

Experiencing Events through User-Generated Media

Raphaël Troncy¹, André Fialho^{*2}, Lynda Hardman² and Carsten Saathoff³

¹ EURECOM, Sophia Antipolis, France, <raphael.troncy@eurecom.fr>

² CWI, Amsterdam, The Netherlands, <firstName.lastName@cwi.nl>

³ WeST Institute, Koblenz, Germany, <saathoff@uni-koblenz.de>

Abstract. Large numbers of websites contain (human-readable) information about scheduled events, of which some may display media captured at these events. This information is, however, often incomplete and always locked into the sites. This prevents users from creating overviews of media associated with an event from multiple websites. We carried out exploratory user studies with potential end-users to guide the design of a web-based environment for supporting event-based services. Based on our results, our goal is to provide support for exploring and selecting events and associated media, and for discovering meaningful, surprising or entertaining connections between events, media and participants by consuming linked data. We assembled a large collection of event and associated media descriptions, which we interlinked with the Linked Open Data cloud. The dataset is obtained from three large public event directories (last.fm, eventful, upcoming) represented with the LOD ontology and from large media directories (flickr, youtube) represented with the Media Ontology. We present the results from the user studies, the conversion, interlinking and publication of the data following the best practices of the Semantic Web community, and our initial application design.

1 Introduction

As with all developing technologies, it is difficult to identify novel user needs that can be satisfied with emerging semantic web technologies. Large scale data integration is often cited as an example where linked data technologies prove to be useful. On the contrary, end-user applications that benefit from these technologies in order to better support users tasks have yet to come. Hence, the community is now putting a lot of effort into developing end-user applications that benefit from consuming linked data and end-user interfaces that go beyond generic linked data browsers. We echo the opinion expressed in a recent provocative statement: *User Interfaces for Semantic Web: Do They Have to Be Ugly?*⁴, and we also argue that unified interfaces for data access is a myth. In this paper, we present a method to guide the design and implementation of a particular system – an event-based environment for users to explore, annotate and

^{*} André Fialho is also affiliated with Delft University of Technology, The Netherlands.

⁴ Andraz Tori, Semantic Technology Conference 2010.

share media, in response to well-identified user needs: relive experiences based on media and background knowledge and support decision making for attending upcoming events, in the context of sparse data locked in various websites.

Events are a natural way for referring to any observable occurrence grouping persons, places, times and activities that can be described [4]. 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. Our goal is to design an application that support users in interacting with events and discover meaningful, entertaining or surprising relationships amongst them. Various websites such as eventful.com, upcoming.org, last.fm/events and facebook.com/events to name a few, provide already interfaces to some of these functionalities but these services have sometimes overlap in terms of coverage of past and upcoming events and provide each their own social networks features to support users in sharing and deciding upon attending events. The information about the events, the social connections and the representative media are therefore all spread and locked in amongst these services providing limited event coverage and no interoperability of the description. We advocate the use of linked data technologies to aggregate these heterogeneous sources of information and we show in this paper how to model and interlink the data.

The work reported here uses an explorative user-centered design approach. We first perform a user study where users are asked about real-world tasks they would like to carry out in order to extract functional and data requirements (section 2). We then select precise tasks we wish to support and we carried out a focus group study in order to ask users opinions and observe how they can realize these scenario using specific technologies that they are familiar with (section 3). Lack of coverage for all event directories and frustration of being locked in a particular site or social network being the recurrent issues, we propose to use linked data technologies as a unified data model for integrating heterogeneous information. We briefly describe how we represent description of events using the LOD ontology and we detail the data scraping and interlinking process together with a large SKOS taxonomy of event categories (Section 4). We present our final application and we provide interfaces based on this dataset to illustrate the functionalities supported (Section 5). Finally, we give our conclusions and outline future work in Section 6.

2 Exploratory User Studies

We follow a user-centered design process done through interaction with potential end-users at different stages of development. The following sections describe two exploratory studies that assess user needs and identify behaviors while discovering, exploring and sharing events with the support of user generated media. Initially, we conducted a user survey and two focus-group sessions to understand end-users' event-related experiences and to collect insights about existing web-based technologies that support related activities [1].

