-SEARCH proposes the research and development of an innovative Rich Unified Content Description (RUCoD). RUCoD will consist of a multi-layered structure, which will integrate descriptors of all of the above types of

content, real-world information, even non-verbal yet implicit, emotional cues and social descriptors, in order to better express what the user wants to retrieve. Another objective of I-SEARCH is the development of intelligent content interaction mechanisms, including personal, social-based and recommendation-based relevance feedback and novel interoperable multimodal interfaces. This will result in a highly user-centric search engine, able to deliver to the end-users only the content of interest, satisfying their information needs and preferences, which is expected to dramatically improve end-user experience and offer new market opportunities. Furthermore, I-SEARCH introduces the use of advanced visual analytic technologies for search results presentation in order to facilitate their fast and easy interpretation and also to support optimal results presentation under various contexts (i.e. user profile, end-user terminal, available network bandwidth, interaction modality preference, etc.). Finally, the search engine will be dynamically adapted to end-user's device, which will vary from a simple mobile phone to a high-performance PC." (from cordis.europe.eu projects archive)

3 Scenarios

3.1 Music retrieval through expressive embodied queries

Chiara is a music-lover, looking for music material that share common affective features. Here, search aims at discovering unexpected filiations and similarities across music artworks. It takes place in an environment equipped with devices enabling the user to express herself through voice, hands and body gesture. Pre-recorded multimedia content can be uploaded from an external device (e.g., a mobile phone). For each digital content in the collection, descriptors related to low-level features, real-world context data, and expressive/emotional/social cues that compose the RUCoD (Rich Unified Content Description of the I-SEARCH project) standard are stored.

Inputs to the search module include text queries, audio capture of live user singing incipit of music piece, audio file recorded on handheld device such as mobile phones. Beat tracking can be captured by tapping on a microphone or through accelerometers embedded in mobile devices. User gestures can be captured using either video camera or accelerometers embedded in user's mobile devices.

Chiara wants to explore music artworks that share affective features with the Ravel's Bolero. She starts by using the I-SEARCH framework to retrieve audio information that share similarities with this audio pattern. Using a tangible acoustic interface [6], she taps the beating of the rhythm, a constant 4/4 time with a prominent triplet on the second beat of every bar, or smaller rhythmic cells (for example the beat and triplet).

The recorded audio fragment is used by the I-SEARCH engine as an audio query to initialize the search. Specifically, the I-SEARCH framework extracts low-level descriptors from the audio content (the tapping resulting audio) and create a query based on the RuCoD format. Through template matching techniques [4], similar audio results related to Bolero rhythmic pattern are retrieved. Related video files and music scores are also retrieved through multimodal annotation propagation. Results are

displayed via visual analytics techniques on Chiara's terminal, using clusters annotated with information like modality type, population size and others.
Chiara picks one of the results returned by the query, listens to it but decides that what she needs is something more energetic, so she closes her fists and starts making sharp, sudden vertical movements on the same Bolero rhythm. Through a video camera or embedded accelerometers the environment captures such expressive features of the gesture, and refine the search, resulting in changes in the displayed results to match it (either by removing the items that don't convey that expression, or by moving the suitable items closer to build a different cluster configuration). One of the results captures Chiara's attention: a drum recording from Italian ethnomusicological repertoire, the 'Ritmo di tamburo', where various drum rhythms are played, sharing indeed the same triplet of Bolero. A little further away she also finds a voice recording.

3.2 Collective DJ - Social music retrieval through expressive embodied queries

Four friends at a party wish to dance together, and to accomplish this they search some music pieces resonating with their (collective) mood. They do not know in advance the music pieces they want, and they use the I-SEARCH tool collaboratively to find their music, and possible associated video. Alternatively, the search process might not necessarily be 'conscious' or intentional, but simply a part of a social game, i.e., in a more fun/entertainment approach.
The friends are at a party, but not necessarily they share the same physical environment. One or more of them can be remotely connected via audiovisual links. GPS and context aware information are available to the ISEARCH search engine.
Users have devices that enable to express themselves through voice, movement, face, and full-body movements.
For each digital music content of the archive under consideration, descriptors related to low-level features, real-world data and expressive/emotional/social cues that compose the RuCoD (Rich Unified Content Description) standard are available.
Users inputs include the following: (i) Rhythmic queries, using hands, clapping, full-body movement; (ii) Context data (GPS, compass, proximity of others etc); (iii) Entrainment/synchronisation and dominance/leadership among users, measured by on-body sensors (eg accelerometers or/and videocameras on their mobiles, or game interfaces) and/or environment videocamera(s), to find a shared, collective information to build the query based on non-verbal social signals; (iv) Gestures to shape the query: again, the gesture are captured using either video camera and/or accelerometers embedded in mobile devices (carried by the user or kept in hand) or in environmental videocameras.
The output of the experience include a shared enjoyment of performances of the retrieved music pieces, possible video clips associated to such music pieces (music videos).

The experience may be described as follows:

1. The four users A, B,C, D start to dance;
2. Their movement acts as selector of music pieces coded in relation to their motoric-affective-social behaviour (slow, fast, dionisiac, ...);
3. If the movements of A,B,C,D are not sympathetic (i.e., low entrainment), an overlapping of different music pieces will emerge, in a rather chaotic sound environment (the different music pieces are heard simultaneously at different changing levels according to users behaviour). As the joint experience goes on, the music continuously changes, and may start to converge to a piece corresponding to the user who results to dominate in the group (dominance/leadership features which are automatically extracted by the system);
4. When the movement of the group obtains a sufficient uniformity (contagion from the one who results to act as a leader), the group will converge to dance on the same music, chosen by such "collective gesture": this is the first level of query result;
5. This entertaining task is integrated and is part of the search task: for example, the search can take the priority as soon as one of the users, once obtained a shared agreement and a single music, tries to trigger a change in the shared general emotion of the group (with a consequent change of musical choices), for example a perturbation to the current situation by a user, to explore music pieces similar to the one obtained and experienced by the group. The effect can be a sort of "game on leadership": the user who is able to triggers a general perturbation of emotion in the group causes a sort of "collective DJ-like" real-time interactive editing of heterogeneous music fragments on the main music piece. The user who is able to act as a leader (detected by the system), has the possibility to inject the consequences of her gestural/movement/affective choices, which will take the power to determine the change of the music context only in relation to the capability of her contagion on the other users. If the others will be captured, they will follow her moving toward a new music piece, by means of an audio and possibly video cross-fade or sudden change of scene. Otherwise, if the user will not have enough power or dominance on the others, her associated new music piece will fade off.

This is a sort of a "Collective DJ" example, in which the collective behaviour provides the source data for the search of the music. It is the opposite of the traditional music experience: here the movement determines its own correct music frame [8].
The music retrieval may also keep into account of GPS and spatial locations of users. For example, GPS information may be kept into account to select music pieces keeping into account their geographical region.

3.3 Discussion

Social search can occur simultaneously or in different moments. In the first case, which is considered in the scenario 2, users collaborate to a common shared objective. In the latter, a user leaves a track of her activity, which can be used later by other users to take inspiration for their search. Of course, mechanisms allowing the

intentional control of access to personal search schemas and data, similarly to shared data in social networks, must be introduced.
In situations of "affective non-verbal querying" the following three types of difficulties emerge:

(i) The user is not familiar to non-verbal search by means of "affect", "expressivity", "embodiment": she is not sure if she is able to perform correctly the task, she does not know if she is able to link an affect with a content. For example, if she sings/hums a song but she has poor singing skills, this can be a case in which the user is not able to express the affect she wants to convey (but associated gesture may help). In a "game-like" scenario, the users may be more inclined to dancing even if they are not professional dancers, for the sake of fun and social interaction. In this case, the results of the scenario should be measured more in terms of user experience than on the technical 'efficiency' and performance of the search task.

