

Confomaton: A Conference Enhancer with Social Media from the Cloud

Houda Khrouf¹, Ghislain Ateazing¹, Thomas Steiner²,
Giuseppe Rizzo¹ and Raphaël Troncy¹

¹ EURECOM, Sophia Antipolis, France,
<firstName.lastName@eurecom.fr>

² Universitat Politècnica de Catalunya, Barcelona, Spain, <tsteiner@lsi.upc.edu>

Abstract. A scientific conference is a type of event for which the structured program is generally known in advance. The Semantic Web community has setup a so-called Semantic Web dog food server that exposes structured data about the detailed program of more and more conferences and their sub-events (e.g. sessions). Conferences are also events that trigger a tremendous activity on social media. Participants tweet or post longer status messages, engage in discussion with comments, share slides and other media captured during the conference. This information is spread over multiple platforms forcing the user to monitor many different channels at the same time to fully benefit of the event. In this paper, we present Confomaton, a semantic web application that aggregates and reconciles information such as tweets, slides, photos and videos shared on social media that could potentially be attached to a scientific conference.

Keywords: Data aggregation, Linked Data consumption, data reconciliation, data visualization

1 Introduction

Just like any other popular event, scientific conferences trigger an ever-growing amount of activities on social media. But in contrast to events like concerts or sport matches, a conference is highly structured, consisting generally of workshops and tutorials, parallel sessions composed of talks, keynotes, panels, posters and demos that all have planned schedules, topics, and allocated rooms. The Semantic Web community is used to model this structured data using RDF and to publish it following the Linked Data principles using a so-called Semantic Web dog food server³ [3]. The social media activities that are shared around the conference consist of slides, photos and videos posted by authors and participants but also status messages published on social networks such as Twitter, Google+ or Facebook. The problem is that these activities are unstructured data, spread over multiple platforms that are just weakly associated to a conference event

³ <http://data.semanticweb.org/>

as a whole as opposed to its fine grained sub-events. Overall, the physical participants or the ones who try to follow the event online are forced to monitor multiple channels to full benefit from a scientific conference.

Exploring this intrinsic connection between structured events and media shared on the web has been the focus of several studies [1,2,5]. They propose different techniques in the area of media classification, data interlinking and event detection, trying to leverage the wealth of user generated knowledge. However, most of these works have mainly targeted a specific social service such as Twitter or Flickr, without any guarantee that they can be valid for others services. We believe that exploiting the diversification of user generated content from different social services inside one application is a challenging task. In this work, we aim at creating a rich environment to enable users navigating events as well as their various representative media such as pictures, slides and tweets. A typical usage is to gather data about a scientific conference and investigate the added value of collecting scientific-related media. A non trivial task in such application is to connect structured data with extremely noisy content, especially in the case of a major conference. In this paper, we present Confomaton, a semantic web application that collects social media activities, reconciles the data and attempts to align it to the various sub-events that compose a conference. We will showcase Confomaton live during the ESWC 2012 conference.

2 Confomaton

The name *Confomaton* is a word play on the French term *Photomaton* (English photo booth) and *conference*. Just like a Photomaton illustrates the scene inside of the booth, the Confomaton illustrates an event such as a conference enriched with social media. Confomaton is a semantic web application that produces and consumes linked data and is composed of four main components: (i) an Event Collector which extracts events descriptions such as the ones available in the Semantic Web Dog Food corpus; (ii) a Media Collector which collects social media content and represents it in RDF using various vocabularies; (iii) a Reconciliation Module playing the role of associating social media with sub-events and external knowledge; (iv) a User Interface powered by an instance of the Linked Data API as a logical layer connecting all the data in the triple store with the front-end visualizations.