2.1 Method

The survey was completed by 28 participants (11 females) with a mean age of 27 (range 23-47). Participants were mostly students and researchers in broad domain areas. The focus group discussions were conducted with two other groups (10 and 25 participants) of students of web and information technology courses. Results from these discussions were used to validate and refine the survey outputs. In both studies, 8 open questions were presented to the participants. The first half aimed at identifying participants' insights based on past experiences⁵. Questions referred to how events were discovered and shared, the decision making process, and about interesting relationships amongst events. The second part collected insights regarding existing web technologies in the context of the activities⁶. These questions explored the perceived benefits and drawbacks of using event directories and media, social networks, and a merger of these services.

2.2 Results and Discussion

Past experiences. According to participants, people's invitations and recommendations are the main means for finding out about events. Traditional media (e.g. posters, advertisements) also played a major role in identifying new events. Search engines were said to be commonly used, specifically when participants knew what to look for. On-line social networks were another commonly reported source, especially amongst students from the group discussions. Participants used event directories, mailing lists, newsletters and forums less often.

When deciding whether or not to attend to an event, participants assigned importance to constraints such as distance (location), availability (time) and budget (price). Social information about "who's joining" the event, and more specifically which friends will attend, also played an important role. The content of the event itself (e.g. type, performer, topic, target audience), subjective factors (e.g. fun, event atmosphere), and user opinions were also commonly mentioned.

Participants shared their experience by sharing stories and taking pictures and videos of attended events. Direct social interaction was the most common and preferred choice for sharing information. Media directories (e.g. Flickr, YouTube) and social networks (e.g. Facebook) were also commonly used.

Participants reported relying on previously attended events as a means to identify new interesting events (e.g. same venue, performer, category). Friends attending other events, target audience and users with similar interests, were other important identified relationships. Finally, future occurrences of repeated events was also seen as a strong relationship.

Existing technologies. Event directories were positively viewed for providing a single source for information overview, allowing opportunistic discovery of events by time, by category or by artist. Other positive features included social features,

⁵ The survey is available at <http://tinyurl.com/eventmedia-s1>

⁶ The survey is available at <http://tinyurl.com/eventmedia-s2>

notification of upcoming events, and links to related services (e.g. tickets). Low coverage of events was seen as a main drawback; others included information overload, unreliability and incompleteness (e.g. lack of location map and videos).

Participants recognized benefits from enriching events through media directories. According to participants, media facilitate reliving and sharing experiences. Additionally, media give a better impression of the environment/atmosphere and provide valuable information for decision support. As for enriching events with social networking information, participants reported that the main benefits are communication and information sharing between users. More importantly, it enhances the attendance information by allowing to identify social proximity (e.g. friends participating) and event popularity. Live event information updates (e.g. real-time tweets) was also seen as a positive feature. For both enhancements (associated media and social information) the drawbacks concerned information overload and privacy issues while sharing personal media.

3 Scenario-based user study

Based on the results of the exploratory user studies, we identified a set of potential use-cases. However, since participants relied on past experiences, it is debatable whether collected insights are representative of real user behaviors. In order to account for this and to allow a better understanding of behaviors in a well-defined scenario, a second study was performed in order to identify participants' strategies, information sources and behavior patterns while enacting four predefined scenarios:

1. Collect information about an event after receiving an invitation and decide whether or not to attend the event.
2. Use the information about an attended event, as well as other people's opinions (i.e. review, ratings) to identify similar events in the future.
3. Discover and decide about currently occurring events based on what other friends are currently doing (i.e. life streams).
4. Upload and share media for an attended event and explore other people's experiences through available media.

Two sessions were conducted where participants were requested to complete a set of scenarios, performing different tasks in order to achieve the presented goals. Fifteen participants (3 females) took part in the study at two research organizations. The mean age was 26 (range 23-49, SD=9.3). All participants used internet on a daily basis and were acquainted with related on-line services. On-line social network, media directory and event directory usage was measured on a 5pt Likert scale ranging from "never used" to "constantly used". Participants reported that they sometimes used social networks (M=3.3, SD=1.3) and media directories (M=3.1, SD=0.7), and never or rarely used event directories (M=1.3, SD=0.5). During each session of around 1.5 hours, participants role-played the four different scenarios while making use of internet access and a list of links to

well-known social networking services, media and event directories. Each participant was requested to self-report his or her experience. After each scenario was completed, participants shared their expectations, strategies, and outcomes of their actions. In addition, collaborative affinity diagramming was conducted by the groups to organize the outputs for each scenario. The affinity diagramming was done using post-it notes with the self-reported actions. During the exercise, participants collaboratively clustered the notes into different action sets on a flip-chart, thus making the strategy and behavior patterns explicit.