(ii) The user may be not capable to express what she is searching for;

(iii) The machine might not be able to execute the search according the user intentions.

Despite these difficulties, emotional/expressive social non-verbal queries, with their inner "blurred" characterisation, can enhance "serendipitous discovery", and can lead to stimulating exploring-style querying, complementary to traditional querying paradigms.

4. Multimodal queries based on non-verbal expressive and social features

Several research challenges emerge from the scenarios sketched above, including the understanding of users multimodal inputs, and in particular the non-verbal multimodal cues conveying users' intentions, and the mapping between user multimodal inputs to features in the audiovisual content.

In this section we focus on how to exploit users' non-verbal expressive and social signals useful to build multimodal queries.

The proposed approach consists of two phases:

(i) Extraction of an array of expressive features describing each user behavior [12];

(ii) Using such expressive features as the inputs to modules which extract social features related users behavior, with particular focus on entrainment and dominance [13, 14, 15].

The resulting array of individual and social features will be a subset of the user component RuCoD descriptors.

4.1 Expressive features

As for the extraction of descriptors on non-verbal expressive behaviour, a particular focus is on full-body movement and gesture, i.e., on recognizing how a gesture is performed, including expressive and affective content [12]. Typically, a single gesture can be performed in several different ways (e.g., fluid, hesitant, impulsive). A collection of features characterizing the expressive qualities of a gesture has been defined, starting from biomechanics, psychology, and humanistic theories [12].
Expressive features include the following:

Quantity of Motion (QoM) is an index of motoric activation that provides an estimation of the amount of overall movement (variation of pixels) the video-camera detects. QoM computed on translational movements only (TQoM) provides an estimation of how much the user is moving around the physical space. Using Laban's Effort [17] terminology, whereas Quantity of Motion measures the amount of detected movement in both the Kinesphere and the General Space, its computation on translational movements refers to the overall detected movement in the General Space only. TQoM, together with speed of barycentre (BS) and variation of the Contraction Index (dCI) are introduced to distinguish between the movement of the body in the General Space and the movement of the limbs in the Kinesphere. Intuitively, if the user moves her limbs but does not change her position in the space, TQoM and BS will have low values, while QoM and dCI will have higher values.

Impulsiveness (IM) is extracted using a model measuring it as a combination of other features, mainly derived from QoM. The first one is the variance of QoM in a sliding time window of 3s, i.e., a user is considered to move in an impulsive way if the amount of movement the video-camera detects on her body changes considerably in the time window. A second group of features is related to the analysis of the shape of QoM along time. Such features include e.g., the ratio between the main peak of QoM in the time window and the time duration of such a peak, the steepness of the attack of a movement phase, the steepness of the main peak, the number of peaks detected in the time window, the ratio between the main peak and the second biggest one, the distribution of the peaks in the time window (i.e., whether they are uniformly distributed along the time window or concentrated over a specific time range). Another feature is related to the content of the QoM spectrum in the frequency band over 5 Hz. Each feature is then weighted and combined in the model in order to provide an overall index of Impulsiveness.

Vertical and horizontal components of velocity of peripheral upper parts of the body (VV, HV) are computed starting from the positions of the upper vertexes of the body bounding rectangle. The vertical component, in particular, is used for detecting upward movements that psychologists (e.g., [16]) identified as a significant indicator of positive emotional expression.

Space Occupation Area (SOA) is computed starting from the movement trajectory integrated over time. In such a way a bitmap is obtained, summarizing the trajectory followed along the considered time window (3s). An elliptical approximation of the shape of the trajectory is then computed. The area of such ellipse is taken as the Space Occupation Area. Intuitively, a trajectory spread over the whole space gets high SOA values, whereas a trajectory confined in a small region gets low SOA values.

Directness Index (DI) is computed as the ratio between the length of the straight line connecting the first and last point of a trajectory (in this case the movement trajectory in the selected 3s time window) and the sum of the lengths of each segment composing the trajectory. It is inspired by the Space dimension of Laban's Effort Theory.

Space Allure (SA) measures local deviations from the straight line trajectory. It is inspired by composer Pierre Schaeffer's Morphology. Whereas DI provides information about whether the trajectory followed along the 3s time window is direct or flexible, SA refers to waving movements around the straight trajectory in shorter time windows. Currently, SA is approximated with the variance of DI in a time window of 1s.

The *Amount of Periodic Movement* (PM) provides a preliminary information about the presence of rhythmic movements. Computation of PM starts from QoM. Movement is segmented in motion and pause phases using an adaptive threshold on QoM [12]; inter-onset intervals are then computed as the time elapsing from the beginning of a motion phase and the beginning of the following motion phase. The variance of such inter-onset intervals is taken as an approximate measure of PM.

Symmetry Index (SI) is computed from the position of the barycenter and the left and right edges of the body bounding rectangle. That is, it is the ratio between the difference of the distances of the barycenter from the left and right edges and the width of the bounding rectangle:

$$SI = \frac{\left||x_B - x_L| - |x_B - x_R|\right|}{|x_R - x_L|}$$

where x_B is the x coordinate of the barycentre, x_L is the x coordinate of the left edge of the body bounding rectangle and x_R is the x coordinate of the right edge.

The expressive features, can be analyzed using video input from videocameras, but other sensor inputs can be considered, e.g. the 3D accelerometers embedded in mobile systems.

Current work aims at refining and extending the set of expressive features, to contribute to the RuCoD standard.

The expressive features are implemented as real-time software modules in the open software platform EyexWeb XMI (www.eyesweb.org).

4.2 Social features

Research on the analysis of social descriptors include the development of models and techniques for measuring entrainment, empathy, dominance, leadership, and salient behaviour in small groups of users. We obtained preliminary results on entrainment and dominance, based on theories of synchronization [13,14,15].

A basic assumption to approach nonverbal social behavior consists of the modeling of a small group of users as a complex system consisting of single interacting components able to auto-organize and to show global properties, which are not obvious from the observation of their individual dynamics.

A number of algorithms and related software modules for the automated analysis in real-time of non-verbal cues related to expressive gesture in social interaction are

currently studied at our centre. Analysis of entrainment and dominance is based on Phase Synchronisation and Recurrence Quantification Analysis. We start from the hypothesis that phase synchronisation is one of the low-level social signals explaining empathy and dominance in a small group of users. Another direction, based on Multi Scale Entropy and other approaches is currently adopted to measure saliency and rarity index in small group of users. Real time implementation of the algorithms developed so far is available in the EyesWeb XMI Social Signal Processing Library (www.eyesweb.org).

5 Conclusions

Some of the main research challenges faced in the EU 7FP ICT I-SEARCH project, and preliminary results in the analysis of users behavior in terms of expressive and social features have been presented, as a contribute to the RuCoD standard in I-SEARCH. Current work includes research on empathy, emotional entrainment, leadership, co-creation, and attention.

Other important directions of the research in I-SEARCH concern the study of descriptors in audiovisual content, and the study of cross-modal descriptors [11].

Acknowledgements

I am deeply grateful to my colleagues and friends Corrado Canepa, Paolo Coletta, Nicola Ferrari, Alberto Massari, Gualtiero Volpe, Donald Glowinski, Maurizio Mancini, Giovanna Varni.
This research is partially supported by the 7FP EU-ICT three-year project I-SEARCH no. 248296.

