



Detection, Representation, and Exploitation
of Events in the Semantic Web

Workshop in conjunction with the
10th International Semantic Web Conference 2011
Bonn, Germany, 23 October 2011

Edited by:
Marieke van Erp
Willem Robert van Hage
Laura Hollink
Anthony Jameson
Raphaël Troncy

DeRiVE 2011 is sponsored by:

_textkernel

Introduction

The goal of this workshop is to strengthen the participation of the Semantic Web community in the recent surge of research on the use of *events* as a key concept for representing knowledge and organising and structuring media on the web. The workshop call for papers invited contributions to three central questions, and the discussion at the workshop itself will aim to formulate answers to these questions that advance and reflect the current state of understanding. Each paper accepted for presentation at the workshop addresses at least one question explicitly, and several are accompanied by a system demonstration. The workshop concludes with a challenge competition in which systems that may address any of the three main questions make use of RDF datasets of event-related media such as EventMedia¹. The challenge prize sponsored by .textkernel².

Why the Topic Is of Particular Interest Now

In recent years, researchers from several communities involved in aspects of the web have begun to realise the potential benefits of assigning an important role to *events* in the representation and organisation of knowledge and media—benefits which can be compared to those of representing entities such as persons or locations instead of just dealing with more superficial objects such as proper names and geographical coordinates. While a good deal of relevant research—for example, on the modelling of events—has been done in the semantic web community, a lot of complementary research has been done in other, partially overlapping communities, such as those involved in multimedia processing and information retrieval. The goal of this workshop is to advance research on this general topic within the semantic web community, both building on existing semantic web work and integrating results and methods from other areas, while focusing on issues of special importance for the semantic web.

¹<http://thedatahub.org/dataset/event-media>

²<http://www.textkernel.com/>

Questions Addressed

The intended outcome of the workshop is to advance understanding of three high-level questions about the role of events in the semantic web. Below we reproduce each of the three main questions (and associated more specific questions) that were included in the call for papers for the workshop. We then indicate how the papers accepted for presentation in the corresponding sections of the workshop address the respective questions.

Question 1: How can events be detected and extracted for the semantic web?

More Specific Questions

- How can events be recognised in particular types of material on the web, such as calendars of public events, social networks, microblogging sites, semantic wikis, and normal web pages?
- How can the quality and veracity of the events mentioned in noisy microblogging sites such as TWITTER be verified?
- How can a system recognise when a newly detected event is the same as a previously detected and represented event?
- How can a system recognise a complex event that comprises separately recognisable subevents?

Contributions of Accepted Papers

One of the core obstacles for using events is that they are often difficult to detect. In text, one can describe and refer to events in a myriad of ways. In video, it is difficult to discern which frames denote interesting or significant events and which are merely fillers. For the event detection track, we received submissions that address a variety of issues in event detection. The papers we have accepted can be divided into two types: automatic event detection approaches (for text) and crowdsourcing approaches (for video and images).

The paper *An Overview of Event Extraction from Text*, by Frederik Hogenboom, Flavius Frasincar, Uzey Kaymak and Franciska de Jong, provides a thorough overview of event detection approaches from text and makes recommendations for choosing the right approach for different problems. An example of a data-driven event detection approach is presented in *Using Semantic Role Labeling to Extract Events from Wikipedia*, by Peter Exner and Pierre Nugues. By using standard text mining tools in a cascaded event detection pipeline, the authors show how they can extract event elements with reasonable precision and recall.

As image and video processing have yet to reach a state where they can be used for event detection, the papers about detecting events from videos and images rely on crowdsourcing. *Crowdsourcing Event Detection in YouTube Videos*

by Thomas Steiner, Ruben Verborgh, Rik Van de Walle, Michael Hausenblas and Joaquim Gabarro Valles describes a three-tiered approach that uses visual processing combined with users' clicking behavior as well as the textual meta-data that accompanies the video to identify different events. *Clues of Personal Events in Online Photo Sharing*, by Pierre Andrews, Javier Paniagua and Fausto Giunchiglia, identifies events by classifying how users organize their photos in albums. By classifying album titles, the authors show it is possible to identify photos about trips or different types of celebrations.

Question 2: How can events be modelled and represented in the semantic web?