3.1 Observations

While completing each scenario, participants used on average 5 different information sources (most-used first): search engine (Google), venue/event website, media directories (e.g. YouTube, Flickr, Picassa), social networks (e.g. Facebook, Twitter), event directories (Last.fm, local city event directories).

Seeking information. The majority of the participants started the scenarios by searching for events using a search engine (e.g. Google). Participants used general terms or information provided by the scenario, usually combining title, venue, performer and other information, such as city or time, to constrain the search. Results directed users to event directories or to specific venues/events. Some participants were acquainted with specific events or venues and tried to reach the website directly. Some participants also searched in social networking services, limiting their strategies to finding events through friends, but had little or no information about the specific events presented in the scenarios. Other participants used event directories (e.g. Eventful, Upcoming, Zevents) but were mostly dissatisfied with results from the event directories due to low coverage for the specific scenarios. However, local directories seem to provide better results. In many cases, participants ended up at specific venue/event websites. Regardless of the sources, participants usually performed several subsequent searches, mainly on search engines to obtain further information (e.g. location on map, images, videos, user comments). While searching for related events, participants used event characteristics (e.g. type, genre, sub-genre, performers) as keywords. Alternatively, few participants relied on related event videos on YouTube or related event artists from Last.fm to identify related events.

Exploring Information. For participants who were redirected to the event website or were able to track the event on the venue website agenda, information about the event in terms of date, description, and performers was readily available as well as some images and videos providing a better illustration about the event. These seem to be the most complete information sources. Some used information from event directories which also provide factual data and few images from the event or the event performers. The directories provide also information about the event attendance as well as comments and reviews which was also appreciated by participants. Social networks such as Facebook were the favorite means for obtaining friends whereabouts, their event opinions and to see if they

were attending the investigated event. Another few participants said they would rely on instant messengers or emails in order to contact their friends directly. Social networking services (e.g. Twitter and Facebook) was also said to be the best source of live information about the events. In most cases, however, participants still searched for other related images and videos on a search engine or on media directories (e.g. YouTube) in order to better convey the experience. Participants said that images and specifically videos allowed a better understanding of the event's atmosphere and environment (e.g. party, disco, cozy).

Sharing Information. Most participants chose to share their event experiences through images on media directories such as Picasa and Flickr. The directories were specifically good for posting whole sets of pictures. These directories are also the main source to find more pictures about an event. Some participants also referred to YouTube to post videos about attended events. Many participants pointed out that they would use social networks such as Facebook to post images and videos on their profile or on the Facebook event page. Contrary to image directories, media is selected more carefully when posted on Facebook. The most interesting, representative or funny pictures and videos are posted. Friends can then comment on these media and share their experiences. Few participants preferred to post media directly on their personal web-sites or blogs. Some participants rely more on face-to-face sharing of media where they can point out specific pictures and discuss them. Few others share some pictures directly through emails and MMS messages.

3.2 Discussion

After completing all scenarios, clustered actions were used as creative input for semi-structured group discussions. During the discussion, participants addressed their expectations, strategies, main challenges and recommendations for exploring and sharing events. The results from these discussions are described below.

Information is spread and decentralized. When exploring events, participants reported that there were too many different information sources. They recognized that in order to fulfill the scenarios, there was a need to access several different on-line services. One participant reported *I don't like always having to go from one site to another to find out things about the event*. Therefore, participants agreed it was easier to use a single search engine that has broader coverage of different information sources. One participant reported *There is so much information that it's difficult to prevent the immediate reaction to go to Google*. However, if the participants knew where to find the information, they would go directly to that information source using bookmarks or known website addresses. Specific venue or event web sites were seen as the best source for information overviews. The sources often provided all necessary integrated information, including media. Social networks such as Facebook were also reported to integrate available information (e.g. photos, attendance list, discussions) to some degree, but not sufficiently. Participants suggested integrating social networks with other services. However, participants also agreed that an ideal solution should not just

be another information source. There should be some means of centralizing all available information. This can be summarized through one participant's comment: *It would be nicer to have a mash-up with the most important information from each website.*

Information seeking and decision making strategies. While searching for interesting events, there was no agreement on the search strategy. However, most participants reported that the most important information was location, time (date), type of event and popularity. A common strategy was to start constraining the factual properties of the event (e.g. type, location and date) in order to filter the available information. Participants showed an interest in specifying these constraints by defining ranges (e.g. max price). Searching by title (if known) and other information were amongst other options. Other few participants suggested alternative methods for identifying events, such as mobile location-based services to track nearby events. Participants also agreed that the strategies were dependent on type of event (e.g. concerts, parties, art exhibitions). Furthermore, social aspects such as people and friends who are attending could have priority over any other available information. Therefore, starting to search by friends was another potential starting point. One participant commented *If your close friends like it, it's more likely that you will enjoy it.*