References

1. Lew, Michael S., Sebe, N., Djeraba, C., and Jain, R.. Content-based multimedia information retrieval: State of the art and challenges. ACM Trans. Multimedia Comput. Commun. Appl., 2(1):1–19, (2006).
2. Laso Ballestreros, I. (Ed.) Research on Future Media Internet, Future Media Internet Task Force, European Commission, 7FP ICT Networked Media Unit, January 2009.
3. Laso Ballestreros, I. (Ed.) User Centric Media in the Future Internet, European Commission, 7FP ICT Networked Media Unit, November 2009.
4. Casey, M.A., R. Veltkamp, M. Goto, M. Leman, C. Rhodes, and M. Slaney. Content-based music information retrieval: current directions and future challenges. *PROCEEDINGS-IEEE*, 96(4):668, 2008.
5. Pentland, A.,: Socially aware, Computation and Communication. Computer, IEEE CS Press (2005)

6. Camurri, A., C.Canepa, S.Ghisio, G.Volpe (2009) Automatic Classification of Expressive Hand Gestures on Tangible Acoustic Interfaces According to Laban's Theory of Effort. In M.S.Dias, S.Gibet, M.W.Wanderley, R.Bastos (Eds.), Gesture-Based Human-Computer Interaction and Simulation, pp.151-162, LNAI5085, Springer, ISSN 0302-9743.
7. Camurri, A., De Poli, G., Leman, M., Volpe, G.: Toward Communicating Expressiveness and Affect in Multimodal Interactive Systems for Performing Art and Cultural Applications, IEEE Multimedia, Vol.12, No.1, pp.43-53, IEEE Computer Society Press (2005)
8. Camurri, A.: Interactive Dance/Music Systems, Proc. *Intl. Computer Music Conference ICMC-95*, pp.245-252, The Banff Centre for the arts, Sept.3-7, Canada, ICMA-Intl.Comp.Mus.Association (1995)
9. Camurri, A., Canepa C., Volpe G.: Active listening to a virtual orchestra through an expressive gestural interface: The Orchestra Explorer. Proc. Intl Conf NIME-2007 New Interfaces for Music Expression, New York University, (2007)
10. Varni, G., Camurri, A., Coletta, P., Volpe G.: Emotional Entrainment in Music Performance. Proc. 8th IEEE Intl Conf on Automatic Face and Gesture Recognition, Sept. 17-19, Amsterdam (2008).
11. Camurri, A., P.Coletta, C.Drioli, A.Massari, G.Volpe (2005). Audio processing in a multimodal framework. Proc. Intl. Conf. AES-05 Audio Engineering Society, Barcelona, May 2005.
12. A.Camurri, B. Mazzarino, G. Volpe (2004) Expressive Interfaces. Cognition Technology & Work, special issue on "Presence: design and technology challenges for cooperative activities in virtual or remote environments", P.Marti (Ed.), Vol.6, pp.15-22, Springer-Verlag.
13. Varni G., Mancini M., Volpe G., Camurri A. (2009) "Sync'n'Move: social interaction based on music and gesture". In Proceedings of the 1st Intl. ICST Conference on User Centric Media (UCMedia 2009), Venice, Italy, December 2009. LNICST vol.40, Springer, 2010. (ISBN 978-3-642-12629-1).
14. Camurri A., Varni G., Volpe G. (2009) Measuring Emotional Entrainment in Small Groups of Musicians. In Proceedings of International Conference on Affective Computing & Intelligent Interaction (ACII 2009), Lisbon.
15. Varni G., Camurri A., Coletta P., Volpe G., (2009) Toward Real-time Automated Measure of Empathy and Dominance. In Proceedings of the 2009 IEEE International Conference on Social Computing SocialCom, Vancouver, Canada, August 2009.
16. Boone, R. T., Cunningham, J. G., (1998) Children's decoding of emotion in expressive body movement: The development of cue attunement, *Developmental Psychology*, 34, 1007-1016.
17. Laban, R., Lawrence, F.C., 1947. Effort. Macdonald & Evans Ltd., London.

Interaction Design for the Exchange of Media Organized in Terms of Complex Events

Anthony Jameson and Sven Buschbeck*

DFKI, German Research Institute for Artificial Intelligence
Saarbrücken, Germany

Abstract. Even the most sophisticated automatic recognition of events must often be paired with an appropriate design of the users' interaction with those events. This paper presents three presumably typical use cases and associated interaction design proposals, which illustrate (a) how untrained users can benefit from the organization of media in terms of complex events; (b) how they can have their own media categorized in this way without having to invest much effort; and (c) how they can even create complex event instances with novel structures, without having to think explicitly about event structures.

1 Introduction

As will be shown by many of the papers that will be presented at the EVENTS 2010 workshop, the automatic identification and processing of events raises many technical challenges. But even before solutions to these problems have been found, we have to consider exactly how people might interact with systems that make use of representations of events. Having a clear idea of use cases, scenarios, and interaction designs can help us to see which technical problems are most important and what requirements need to be met.

This workshop paper considers how the recognition and representation of events can enhance interaction in a particular type of system: a media marketplace in which professional and amateur users contribute and exchange various types of media, most typically photos and videos (but also other types, such as audio files and text documents). One underlying idea is that it is often helpful for such media to be indexed and organized in terms of events that they depict or describe, in addition to more familiar indexing on the basis of time, location, tags, and named entities (such as people).

More specifically, we consider how interaction in such a marketplace can be enhanced if not only atomic events but also *complex events* are represented: Such an event may extend over a considerable period of time and consist of subevents, some of which in turn may be complex events. A simple example of a complex event is a soccer tournament, which comprises two or more rounds and a number of games, each of which can in turn be viewed as a complex event.

* The research described in this position paper is being conducted in the context of the 7th Framework EU Integrating Project *GLOCAL: Event-based Retrieval of Networked Media* (http://www.glocal-project.eu/) under grant agreement 248984.

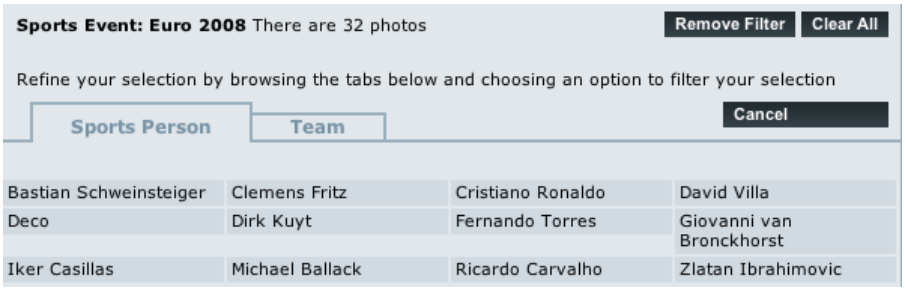

Fig. 1. *Partial screenshot from the photo exchange site Sport Photo Gallery (http://www.sportphotogallery.com/). Though the site offers many photos of the Euro 2008 tournament, it is not possible to navigate among them in terms of the structure of the tournament.*

We will present several scenarios and interaction designs that should help to stimulate thought on the following questions:

1. How could users benefit from the representation in the system of complex events, as opposed to having only simple events represented?
2. How can a user and a system collaborate to build up and maintain a representation of complex events, without any requirement for users to invest more than a minimal amount of effort?

This work is being done in the context of the integrating project GLOCAL.[1]

2 Why Do We Need Complex Events?

Suppose you are an (amateur or professional) photographer or journalist who wants to share, buy, or sell media about the first half of the final game of the 2008 European Cup soccer tournament. Media concerning this event can be found in a number of media exchange sites, including Flickr.[2]

Citizenside.com[3] is an example of a site that specifically supports selling of the media by amateur photographers to professional organizations, such as news agencies. Although this site organizes and indexes media in quite sophisticated ways, you would run into difficulty if you wanted to think in terms of parts of particular tournaments: The site does not organize media in terms of complex events like tournaments.

In the Sport Photo Gallery site,[4] which is dedicated to sports photos (Figure 1), you can find the "Event" Euro 2008, but the media about it are indexed only in terms of players and teams, not parts of the tournament.

[1] Since a special session of the EVENTS 2010 workshop is being devoted to this project, we assume that the workshop proceedings will contain an introductory overview of the project; therefore, we do not include such an overview in this submission. If necessary, we can add such an overview in the final version of this paper.