Event Collector: it takes as input the Dog Food corpus described using the SWC ontology and converts all events into the LODE ontology⁴, a minimal model that encapsulates the most useful properties for describing events. We use the Room ontology⁵ for describing the rooms contained in a conference center. An explicit relationship between an event and its representative media (photo, slide, tweet, etc.) is realized through the `lode:illustrate` property. For describing

⁴ <http://linkedevents.org/ontology/>

⁵ <http://vocab.deri.ie/rooms>

those media, we re-use two popular vocabularies: the W3C Ontology for Media Resources⁶ for photos and videos, and SIOC⁷ for tweets, status, posts and slides.

Media Collector: it has the purpose to search from various social networks and media platforms for event-related media items such as photos, videos, and slides. We currently support 4 social networks (Google+, MySpace, Facebook, and Twitter) and 7 media platforms (Instagram, YouTube, Flickr, MobyPicture, img.ly, yfrog and TwitPic). Our approach being agnostic of media providers, we offer a common alignment schema for all of them containing information such as the deep link of the media, the media type, the story URL, the story content, the author profile URL, the timestamp, etc. In order to retrieve data from media providers, we use the particular media provider’s search Application Programming Interfaces (API) where they are available, and fall back to screen scraping the media provider’s website if not.

Reconciliation Module: it aims to align the incoming stream of social media with their appropriate events and to interlink some descriptions with general knowledge available in the LOD cloud (e.g. people and institutions descriptions). Attaching social media to fine-grained event is a challenging problem. We tackle it by pre-processing the data with two successive filters in order to reduce the noise: one of them relies on keyword search applied to some fields such as title and tag, while the other one filters data based on temporal clues. The reconciliation is then ensured through a pre-configured mapping between a set of keywords and hashtags and their associated events. Furthermore, we extract named entities from the microposts using the NERD framework [4] and we develop a specific heuristic for aligning tweets with sub-events.

User Interface: it is built around four perspectives (tabs in the UI) characterizing an event: (i) “*Where does the event take place?*”, (ii) “*What is the event about?*”, (iii) “*When does the event take place?*”, and finally (iv) “*Who are the participants of the event?*”. In addition, the UI offers full text search for these four dimensions. On the left side of the main view, the user can select the main conference event or one of the sub-events as provided by the Dog Food metadata corpus. On the center, the default view is a map centered on where the event took place and the user is also encouraged to explore potential other type of events (concerts, exhibitions, sports, etc.) happening nearby, this data being provided by EventMedia [5]. The *What* tab is media-centered and allows to quickly see what illustrates a selected event (tweets, photos, slides). Zooming in an event triggers a popup window that contains the title and timetable of the event, the precise room location and a slideshow gallery of all the medias collected for this event. For the *When* tab, a timeline is provided in order to filter events according to a day time period. Finally, the *Who* tab aims at showing all the participants of the conference. This is intrinsically bound to a social component, aiming not only to present relevant information about a participant (his affiliation, homepage, or role at the conference) but also the relationships between the participants between themselves and with the events.

⁶ <http://www.w3.org/TR/mediaont-10/>

⁷ <http://rdfs.org/sioc/spec/>

The UI is powered by the Linked Data API⁸ which provides a configurable way to access RDF data using simple RESTful URIs that are translated into queries to a SPARQL endpoint. More precisely, we use the Elda⁹ implementation developed by Epimorphics. Elda comes with some pre-built samples and documentation which allow to build specification to leverage the connection between the back-end (data in the triple store) and the front-end (visualizations for the user). The API layer helps to associate URIs with processing logic that extract data from the SPARQL endpoint using one or more SPARQL queries and then serialize the results using the format requested by the client. A URI is used to identify a single resource whose properties are to be retrieved or to identify a set of resources, either through structure of the URI or through query parameters.

We have deployed the application using the data describing the ISWC 2011 conference (Figure 1). The application is available at <http://eventmedia.eurecom.fr/iswc2011/>.