Relationships and recommendations. Similarities among events that were said to be interesting include: location, date, event type and genre. However, participants agreed that the most valuable relationship was based on the common interests of people who attend the event. Recommendations were also seen as a potential feature, where users could receive interesting events based on popularity and ratings, people with similar interests/behaviors, friends' attendance, or on the user's past attendance, by keeping a user history.

Participants agreed that the presentation of information about events should be sorted (e.g. time, relevance, popularity) according to their needs (customizable). While displaying large numbers of related media, these could be clustered depending on the event type, media owner (e.g. friends) or visual similarity. Another suggested option was to show only the most popular media or filter the media that belong to known friends. Participants indicated that the most important information is factual information (e.g. what-where-when) and that any other relevant information should be one click away. Most participants argued they would like to have an instant overview of the event through associated media. Participants also agreed that while conducting the scenarios, associated media was the best way to easily illustrate the whole experience. One participant commented *You can have an idea about how the event looks like and what kinds of people go there. It's kind of like a preview.* Excerpts of songs, for a musical event (audio or video), would also be highly appreciated. Additional information such as ratings for events, weather conditions, distance from current location, travel information (public transportation) and accessibility are also said to be very useful. Being able to see user attendance was mentioned. Additionally, user profile pictures can convey a better idea about the "type of people" or target audience of the event.

3.3 Conclusions

We are aware of a number of limitations with this study. The events selected for the scenarios may not have corresponded with users' interests. We feel, however, that our results are sufficiently valid for guiding the development of the functionality of the application. Our results from the studies suggest the need for services that combine information from different event directories, social networks and media sharing platforms. Since information is spread and locked in different services, users express the need for a single resource to explore to experience events. Although this benefit is recognized for existing event directories, their lack of event coverage and information completeness affects the user experience.

Users also pointed out the benefits of merging different information sources. Social factors are a strong component when identifying, deciding and sharing experiences about events. Participants rely on other people not only to receive invitations and recommendations about events, but also to decide whether or not to attend. Social information obtained by attending an event (e.g. "who else is going?") or by sharing experiences (comments, reviews and ratings) provide valuable support for the decision making process. Shared interests amongst users is also the main identified means for obtaining event recommendations. "A picture is worth a thousand words" and associated media were also described as the perfect means for representing events. Images and video provide a powerful means for identifying several event characteristics, to convey the experience and to provide decision support. While participants request more information about the events, there is a common concern about information overload. This issue suggests that interfaces should avoid cluttered information and provide only timely and necessary information. Furthermore, there is a need to support different visualizations and improved browsing options that depend on user interests and constraints.

4 Data Scraping and Interlinking

In this section, we detail how the data from events and media directories is represented and interlinked. Overall, the dataset collected contains more than 30 million triples [5]. We provide a dump at <http://www.eurecom.fr/~troncy/ldtc2010/> and a SPARQL endpoint at <http://data.linkedevents.org/sparql>.

We have described and thoroughly compared numerous event ontologies in [4] which has lead to the development of the LODE ontology⁷, a minimal model that encapsulates the most useful properties for describing events. The goal of this ontology is to enable interoperable modeling of the "factual" aspects of events: *What* happened, *Where* did it happen, *When* did it happen, and *Who* was involved. 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 such as MO, CIDOC-CRM, DOLCE, SEM to name a few. The Ontology for Media Resource

⁷ <http://linkedevents.org/ontology/>

currently developed by W3C is a core vocabulary which covers basic metadata properties to describe media resources⁸. 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 Flickr photo and YouTube video descriptions.

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 categorized as a *Concert* has been given on the 24th of January 2010 at 20:00 PM in the *Henry Fonda Theater* featuring the *Radiohead* rock band. 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. Hence, we see that the video hosted on YouTube has for *ma:creator* the user *aghorrarag*.