[2] http://www.flickr.com/

[3] http://www.citizenside.com/en/sell-share-photos-videos.html

[4] http://www.sportphotogallery.com/

It may help to look at this absence of complex events in terms of an analogy: The way in which photos and videos can be embedded in a Google Map—say, of Athens—shows that it is feasible and useful to organize media in terms of a large, coherent structure—in this case, the map of a city. But suppose that some of these media concern events at a conference—for example, a talk in a session of EVENTS 2010, which is in turn a subevent of SETN 2010. Google Maps can show the conference building, but it has no way of representing the additional dimension: the structure of the "conference event".

3 Use Case A: Navigating Via Event Structures

Suppose now that we have a media marketplace that includes:

- structures for complex events;
- media attached to particular events.

(We will discuss in below how the structures and the media will get into the system.)

Then a user can:

- 1. . . . find a complex event with some combination of keyword search, use of a map and a calendar, and/or providing an example medium about that event; Although finding an optimal interaction design for this sort of event search is an interesting challenge, it is not very difficult to find an acceptable solution, so we do not provide any concrete examples in this paper.
- 2. . . . navigate down the hierarchical structure of the complex event to find the part that they are interested in. One way of allowing this sort of navigation is to visualize the complex event as a tree structure in which each node represents an event or a subevent. In the hypothetical screen Figure 2, the user is focusing on the node for the subevent "first half of the final game", and the media associated with that subevent are shown on the right-hand side of the screen. Nodes representing higher-level events can also have media associated with them, for example a video that covers the entire game.[5]

4 Use Case B: Inserting New Media Into an Event Structure

Even if we grant that users could benefit from this type of organization, the question arises of how media are going to get organized in this way. Realistically speaking, we cannot expect most users to spend a lot of time carefully creating complex event structures and assigning media to particular parts of these structures. So on the one hand, we need system-side processing that can handle a lot of the work of creating and populating complex event structures. On the other hand, since we cannot assume that a

[5] The visualizations this paper were created with the MindManager software; they therefore do not reflect the appearance of the interfaces that will ultimately appear in the GLOCAL system.

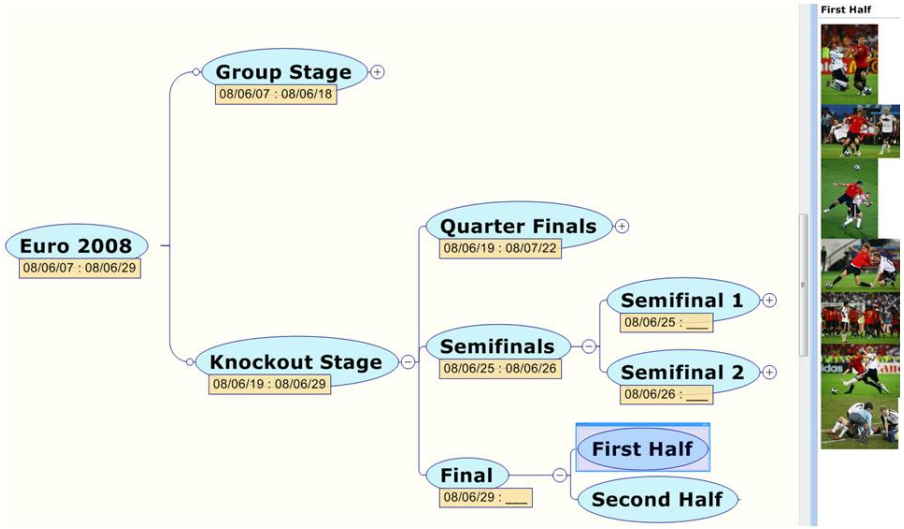

Fig. 2. *Proposed visualization of the structure of a complex event in such a way that it can be used for browsing for media associated with subevents.*

fully automatic solution will be satisfactory, we have to design the user interaction in such a way that users can help the system out without investing much effort.

In this use case, we consider how users might insert media into an existing complex event structure. (The problem of creating such a structure in the first place will be considered below.)

Suppose, concretely, that a photographer has created photos and videos of the Euro 2008 final and would like to add them to the Glocal site (e.g., to sell them or to share them with friends).

In Figure 3, she opens up a new node "New Media" under the "Final" event and uploads the media to the space on the right (which serves as a sort of inbox).

The user could in principle specify by hand whether each medium belongs to the first half, the second half, or the whole game (as with a video that includes highlights form both halves). But the system should be able to do this work largely automatically. Essentially, it can compare the space and time coordinates of the new media—and the low-level properties of their images—with those of the already categorized media.

In Figure 4, the left-hand side of the screenshot shows the system's tentative sorting of the images. The small blue and white icons indicate the system's confidence level: the more blue, the higher the confidence.

The right-hand side of the screenshot shows why it can be important to leave the last word to the user: The user has now deleted two of the low-confidence images (which she now recognizes as being largely irrelevant) and accepted the system's classification of the other images. This example illustrates that, if the user can count on a reasonable amount of intelligence on the part of the system, the user can save some of her own time, even if the system's performance is imperfect. With a bit of effort, the user could have recognized by herself that the photos of the team lining up before the game and

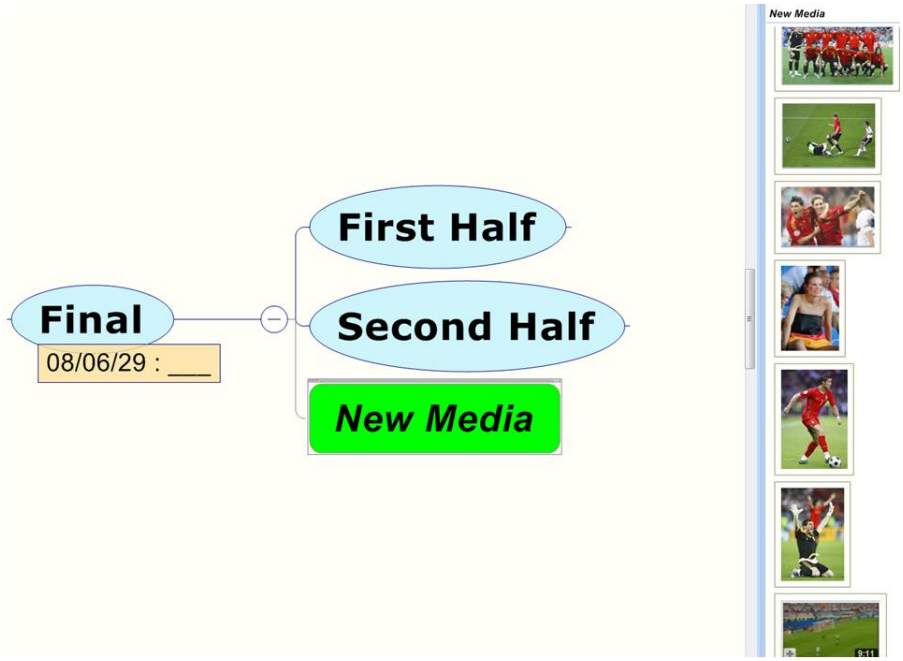

Fig. 3. *Hypothetical screenshot of a situation in which a user is preparing to insert a number of new media into the structure for a complex event.*

of the young lady in the stands do not really belong in the same category as the other photos and videos. But if she knows that the system will make it easy for her to remove any superfluous photos, she doesn't have to be so selective when offering them in the first place.

5 Use Case C: Creating a New Complex Event

But what if the user's new media concern a complex event that is not already represented in the system—maybe because it is of only local interest?

Specifically, assume that a mother has taken photos and videos of her 14-year-old daughter's local soccer tournament. The user will have to create a new complex event instance with an appropriate structure. So in principle, she needs either to find an existing event structure that she can instantiate or create a (partially) new structure that is suitable for describing her event.

The main challenge lies in the fact that most users won't be willing or able to reason in terms of event structures.

The approach that we propose is to support a "copy, paste, and modify" style of event creation.

