Linking Events with Media

Raphaël Troncy
EURECOM
Sophia Antipolis, France
raphael.troncy@eurecom.fr

Bartosz Malocha
EURECOM
Sophia Antipolis, France
bartosz.malocha@eurecom.fr

André T. S. Fialho*
CWI Amsterdam
The Netherlands
andre.fialho@cwi.nl

ABSTRACT

We present a large dataset composed of events descriptions together with media descriptions associated with these events and interlinked with the larger Linked Open Data cloud. We are constructing a web-based environment that allows users to explore and select events, to inspect associated media, and to discover meaningful, surprising or entertaining connections between events, media and people participating in events. The dataset is obtained from three large public event directories (last.fm, eventful and upcoming) represented with the LODE ontology and from large media directories (flickr, youtube) represented with the Media Ontology. We describe how the data has been converted, interlinked and published following the best practices of the Semantic Web community.

Categories and Subject Descriptors

H.5.1 [**Multimedia Information System**]: Audio, Video and Hypertext Interactive Systems; I.7.2 [**Document Preparation**]: Languages and systems, Markup languages, Multi/ mixed media, Standards

General Terms

Languages, Hyperlinks, Web, URI, HTTP

Keywords

Events, LODE, Media Ontology, dataset

1. INTRODUCTION

Events are a natural way for referring to any observable occurrence grouping persons, places, times and activities that can be described [7, 5, 1]. Events are also observable experiences that are often documented by people 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. In the context of the Petamedia[1] Network of Excellence, we are 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.

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 [1]. 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.

In this paper, we present how we intend to represent description of events using the LODE ontology which ensures interoperability among various event ontologies (Section 2). We describe the data scraping and interlinking process as well as a large SKOS taxonomy of event categories (Section 3). We provide interface mockups to illustrate the types of task support and expected functionality we will be providing in the coming weeks based on this dataset (Section 4). Finally, we give our conclusions and outline future work in Section 5.

2. THE LODE ONTOLOGY

The LODE ontology[2] is a minimal model that encapsulates the most useful properties for describing events [5]. 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,

*André Fialho is also affiliated with Delft University of Technology in Delft, The Netherlands.

I-SEMANTICS 2010 September 1-3, 2010, Graz, Austria

[1] http://www.petamedia.eu
[2] http://linkedevents.org/ontology/

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 rhetorically produced will depend on one's epistemological stance. We enhance the LODE descriptions with properties for categorizing events and for relating them to other events through parthood or causal relations using the Descriptions and Situations approach of the Event-Model-F [4].

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.

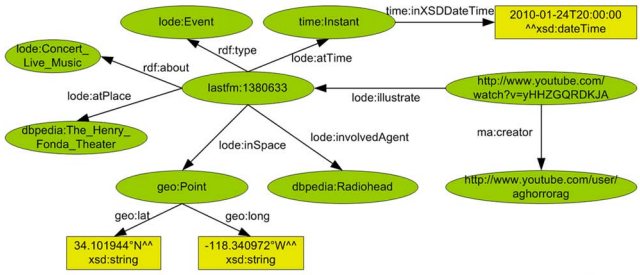

Figure 1: The *Radiohead Haiti Relief Concert* described with LODE

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. Hence, the ontology contains numerous OWL axioms stating classes and properties equivalence between models such as MO [3], CIDOC-CRM, DOLCE, SEM [6] to name a few. 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. DATA SCRAPING AND INTERLINKING

In this section, we detail how the data from events and media directories is scraped and interlinked. Furthermore, we describe how we build a large SKOS thesaurus of event categories.

3.1 Event Directories

We first explore the overlap in metadata between four popular web sites, namely Flickr as a hosting web site for photos and videos and Last.fm, Eventful and Upcoming as a documentation of past and upcoming events. Explicit relationships between events and photos exist using machine tags such as `lastfm:event=XXX`. Hence, we have been able to convert the description of more than 1.7 million photos which are indexed by nearly 140.000 events.

We use the Last.fm, Eventful and Upcoming APIs to convert each event description into the LODE ontology (Section 2). We mint new URIs into our own namespace for events (`http://data.linkedevents.org/event/`), agents (`http://data.linkedevents.org/agent/`) and locations (`http://data.linkedevents.org/location/`). A graph representing an event is composed of the type of the event, a full text description, the agents (e.g. artists) involved, a date (instant or interval represented with OWL Time [2]), a location in terms of both geographical coordinates and a URI denoting the venue and users participation. A graph representing an agent or a location is composed of a label and a description (e.g. the artist's biography).

Event directories have overlap in their coverage. We interlink these events descriptions when they involve the same agents at the same date or when they happen at the same venue at the same date. We invoke additional semantic web lookup services such as dbpedia, geonames and freebase in order to enrich the descriptions of the agents and the locations. Hence, the agent URI which has for label "Radiohead" is interlinked with the dbpedia URI (`http://dbpedia.org/page/Radiohead`) 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 (`http://data.nytimes.com/N12964944623934882292`) which provides information about the 38 associated articles from this newspaper to this band. The venue has also been converted into a dbpedia URI (`http://dbpedia.org/page/The_Henry_Fonda_Theater`) but has been augmented with geo-coordinates from last.fm which was not originally present in dbpedia thus increasing the amount of information available in the LOD cloud for the benefit of all semantic web applications. We observe that a few venues have been interlinked with the LOD cloud. We are now investigating further linkage with Foursquare which has a much broader coverage of event venues. 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 simple user tasks such as searching and browsing enriched media collections.