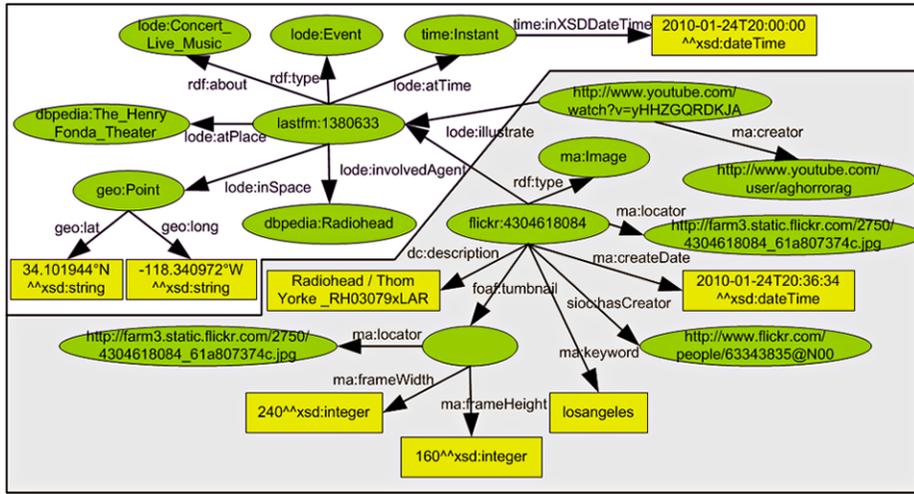


Fig. 1. The *Radiohead Haiti Relief Concert* described with LODE (*top*) and illustrated with media described by the Media Ontology (*bottom*)

We have populated these ontologies by scraping and semantifying data from event directories. We have been able to convert the description of more than 1.7 million photos which are indexed by over 108.000 events (table 1). We initially explored overlap in metadata between three websites (flickr, upcoming and last.fm) looking into explicit relationships between events and photos using machine tags. Furthermore, event descriptions from eventful were extracted by a 3 years sampling filtered by events tagged with the “music” keyword in order to maximize the likelihood of interlinking with the other event directories. Finally, we extract associated event videos from YouTube employing various querying strategies (e.g. event title, venue, performers) while using the date as a filter. It is important to note that we did not scrap all data that was available but

⁸ <http://www.w3.org/TR/mediaont-10/>

only the data that was mentioned as useful during our user studies. 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. More details about the dataset are available in [5]. 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 media collections.

	Event	Agent	Location	Media	User
Last.fm	57,258	50,150	16,471	1,425,318	18,542
Upcoming	13,114	0	7,330	347,959	4,518
Eventful	37,647	6,543	14,576	0	0
Total	108,019	56,693	38,377	1,773,277	23,060

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

5 End-User Application

Back-end Architecture The back-end of the system consists of a Sesame2⁹ SPARQL endpoint, a distributed query engine, a RESTful API, and a web server. All URIs minted in the dataset are dereferencable and are served as either static RDF files serialized in N3 or as JSON by the RESTful API. We implemented content negotiation in order to let clients decide about the desired representation. Clients requesting a JSON representation are redirected to the RESTful API, which is implemented using the Restlet library¹⁰ and runs within an OSGi runtime environment. Within this environment, a local RDF repository is used that is set up as a federator using a Distributed SPARQL implementation [3]. Besides serving JSON representations of resources available in the dataset, the RESTful API also provides convenience methods exposed as additional resources, which are not explicitly represented in the dataset. Examples for such functionalities include search over the dataset using different parameters such as keyword, time, location. We also allow dataset updates, for instance, by being able to specify attendance information or link additional media to existing events from the front-end.

Discovering Events Users wish to discover events either through invitations and recommendations, or by filtering available events according to their interests and constraints. Therefore, the interface allows constraining different event properties (e.g. time, place, category). Mechanisms for providing this desired support include restricting a time period through a timeline slider control input (Figure 2). Categories and location can be filtered using hierarchical faceted metadata [2], allowing users to browse through different dimensions of the collection. The hierarchical facets are presented according to the predefined event