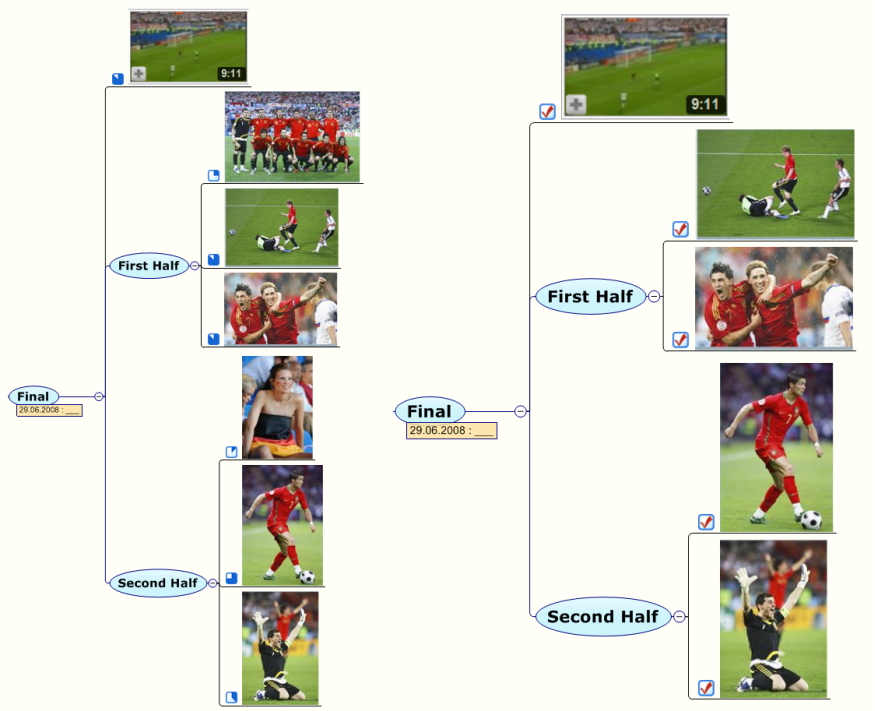

Fig. 4. *Illustration of how* GLOCAL *can propose a classification of a user's new media (left) and how the user can second-guess the system (right).*

A familiar-sounding example of this general approach is an author who creates a properly formated submission to the SETN 2010 conference by taking a Word document with a submission to the SETN 2008 conference:

- If the structure of the author's new submission is exactly parallel to the structure of the old submission, all the author has to do is replace the original content with his own content. He may not have to think explicitly about the structure at all.
- Even if the structure of the old document is not quite right, the author can adjust it in an ad hoc way in the new document, without having to think in general terms about document structures. For example, he might add an appendix using the same format as for one of the normal sections of the paper.
- An intelligent system could support this type of activity by comparing the user's new document with other SETN 2008 (or similar) papers and perhaps suggesting improvements in the structure (e.g., a slightly different way of formatting a section that has the title "Appendix" and comes at the end of the paper).

In Figure 5, we assume that the user who wants to add media of her daughter's soccer tournament has already seen the event structure for Euro 2008 and has therefore decided to copy it as a starting point for the new tournament. She has recognized the need to simplify the structure somewhat and has renamed a couple of the subevents. For

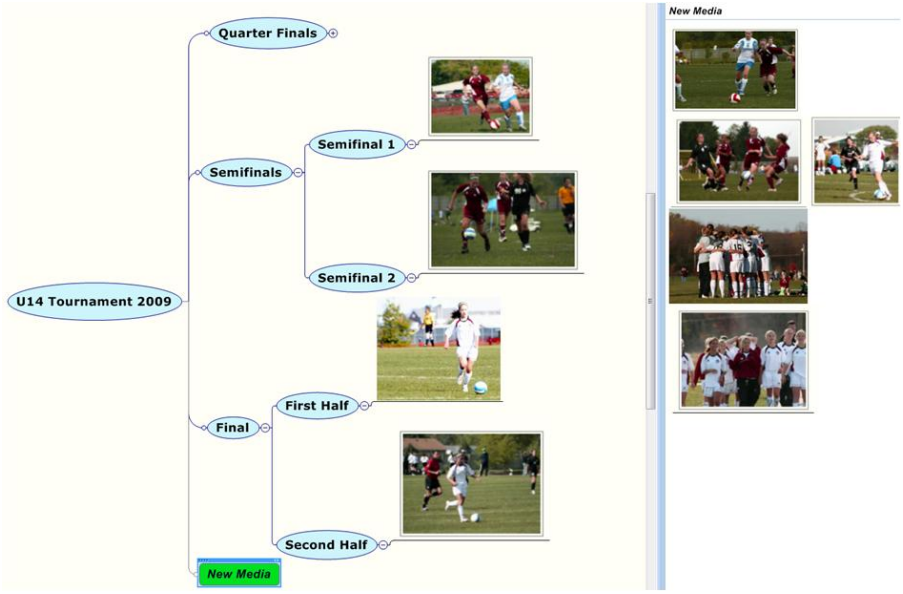

Fig. 5. *Illustration of a situation in which a user has (a) created a new complex event instance, using an existing event instance as a starting point; and (b) inserted a small number of media into the new structure so as to enable the system to insert the other new media.*

example, the youth soccer tournament does not have a distinction between a "Group Stage" and a "Knockout Stage"; it begins directly with the quarterfinals.

The figure shows the state of the system after the user has (as in the previous use case) uploaded her "new media", which concern various games in the tournament, and assigned one medium to each leaf node in the hierarchy. Note that it is necessary for the user to do this initial work of placing some media in the appropriate places, since in this situation the system initially does not know any details about the subevents represented by the nodes and can therefore not perform an initial tentative categorization of new media, as it did in the previous use case.

The system now has some information about the times and places of the games, about the colors of the teams' uniforms in each game, etc. Given this information, the system can guess at the classification of the remaining media, as before (the confidence levels are not shown in the figure).

But it is unlikely that all of the media will fit naturally into the structure that the user has just created, given that this structure was simply created ad hoc on the basis of a structure for another complex event. We must assume that there may be media that call for some adaptation of the event structure.

In our example, as shown in Figure 6, the system notes that the last two photos don't seem to fit into any subevent. The system might conceivably ask the user to extend the event structure to create a slot for them, but most users would find this operation difficult.

Fig. 6. *Illustration of how the system might suggest an improvement on the users structuring of the complex event, making use of existing representations of similar events.*

So instead, the system examines the structures of other complex events (in this case: soccer tournaments) that have been created and used in the past. It notices that some of these events have included a "Celebration" event right after the end of the final game.

So it tentatively introduces this event node, putting the questionable media under it and offering an explanation of why the new subevent seems reasonable.

If the user doesn't like the suggestion, she can ask the system to suggest other subevents in a similar way (or she can just delete the photos, if she can see that they are irrelevant).

6 Related Work

A great deal of research on support for photo annotation—mostly not involving indexing in terms of events—has yielded many ideas about effective combinations of

backend processing and interaction design (see, e.g., [5], [1], for individual contributions and [3] for a brief synthetic overview). Some of the work in this area also refers to indexing in terms of events. Some research (e.g., that of [2]) focuses on the technical aspects of event clustering. [4] likewise explore event clustering somewhat similar to the type of clustering assumed in the scenarios in this paper, also providing evidence for the viability of the sort of collaboration between user and system that is proposed here.

7 Conclusions and Next Steps

These scenarios and hypothetical examples illustrate how it may be possible and natural for untrained users to (a) benefit from an organization of media in terms of complex event structures and even (b) to create new event structures themselves, as a natural by-product of organizing their own media.

We are currently working on variants of these scenarios, which will then be presented to typical potential users, whose responses will presumably suggest desirable changes. The subsequent step will be the implementation of mockups that allow the interaction design to be tested.