3.2 Media Directories

The Ontology for Media Resource currently developed by W3C is a core vocabulary which covers basic metadata properties to describe media resources[3]. It also contains a formal set of axioms defining mapping between different metadata formats for multimedia. We use this ontology together with properties from SIOC, FOAF and Dublin Core to convert into RDF the Flickr photo descriptions (Figure 2). The link between the media and the event is realized through the `lode:illustrate` property, while more information about the `sioc:UserAccount` can be attached to his URI. 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 the video. We use the patterns defined in the Event-Model-F to represent the actual role of various users for a given media: the uploader of the video, the user who has tagged, or commented the video, etc. In summary, using these patterns, we can extend the LODE and Media ontologies with provenance information, making

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

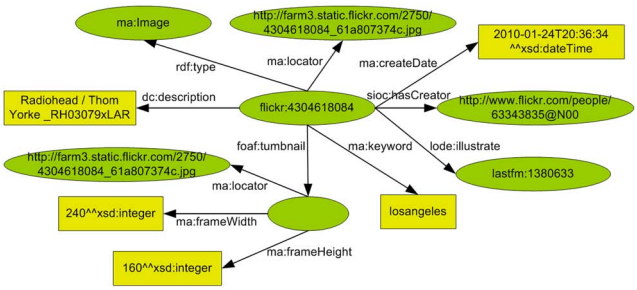

Figure 2: A photo taken at the *Radiohead Haiti Relief Concert* described with the Media Ontology

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.

3.3 Event Categories

Events are generally categorized in lightweight taxonomies that provide facets when browsing event directories. We have manually analyzed the taxonomy used in various sites, namely facebook, eventful, upcoming, zevents, linkedin, eventbrite and ticketmaster, and used card sorting techniques in order to build a rich SKOS thesaurus of event categories. This SKOS thesaurus contains axioms expressing mappings relationships with these taxonomies while the terms are defined in our own namespace (`http://data.linkedevents.org/category/`). The top level categories are Sports, Music, Food, Arts, Movies, Family, Social Gathering, Community and Professional but alignment with other classification such as the IPTC News Codes for sports of the last.fm genres for music is also provided.

	Event	Agent	Location	Media	User
Last.fm	37,647	50,151	16,471	1,393,039	18,542
Eventful	37,647	6,543	14,576	52	12
Upcoming	13,114		7,330	347,959	4,518

Table 1: Number of event/agent/location and media/user descriptions in the dataset

Overall, the dataset collected contains more than 30 million triples. A dump is available at (`http://www.eurecom.fr/~troncy/ldtc2010/`) while a temporary SPARQL endpoint is available at (`http://data.linkedevents.org/sparql`).

4. APPLICATION

In this section, we briefly illustrate initial interface possibilities derived from a user study described in [1]. Unsurprisingly, the sketches below correspond to the basic properties defined in the LODE ontology.

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[4] to represent clustered result sets or convey sorting by size on different contexts (Figure 3a).

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

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

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 hotspots on a map, indicate visual cues of popularity according to number of attendees.

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

Figure 4: Interface illustrating an event instance view for a Radiohead concert

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.

5. CONCLUSION AND FUTURE WORK

In this paper, we have shown how linked data technologies can be used for integrating information contained in event and media directories. We used the LODE and Media Ontology respectively for expressing linked data description of events and photos. Ultimately, we aim at providing an event-based environment for users to explore, annotate and share media and we present some sketches of user interfaces that we will develop in the coming weeks.

We are currently consolidating and cleaning our dataset with more sources (e.g. YouTube videos) and more linkage (e.g. description of recurring event, artist and venue from hubs such as freebase or dbpedia). We intend to provide soon user participation at events from public Foursquare check-in or live Tweets. Our priority is also to express the right licensing and attribution information to the data that has been rdf-ized. Finally, we will release soon a voiD description of the complete dataset. We truly believe that multimedia will then be finally added back to the Semantic Web.

6. ACKNOWLEDGMENTS

The authors would like to thank Lynda Hardman (CWI), Carsten Saathoff (WeST Institute), Ansgar Scherp (WeST Institute) and Ryan Shaw (Berkeley University) for fruitful discussions on the infrastructure and the design of the EventMedia interfaces. The research leading to this paper was supported by the European Commission under contract FP7-216444, Petamedia Peer-to-peer Tagged Media.

7. REFERENCES

[1] A. Fialho, R. Troncy, L. Hardman, C. Saathoff, and A. Scherp. What's on this evening? Designing User Support for Event-based Annotation and Exploration of Media. In *1st International Workshop on EVENTS - Recognising and tracking events on the Web and in real life (EVENTS'10)*, pages 40–54, Athens, Greece, 2010.

[2] J. Hobbs and F. Pan. Time Ontology in OWL. W3C Working Draft, 2006. `http://www.w3.org/TR/owl-time`.

[3] Y. Raimond, S. Abdallah, M. Sandler, and F. Giasson. The Music Ontology. In *8th International Conference on Music Information Retrieval (ISMIR'07)*, Vienna, Austria, 2007.

[4] 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.

[5] R. Shaw, R. Troncy, and L. Hardman. LODE: Linking Open Descriptions Of Events. In *4th Asian Semantic Web Conference (ASWC'09)*, 2009.

[6] W. van Hage, V. Malaisé, G. de Vries, G. Schreiber, and M. van Someren. Combining Ship Trajectories and Semantics with the Simple Event Model (SEM). In *1st ACM International Workshop on Events in Multimedia (EiMM'09)*, Beijing, China, 2009.

[7] U. Westermann and R. Jain. Toward a Common Event Model for Multimedia Applications. *IEEE MultiMedia*, 14(1):19–29, 2007.