⁹ <http://openrdf.org/>

¹⁰ <http://www.restlet.org/>

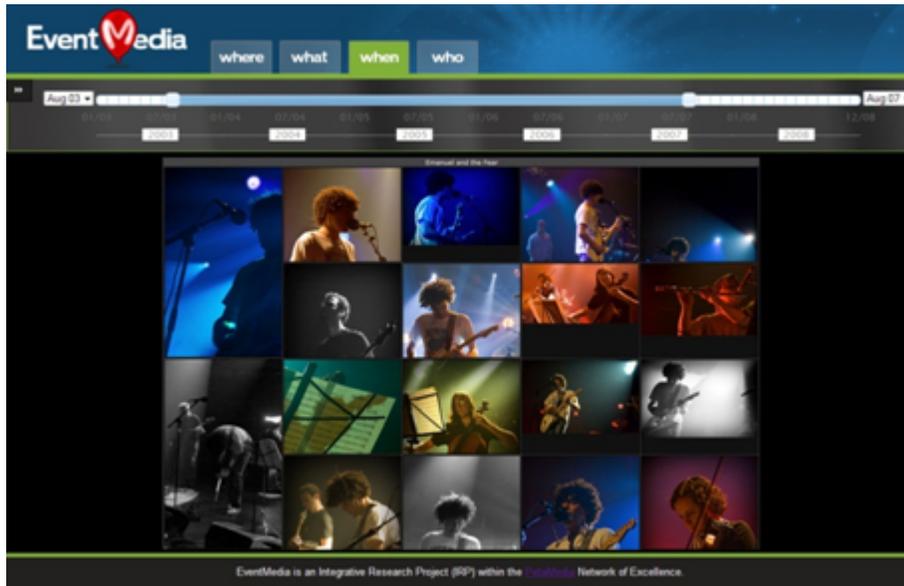


Fig. 2. Interface illustrating a set of media associated to an event for a period of time

categories in the SKOS taxonomy, and through an event's geo-location information. These properties allow the combination of different event types and locations while visually guiding the user through an interactive query refinement process. Faceted browsing also avoids empty results by restricting the available filtering options to display only non-empty results. Since users are likely to revisit information they have viewed in the past [2], we will also support simple history mechanisms, by saving a list of recently viewed events. To aid search, input boxes with dynamic term suggestions (auto-completion) is used to provide user feedback by suggesting a list of matching terms while typing.

Displaying Information After an event is selected, all associated information is displayed. Media are presented to convey the event experience, along with social information to provide better decision support. According to user interests, social proximity should be emphasized while displaying event attendance (e.g. friends attending). Other information that should be presented includes: performers, topic, genre, price. While scraping the data, some events such as popular music festivals were associated with more than 2,000 photos and videos. In order to deal with this large number, pagination is used while ordering media according to different contexts (e.g. by popularity, time, or social proximity). Alternatively, pictures are clustered according to context or visual similarity, and representative images are shown through Treemaps to present a varied sample of associated media. To allow searching for events in a variety of contexts, users also requested to be able to see events according to different views. We

explore these different views according to the basic event properties defined in the LODÉ ontology. **What** - Displays the most relevant associated media. A treemap is used to represent different event sets, where size can convey different properties (e.g. relevance, popularity)¹¹. **When** - Shows the chronological relations among events. A time axis is used to visualise the start-time and duration of events. **Where** - Indicates where events occur geographically to orient the user and convey distance. Maps are commonly used to visualize such information.

6 Conclusion and Future Work

We described an event-based approach for users to explore, annotate and share media. We conducted two user studies, where users were asked about real-world tasks they would like to carry out. We observe and identify participants' strategies, information sources and behavior patterns while enacting predefined scenarios. We used and consumed linked data technologies for integrating information contained in event and media directories. We described the scraping and interlinking process yielding a unique and dense dataset of more than 30 million triples. Finally, we present the architecture and the user interfaces of the application available at <http://eventmedia.cwi.nl/demo>.

7 Acknowledgments

The authors would like to thank Fabien Duchateau (CWI), Hyowon Lee (DCU), Bartosz Malocha (EURECOM) and Ansgar Scherp (WeST Institute) 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.

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*, pages 40–54, Athens, Greece, 2010.
2. M. Hearst. *Search User Interfaces*. Cambridge University Press, 2009.
3. S. Schenk, C. Saathoff, S. Staab, and A. Scherp. SemaPlorer - Interactive Semantic Exploration of Data and Media based on a Federated Cloud Infrastructure. *Journal of Web Semantics*, 7(4):298–304, 2009.
4. R. Shaw, R. Troncy, and L. Hardman. LODÉ: Linking Open Descriptions Of Events. In *4th Asian Semantic Web Conference (ASWC'09)*, pages 153–167, Shanghai, China, 2009.
5. R. Troncy, B. Malocha, and A. Fialho. Linking Events with Media. In *6th International Conference on Semantic Systems (I-SEMANTICS'10)*, Graz, Austria, 2010.

¹¹ See for example <http://www.jinni.com> OR <http://www.ted.com/talks>.