These scenarios do make some strong assumptions about the capabilities of GLOCAL's backend processing, which is being developed in parallel in other parts of the GLOCAL project. Understanding of how the interaction can work helps to guide the development of the backend processing, and vice versa.

References

1. Barthelmess, P., Kaiser, E., McGee, D.: Toward content-aware multimodal tagging of personal photo collections. In: Proceedings of the Ninth International Conference on Multimodal Interfaces. pp. 122–125 (2007)
2. Cooper, M., Foote, J., Girgensohn, A., Wilcox, L.: Temporal event clustering for digital photo collections. ACM Transactions on Multimedia Computing, Communications and Applications 1(3), 269–288 (2005)
3. Hasan, T., Jameson, A.: Bridging the motivation gap for individual annotators: What can we learn from photo annotation systems? In: Proceedings of the First Workshop on Incentives for the Semantic Web at the 2008 International Semantic Web Conference. Karlsruhe, Germany (2008)
4. Suh, B., Bederson, B.B.: Semi-automatic photo annotation strategies using event based clustering and clothing based person recognition. Interacting with Computers 19, 524–544 (2007)
5. Tuffield, M.M., Harris, S., Dupplaw, D., Chakravarthy, A., Brewster, C., Gibbins, N., O'Hara, K., Ciravegna, F., Sleeman, D., Shadbolt, N., Wilks, Y.: Image annotation with Photocopain. In: Proceedings of the First International Workshop on Semantic Web Annotations for Multimedia, held at the World Wide Web Conference (2006)

Exploiting a region-based visual vocabulary towards efficient concept retrieval

Evaggelos Spyrou, Yannis Kalantidis, and Phivos Mylonas

Image, Video and Multimedia Systems Laboratory,
School of Electrical and Computer Engineering
National Technical University of Athens
9 Iroon Polytechniou Str., 157 80 Athens, Greece,
espyrou@image.ece.ntua.gr,
WWW home page: http://www.image.ece.ntua.gr/∼espyrou/

Abstract. This paper presents our approach for semantic concept retrieval based on visual characteristics of multimedia content. The former forms a crucial initial step towards efficient event detection, resulting into meaningful interpretation of available data. In the process, a visual vocabulary is constructed in order to create a representation of the visual features of still image content. This vocabulary contains the most common visual features that are encountered within each still image database and are referred to as "region types". Based on this vocabulary, a description is then formed to capture the association of a given image to all of its region types. Opposite to other methods, we do not describe an image based on all region types, but rather to a smaller representative subset. We show that the presented approach can be efficiently applied to still image retrieval when the goal is to retrieve semantically similar rather than visually similar image concepts by applying and evaluating our method to two well-known datasets.

1 Introduction

It is true that the main obstacle in order to successfully implement the task of (semantic) concept retrieval in multimedia is that the actual semantic description of image objects or even of entire image scenes, is rather difficult to grasp. Several research approaches exist in the literature and they range from text-based to content-based ones. The former tend to apply text-based retrieval algorithms to a set of usually (pre-)annotated images including keywords, tags, or image titles, as well as filenames. The latter typically apply low-level image processing and analysis techniques to extract visual features from images, whereas their scalability is questionable. Most of them are limited by the existing state-of-the-art in image understanding, in the sense that they usually take a relatively low-level approach and fall short of higher-level interpretation and knowledge.

In this paper, we shall provide our research view on modelling and exploiting visual information towards efficient semantic interpretation and retrieval of multimedia content. Our goal is to create a meaningful representation of visual features of images by constructing a visual vocabulary, which will be used at a later stage for efficient concept detection. The proposed vocabulary contains the most common region types

encountered within a large-scale image database. A model vector is then formed to capture the association of a given image to the visual dictionary. The goal of our work is to retrieve semantically similar images through the detection of semantic similar concepts within them. This means that given a query image, depicting a semantic concept, only the returned images that contain the same semantic concepts will be considered as relevant. Thus, images that appear visually similar, without containing the semantic concept of the query image will be considered irrelevant.

The idea of using a plain visual dictionary in order to quantize image features has been used widely in both image retrieval and high-level concept detection. In [1] images are segmented into regions and regions correspond to visual words based on their low level features. Moreover, in [2] the bag-of-words model is modified in order to include features which are typically lost within the quantization process. In [3], fuzziness is introduced in the process of the mapping to the visual dictionary. This way the model does not suffer from the well-known "curse of dimensionality". In [4] images are divided into regions and a joint probabilistic model is created to associate regions with concepts. In [5] a novel image representation is proposed (bag of visual synset), defined as a probabilistic relevance-consistent cluster of visual words, in which the member visual words induce similar semantic inference towards the image class. The work presented in [6] aims at generating a less ambiguous visual phrase lexicon, where a visual phrase is a meaningful spatially co-occurrent pattern of visual words. However, as it will be showed in the following sections, all above references lack significantly in comparison to the herein proposed work, both in terms of representation modelling and scalability/expressiveness.

The rest of this paper is structured as follows: Section 2 discusses the idea of using a visual vocabulary in order to quantize image features and presents the approach we adopt. Section 3 presents the algorithm we propose in order to create a model vector that will describe the visual properties of images. Experiments are presented in Section 4 and brief conclusions are finally drawn in Section 5.

2 Building a Visual Vocabulary

As it has already been mentioned, the idea of using a visual vocabulary to quantize image features has been used in many multimedia problems. In this Section we discuss the role of the visual vocabulary and we present the approach used in this work for its construction.

Given the entire set of images of a given database and their extracted low-level features, it may easily be observed that for concepts that can be characterized as "scenes" or "materials" regions that correspond to the same concept have similar low-level descriptions. Also, images that contain the same high-level concepts are typically consisted of similar regions. For example, regions that contain the concept *sky* are generally visually similar, i.e. the color of most of them should be some tone of "blue". On the other hand, images that contain *sky*, often are consisted of similar regions.

The aforementioned observations indicate that similar regions often co-exist with some high-level concepts. This means that region co-existences should be able to provide visual descriptions which can discriminate between the existence or not of certain

high-level concepts. As indicated in the bibliography, by appropriately quantizing the regions of an image dataset, we can create efficient descriptions. Thus, this work begins with the description of the approach we follow in order to create a visual vocabulary of the most common region types encountered within the data set. Afterwards, each image will be described based on a set of region types.

In every given image I_i we first apply a segmentation algorithm, which results to a set of regions R_i. The segmentation algorithm we use is a variation of the well-known RSST [7], tuned to produce a small number of regions. From each region r_{ij} of I_i we extract visual descriptors, which are then fused into a single feature vector f_i as in [8]. We choose to extract two MPEG-7 descriptors [9], namely the Scalable Color Descriptor and the Homogeneous Texture Descriptor, which have been commonly used in the bibliography in similar problems and have been proved to successfully capture color and texture features, respectively.

After the segmentation of all images of the given image dataset, a large set $\mathcal{F}$ of the feature vectors of all image regions is formed. In order to select the most common region types we apply the well-known K-means clustering algorithm on $\mathcal{F}$. The number of clusters which is obviously the number of region types N_T is selected experimentally.

We define the visual vocabulary, formed by a set of the region types as

$$T = \Big\{ w_k \Big\},\ k = 1, 2, \ldots N_T,\ w_k \subset \mathcal{F}\ , \tag{1}$$

where w_k denotes the k-th region type. We should note here that after clustering the image regions in the feature space, we chose those that lie nearest to the centroid of each cluster.

We should emphasize that although a region type does not contain conceptual semantic information, it appears to carry a higher description than a low-level descriptor; i.e. one could intuitively describe a region type as "green region with a coarse texture", but would not be necessarily able to link it to a specific concept such as *vegetation*, which neither is necessary a straightforward process, nor falls within the scope of the presented approach.