More Specific Questions

- How can we improve the interoperability of the various event vocabularies such as EVENT,³ LODE,⁴ SEM,⁵ and F?⁶
- How can aspects of existing event representations developed in other communities be adapted to the needs of the semantic web?
- What are the requirements for event representations for qualitatively different types of events (e.g., historical events such as wars; cultural events such as upcoming concerts; personal events such as family vacations)?
- To what extent can/should a unified event model be employed for such different types of events?

Contributions of Accepted Papers

The term “event” has several meanings. It is used to mean both phenomena that have happened (e.g., things reported in news articles or explained by historians) and phenomena that are scheduled to happen (e.g., things put in calendars and datebooks). Events are also a natural way for referring to any observable occurrence grouping persons, places, times and activities that can be described. Hence, a number of different RDFS+OWL ontologies providing classes and properties for describing the “factual” aspects of events (*What* happened, *Where* did it happen, *When* did it happen, and *Who* was involved) have been proposed and compared.

The papers we have accepted can again be divided into two types: the ones that have been applied in practical applications such as museum narratives or e-Science and the ones who present more theoretical work for representing relationships between events. Paul Mulholland, Annika Wolff, Trevor Collins and Zdenek Zdrahal in *An event-based approach to describing and understanding*

³<http://motools.sourceforge.net/event/event.html>

⁴<http://linkedevents.org/ontology/>

⁵<http://semanticweb.cs.vu.nl/2009/11/sem/>

⁶<http://isweb.uni-koblenz.de/eventmodel>

museum narratives presents the Curatorial Ontology (CO) for describing curatorial narratives. This ontology draws on structuralist theories that distinguish between story (i.e. what can be told), plot (i.e. an interpretation of the story) and narrative (i.e. its presentational form). Lianli Gao and Jane Hunter in *Publishing, Linking and Annotating Events via Interactive Timelines: an Earth Sciences Case Study* describe two ontologies: Event, Timeline, Annotation and TemporalRelation for relationships between events. They also developed a semantic annotation system that enables the discovery, retrieval and ontology-based markup of such event data via interactive timelines.

Ilaria Corda, Brandon Bennett and Vania Dimitrova in *A Logical Model of an Event Ontology for Exploring Connections in Historical Domains* describe a formal model for representing events and comparing temporal dimensions as the backbone for drawing connections and exploring relationships between happenings. Stasinios Konstantopoulos in *Using On-the-Fly Pattern Transformation to Serve Multi-Faceted Event Metadata* proposes the SYNC3 Ontology which is based on both the DOLCE Ultralite ontology and the F model and contains a number of conversion rules to the common LODDE ontology.

Question 3: How can events be exploited for the provision of new or improved services?

More Specific Questions

- How can event representations be better exploited in support of activities like semantic annotation, semantic search, and semantically enhanced browsing?
- What application areas for semantic technologies can benefit from an increased use of event representations?
- How can we improve existing methods for visualising event representations and enabling users to interact with them in semantic web user interfaces?
- What requirements for event detection and representation methods (Questions 1 and 2 above) are implied by advances in methods for exploiting events?

Contributions of Accepted Papers

The four accepted papers for this part of the workshop mostly contribute new ideas about forms of exploitation and application areas, though there is also some attention to interaction design and visualisation.

Linked Open Piracy, by Willem R. van Hage, Véronique Malaisé, and Marieke van Erp, shows in detail how formally represented events can be used to support the creation of mashups and visual analytics. Referring to the specific application goal of analysing pirate attacks on shipping, the authors show how piracy reports intended for human reading can be augmented with semantic representations that in turn make possible a variety of visualisations and statistical analyses.

A different application area—web archiving—is discussed in *Using Events for Content Appraisal and Selection in Web Archives*, by Thomas Risse, Stefan Dietze, Diana Maynard, Nina Tahmasebi, and Wim Peters. The authors address the goal of archiving material from the web in a relatively structured and selective way, aiming to capture material related to events (and other entities) in a way reminiscent of a “community memory”, exploiting the wisdom of the crowd. A good deal of the paper discusses strategies for overcoming the challenges for event extraction and detection that arise when this goal is pursued.

An application in the area of cultural heritage is presented in *Hacking History: Automatic Historical Event Extraction for Enriching Cultural Heritage Multimedia Collections*, by Roxane Segers, Marieke van Erp, Lourens van der Meij, Lora Aroyo, Guus Schreiber, Bob Wielinga, Jacco van Ossenbruggen, Geertje Jacobs, and Johan Oomen. The authors show how linking cultural artifacts to explicitly modelled events (and other entities) can support new forms of browsing and searching. The paper also discusses the challenges involved in extracting the relevant historical events from texts.