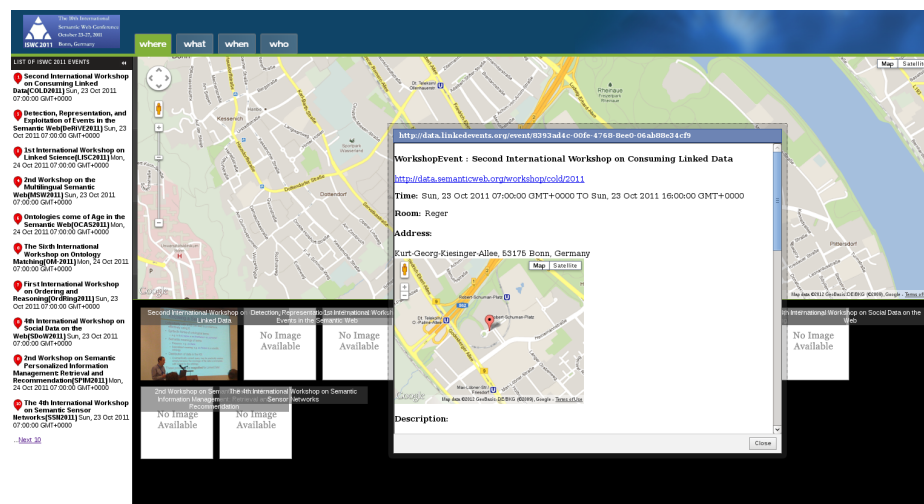


Fig. 1. A showcase of Confomaton with the ISWC 2011 data

3 Conclusion

In this paper, we have presented Confomaton, a semantic web application using the diversity of media resources generated by users, that can potentially be linked with more structure metadata such as a detailed program of a scientific conference as exposed by the Dog Food corpus. We show that collecting

⁸ <http://code.google.com/p/linked-data-api/wiki/Specification>

⁹ <http://code.google.com/p/elda>

and reconciling media items from many services (Twitter, SlideShare, Flickr, Google+, etc.) enables to provide a better conference experience including visual conference summarization or explorative search during and after an event.

Confomaton is an ambitious project that shows well the difficulty to use Linked Data technologies in a real setting. The solution we propose makes use of many services starting from scraping, aggregating data and their reconciliation. It unifies them and exposes the aggregated information as linked data. However, the more services are handled and the more issues one has to deal with, generally with the API provided by those services (e.g. the number of requests of some APIs). Concerning the Linked Data API, at the time we developed the paper, it was not possible to handle queries using selectors with *DISTINCT* and *GROUP BY* queries, although there are means to go through this limitation using *UNION*. We also face the problem of the objective criteria to select a particular vocabulary for modeling the data. Finally, Confomaton is a flexible solution to encompass the data fragmentation due to the proliferation of services used by the conference guests. Confomaton will be deployed with the data describing the ESWC 2012 conference and we will invite all physical and remote participants to provide suggestions for a better user experience.

Acknowledgments

This work is partially supported by the project AAL-2009-2-049 “Adaptable Ambient Living Assistant” (ALIAS) and by the European Commission under Grant No. 248296 FP7 I-SEARCH project.

References

1. H. Becker, M. Naaman, and L. Gravano. Learning similarity metrics for event identification in social media. In *3rd ACM International Conference on Web Search and Data Mining (WSDM'10)*, pages 291–300, New York, USA, 2010.
2. X. Liu, R. Troncy, and B. Huet. Using Social Media to Identify Events. In *3rd Workshop on Social Media (WSM'11)*, Scottsdale, Arizona, USA, 2011.
3. K. Möller, T. Heath, S. Handschuh, and J. Domingue. Recipes for Semantic Web dog food - The ESWC and ISWC metadata projects. In *6th International Semantic Web Conference (ISWC'07)*, pages 802–815, Busan, Korea, 2007.
4. G. Rizzo and R. Troncy. NERD: A Framework for Unifying Named Entity Recognition and Disambiguation Extraction Tools. In *Proceedings of the Demonstrations at the 13th Conference of the European Chapter of the Association for Computational Linguistics (EACL'12)*, pages 73–76, Avignon, France, 2012.
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.