3 Construction of Model Vectors

In this Section we will use and extend the ideas presented in [10] and [11], in order to describe the visual content of a given image I_i using a model vector m_i. This vector will capture the relation of a given image with the region types of the visual vocabulary. For the construction of a model vector we will not use the exact algorithm as in [10]. Instead and for reasons that will be clarified later we will modify it, in order to fit in the problem of retrieval.

Let R_i denote the set of the regions of a given image I_i after the application of the aforementioned segmentation algorithm. Moreover, let N_i denote its cardinality and r_{ij} denote its j-th region. Let us also assume that a visual vocabulary $T = \{w_i\}$ consisting of N_T region types has been constructed following the approach discussed in Section 2.

In previous work we constructed a model vector by comparing all regions R_i of an image to all region types. For each region type, we chose to describe its association to the given image by the smallest distance to all image regions. Let

$$m_i = \{m_i(1)\ m_i(2)\ \ldots\ m_i(N_T)\}\ , \tag{2}$$

denote the model vector that describes the visual content of image I_i in terms of the visual dictionary. We calculated each coordinate as

$$m_i(j) = min_{r_{ij} \in R_i}\{d(f(w_j), f(r_{ij})\}\ , j = 1, 2, \ldots, N_T\ . \tag{3}$$

In this work, instead of m_i we calculate a modified version of the model vector which will be referred to as $\hat{m}_i$. After calculating the distances among each region r_{ij} and all the region types, let $\mathcal{W}_{ij}$ denote an ordered set that contains all the region types with an ascending order, based on their distances d_{ij} to r_{ij}, as

$$\mathcal{W}_{ij} = \{w_{ij} \mid \forall k, l \leq N_T, k \leq l : w_{ik} \leq w_{il}\}\ . \tag{4}$$

For each region r_{ij} we select its closest region types, which obviously are the first K elements of $\mathcal{W}_{ij}$. This way and for each region we define the set of its K closest region types as

$$\mathcal{W}_i^K = \{w_{ij} : j \leq K\}\ . \tag{5}$$

To construct a model vector $\hat{m}_i$, instead of using the whole visual vocabulary, we choose to use an appropriate subset. This will be the union of all ordered sets $\mathcal{W}_i^K$

$$W^K = \bigcup_i \mathcal{W}_i^K\ . \tag{6}$$

This way, the set W^K consists of the closest region types of the visual dictionary to all image regions. We will construct the model vector using this set, instead of the set of all region types. Again,

$$\hat{m}_i = \{\hat{m}_i(1)\ \hat{m}_i(2)\ \ldots\ \hat{m}_i(N_T)\}\ . \tag{7}$$

We define as $\hat{m}_i(j)$ the minimum distance of a region type to all image regions, thus it is calculated as

$$\hat{m}_i(j) = \begin{cases} min\{d(f(w_{ij}), f(r_{ij}))\} & \text{if } w_{ij} \in W^K \\ 0 & \text{else} \end{cases}\ . \tag{8}$$

If we compare Eq.8 with Eq.3 we can easily observe that the resulting model vector $\hat{m}_i$, it becomes obvious that it is not constructed based on the full visual vocabulary. Instead, our method selects an appropriate subset.

The method we followed in order to construct $\hat{m}_i$ contains an intermediate step when compared to the one for the construction m_i. The latter has been used successfully in a high-level concept detection problem. The use of a neural network classifier practically assigned weights to each region type. Thus, those that were not useful for

the detection of a certain concept had been ignored. However, in the case of the retrieval we do not assign any weights to the region types. This means that if the model vector consisted from all region types, those with a small distance to the image regions would act as noise. In this case, retrieval would fail, as many images would have similar descriptions despite being significantly different in terms of their visual content.

To further explain the aforementioned statement, we also give a semantic explanation on why the choice of K instead of one region types for each image region is meaningful and crucial. From a simple observation of a given data set, but also intuitively, it is obvious that many high-level concepts are visually similar to more than one region types. For example, let us assume that the concept *sand* appears "brown" in an image of the database and "light brown" in another. Let us now consider a query image containing the concept *sand*. If the given visual vocabulary contains both a "brown" and a "light brown" region types, in order to retrieve both the aforementioned images of the database, their description should contain both region types and not the most similar. Thus, this way we tackle the problem of quantization.

An artificial example of the K most similar region types to each image region is depicted in Fig.1, for the case of $K = 2$.

Fig. 1. A segmented image and the 2 most similar region types to each region.

4 Experimental Results

In order to test the efficiency of the proposed approach, we selected two descriptive image collections, one dataset[1] created by Oliva and Torralba and one comprised by images derived from the Corel image collection [12]. The first collection was used in a scene recognition problem and is annotated both globally and at a region level. A sample of the first dataset is depicted in Fig.2. We used only the global annotations for 2688 images, as well as all 8 categories of the dataset to evaluate our approach, namely: *coast*, *mountain*, *forest*, *open country*, *street*, *inside city*, *tall buildings* and *highways*.

[1] `http://people.csail.mit.edu/torralba/code/spatialenvelope/`

A similar procedure was followed for the second dataset, containing 750 images and 6 concepts, namely: *vegetation*, *road*, *sand*, *sea*, *sky* and *wave*.

In order to meaningfully evaluate our work, we calculated the mean Average Precision (mAP) measure for each concept. At this point we should remind the reader that given a query image belonging to a certain semantic category, only those images within the results that belong to the same category were considered to be relevant. In addition, the well-known Euclidean distance was applied in order to compare the visual features between regions and region types. The mAP that has been achieved for several visual vocabularies and for several cases of the region types that were considered to be similar to the image regions is depicted in the following Tables.

	Nt=150, K=1	Nt=150, K=2	Nt=150, K=4	Nt=270, K=1	Nt=270, K=2	Nt=270, K=5
coast	0.317	0.336	0.360	0.460	**0.660**	0.450
mountain	0.320	0.287	0.311	0.317	0.428	**0.459**
forest	**0.275**	0.146	0.170	0.270	0.230	0.180
open country	0.134	0.109	0.133	0.121	0.111	**0.158**
street	0.063	0.106	0.130	0.060	0.090	**0.140**
inside city	0.094	0.098	0.121	0.145	0.130	**0.204**
tall buildings	0.084	0.081	0.105	0.124	0.131	**0.152**
highways	0.067	0.066	0.090	0.060	0.100	**0.140**

Table 1. The mAP calculated for six different visual vocabularies, whose size is denoted as N_T and for six cases of closest region types K for the **Oliva/Torralba dataset**.

Table 1 summarizes the results of the application of our method to the first aforementioned dataset. We may observe that the proposed retrieval algorithm achieved its best performance in concepts *coast* and *mountain*. Concept *forest* appears to be somewhere in the middle range, whereas mAPs for concepts *open country*, *street*, *inside city*, *tall buildings* and *highways* were not as high, with all of them ranging equal or below value 0.20. This result can be explained if we consider the visual properties of these concepts. In the case of *coast* and *mountain*, the segmentation algorithm created regions which can easily discriminate those concepts, while in the images depicting the rest of the concepts, segmented regions are more similar to each other and thus not discriminated thoroughly.

We also investigated the effect of the number K of region types which are considered to be similar to the image regions, to the mAP that is achieved. Fig. 3 depicts the evolution of mAP vs. K and N_T for all sets of concepts utilized. It is obvious that in the case of low mAPs (e.g. for all 5 concepts mentioned above), mAP values increase for higher values of K, while we observe an intermediate behavior for concepts with significantly better mAPs like *coast* or *forest*. This leads to the conclusion that the concepts that may be considered as intuitively "simpler", can be efficiently described and retrieved by a smaller value K of their closest region types.