More attention to new forms of interaction with event representations is found in the paper *New Forms of Interaction With Hierarchically Structured Events*, by Sven Buschbeck, Anthony Jameson, and Tanja Schneeberger. The user interface presented differs from the more familiar timelines in that (a) it supports interaction with arbitrarily deep hierarchies of events linked via a “subevent” relation and (b) it offers functionality inspired by mind mapping applications to enable flexible browsing, searching, and media curation in a repository of events and associated media.

Challenge Competition

For the challenge part of the workshop, a dataset was made available consisting of over 100,000 events from the EventMedia LOD dataset (including events from Last.fm, Eventful, and Upcoming). Next to events, they contain artists, venues and location, description and time information. Some links between the instances of these three sources are provided.

This challenge dataset is intended to encourage participation by researchers who do not have an event dataset at their disposal and to increase shared understanding of the issues involved in working with data of this type. The application that makes best use of the provided datasets was awarded *The DeRiVE 2011 Challenge Prize*, which was sponsored by _textkernel. Submissions are judged by their (a) scientific contribution and (b) societal impact (e.g., how much the work contributes to useful applications by providing data or services).

Contributions of Contesters

The three accepted competition entries deal all with event background knowledge in some way. Two of them build new links to related concepts while one investigates how complex queries that use these relations to background knowledge can be executed in real time.

Kristian Slabbekoorn, Laura Hollink and Geert-Jan Houben study the problem of linking data to large, heterogeneous Linked Data sets in their paper *Domain-aware matching of events to DBpedia*. They use DBpedia Spotlight to create a baseline of matches between the artists in the EventMedia dataset and DBpedia resources. They show that knowledge of the domain in terms of relevant DBpedia categories and classes can increase the quality of the matches, and that this domain knowledge can be automatically derived. The resulting 19,840 links to DBpedia are made available for download.

In *Events Retrieval Using Enhanced Semantic Web Knowledge*, Pierre-Yves Vandenbussche and Charles Teissèdre demonstrate the benefit data enrichment in a retrieval system. They link the events to several external sources: city, country and address information, images associated to the events, and links to people and bands in DBpedia. They build a retrieval system that parses natural language queries containing agents, places, and complex temporal expressions. The resulting events and their images are visualised on a timeline.

In *Fusion of Event Stream and Background Knowledge for Semantic-Enabled Complex Event Processing*, Kia Teymourian, Malte Rohde, Ahmad Hassan and Adrian Paschke present research on how to apply reactive semantic complex event processing to event streams. By means of query pre-processing given a static knowledge base their Prova-based system is able to answer complex queries about events in real time.

Programme Committee

The following colleagues kindly served in the workshop's program committee. Their joint expertise covers all of the questions addressed in the workshop, and they reflect the range of relevant scientific communities.

- Jans Aasman, Franz Inc.
- Klaus Berberich, Max Planck Institute for Computer Science, Germany
- Fausto Giunchiglia, University of Trento, Italy
- Christian Hirsch, University of Auckland, New Zealand
- Ramesh Jain, University of California, Irvine, USA
- Krzysztof Janowicz, Pennsylvania State University, U.S.A.
- Jobst Löffler, Fraunhofer IAIS, Germany
- Marco Pennacchiotti, Yahoo! Labs, U.S.A.
- Yves Raimond, BBC Future Media & Technology, UK
- Ansgar Scherp, Universität Koblenz-Landau, Germany
- Nicu Sebe, University of Trento, Italy

- Ryan Shaw, University of North Carolina, U.S.A.
- Michael Sintek, DFKI, Germany
- Alan Smeaton, Dublin City University, Ireland
- Nenad Stojanovic, Forschungszentrum Informatik, Germany
- Denis Teyssou, AFP, France

Organising Committee

This workshop was organised by:

- Marieke van Erp, VU University Amsterdam, The Netherlands
- Willem Robert van Hage, VU University Amsterdam, The Netherlands
- Laura Hollink, Delft University of Technology, The Netherlands
- Anthony Jameson, DFKI, Germany
- Raphaël Troncy, EURECOM, France