Table 2 presents the corresponding results of the application of our method to the second dataset. In this case the proposed retrieval algorithm worked more efficiently,

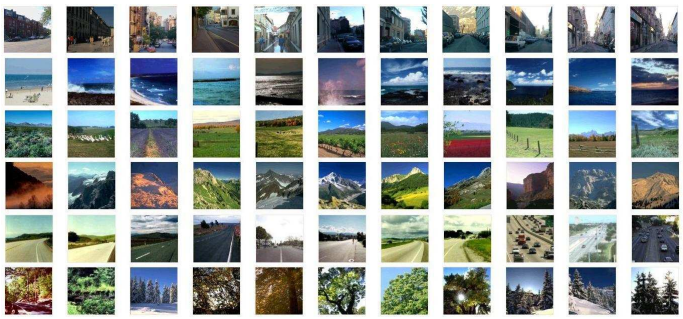

Fig. 2. A subset of the Torralba dataset.

Fig. 3. A subset of the Corel dataset.

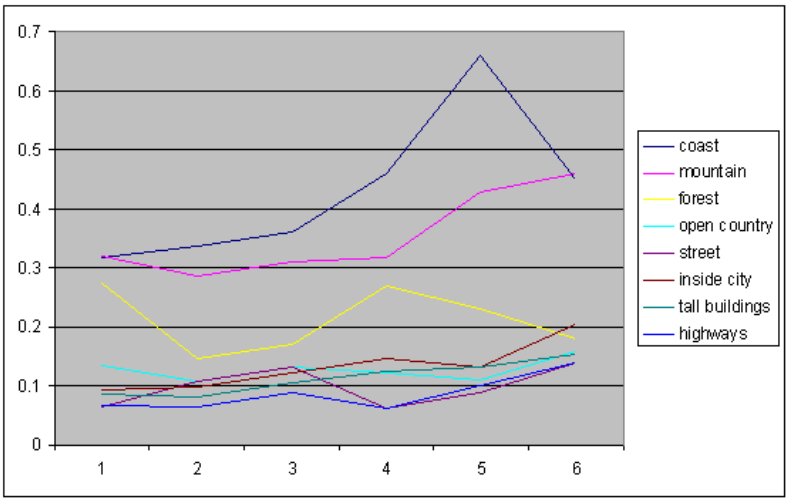

Fig. 4. The achieved mAP for all Torralba concepts, while increasing the number of the closest region types K and the size of visual vocabularies N_T.

especially with respect to concept *sea*, which is to be explained due to the actual nature of the concepts themselves. More specifically, concepts *vegetation* and *sky* performed also very well (i.e. mAP above 0.70), whereas mAP values obtained for concepts *road* and *sand* were average. On the other hand, mAPs for concept *wave* was not as high. This result can again be explained based on the actual visual properties of the particular

	Nt=150, K=1	Nt=150, K=2	Nt=150, K=4	Nt=270, K=1	Nt=270, K=2	Nt=270, K=5
vegetation	0.641	0.664	0.688	**0.721**	0.688	0.670
road	0.581	0.605	0.628	0.607	0.629	**0.657**
sand	0.497	0.520	0.544	**0.574**	0.544	0.552
sea	0.815	0.838	0.862	**0.891**	0.863	0.786
sky	0.618	0.641	0.665	0.673	0.667	**0.734**
wave	0.405	0.429	0.453	**0.457**	0.451	0.401

Table 2. The mAP calculated for six different visual vocabularies, whose size is denoted as N_T and for six cases of closest region types K for the **Corel dataset**.

concept, i.e. a *wave* is difficult to segment and discriminate in a visual manner. Fig. 5 depicts again the evolution of mAP vs. K and N_T for all Corel concepts. In this case, we observe a more unified distribution of mAPs for higher values of K, which results to rather small variations to the actual values, e.g. ranging from a low of 0.401 up to 0.457 for concept *wave* or a low of 0.786 up to 0.891 for concept *sea*.

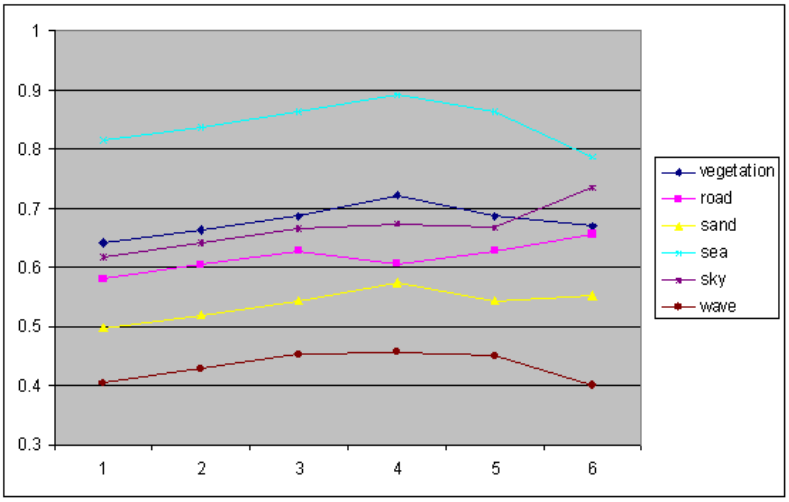

Fig. 5. The achieved mAP for all Corel concepts, while increasing the number of the closest region types K and the size of visual vocabularies N_T.

5 Conclusions

In this paper we presented an approach for semantic image retrieval by exploiting a region-based visual vocabulary. More specifically, we introduced an enhanced bag-of-

words model for capturing the visual properties of images and instead of using the entire vocabulary, we selected a meaningful subset consisting of the closest region types to the image regions. This led to a simple yet effective representation of the image features that allow for efficient retrieval of semantic concepts. Early experimental results on two well-known still image datasets are promising.

References

1. Duygulu, P., Barnard, K., De Freitas, J., Forsyth, D.: Object recognition as machine translation: Learning a lexicon for a fixed image vocabulary. Lecture Notes in Computer science (2002)
2. Philbin, J., Chum, O., Isard, M., Sivic, J., Zisserman, A.: Lost in quantization: Improving particular object retrieval in large scale image databases. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR). (2008)
3. van Gemert, J., Geusebroek, J., Veenman, C., Smeulders, A.: Kernel codebooks for scene categorization. In: European Conference on Computer Vision (ECCV), Springer (2008)
4. Lavrenko, V., Manmatha, R., Jeon, J.: A model for learning the semantics of pictures. In: in NIPS, MIT Press (2003)
5. Zheng, Y., Neo, S., Chua, T., Tian, Q.: Object-Based Image Retrieval Beyond Visual Appearances. Lecture Notes in Computer Science **4903** (2008) 13
6. Yuan, J., Wu, Y., Yang, M.: Discovery of collocation patterns: from visual words to visual phrases. In: Proc. IEEE Conf. on Computer Vision and Pattern Recognition. Volume 1., Citeseer (2007)
7. Avrithis, Y., Doulamis, A., Doulamis, N., Kollias, S.: A Stochastic Framework for Optimal Key Frame Extraction from MPEG Video Databases. Computer Vision and Image Understanding **75**(1) (1999) 3–24
8. Spyrou, E., Le Borgne, H., Mailis, T., Cooke, E., Avrithis, Y., O Connor, N.: Fusing mpeg-7 visual descriptors for image classification. In: International Conference on Artificial Neural Networks (ICANN). (2005)
9. Chang, S., Sikora, T., Purl, A.: Overview of the MPEG-7 Standard. IEEE Transactions on Circuits and Systems for Video Technology **11**(6) (2001) 688–695
10. Spyrou, E., Tolias, G., Mylonas, P., Avrithis, Y.: Concept detection and keyframe extraction using a visual thesaurus. Multimedia Tools and Applications **41**(3) (2009) 337–373
11. Mylonas, P., Spyrou, E., Avrithis, Y., Kollias, S.: Using Visual Context and Region Semantics for High-Level Concept Detection. IEEE Transactions on Multimedia **11**(2) (2009) 229
12. J.Z. Wang, J. Li, G.W.: SIMPLIcity: Semanticssensitive Integrated Matching for Picture LIbraries. In: IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. 23, No.9, IEEE (2001) 947–963