

A Services Framework for Using Social Web as Social Sensor Web

D. Kotzinos, S. Kleisarxaki, L. Theodoridou, V. Christophides, N. Petalidis, G. Kariotis

Abstract— Online social networking sites carry a lot of information organized around the people who actually submit it. They rely a lot on the interconnections between the actors of the network and they relate the generated information in terms of connections or ties among those actors. On the other hand, social networks carry a lot of information in real time or almost real time since people “report” information as they see it unfolding before their eyes. We are targeting this information and we try to understand when this information can be related to refer to a single event, which has both a spatial and a temporal dimension. These social sensors can alert us on what is happening in the society when it happens and by aggregating the different reports from people who are on the scene. In order to achieve that we suggest and describe a set of services that can be used for collection of the information, identifying the discussion topics and provide at the end an alert if the discussion qualifies for and contains enough information to define an event. The goal is for the system to work with no prior knowledge of the events we seek.

Keywords— events, real time, social sensor web, social networks, twitter

I. Introduction and Main Idea

Social networking sites are becoming increasingly popular and this is measured in terms of both number of users and number of posts, which means that more people are now involved but these people are also more active [19]; these form what we call the social web. But the vast increase in users' number has brought to the social networking sites a lot of attention along with new requirements. Users demand new applications and services based on the social networking platforms and the social networks respond to that by providing a set of open (or not so open) APIs that allow programmers to access the data posted in the sites and build new applications or services on top. But these APIs only provide access to the available information (and sometimes not to the whole of it); they do not provide any intelligent processing of the posts besides some statistical information, e.g. the most popular or highly used terms. Additionally the provided information remains inherently human centered, i.e. the interest is on who to follow / connect with and thus users are expected to know beforehand what is needed and as a consequence they cannot respond to the real time changing world. In this work we propose a new event based approach to handle the information being made available in the social networks, i.e. we want to focus on what is of interest to the users without expecting the users to know before where this information will come from. In that sense we claim that it is increasingly important to be able to detect events as they happen and relate the available information to these events. We feel that we can combine information available in social networks with semantic event based models like CIDOC CRM [18] and thus describe the relationships among the event, the

users and the posts in a meaningful way, having the ability to reason on these events and connect them to other pieces of information also described in a semantically rich way. Thus events can be documented and made directly usable by others.

A useful analogy is the one with the (semantic) sensor web, another effort that has received a lot of attention lately [1, 2]. We want to use posts, comments, photos, documents, etc. as a network of social web sensors, which can be used to “sense” or detect events as they occur based on a more complex and intelligent process than the one in the actual sensor web, where most of the events of interest can be predefined – especially in controlled environments. Additionally we plan to exploit the efforts that aim to bring closer the social and the semantic web [3] in order to be able to use metadata either provided by the social networking site or extracted dynamically from the posts themselves [8, 9, 10].

We describe the development of a framework that uses and supports the extensive network of social web sensors, which will allow for the automatic detection of events as they occur and are “reported” by the users.

Dimitris Kotzinos
Department of Computer Science
UCP, Pontoise, France

Sofia Kleisarxaki
University of Crete
Greece

Lila Theodoridou
TEI of Central Macedonia
Greece

Vassilis Christophides
University of Crete
Greece

Nicolaos Petalidis
TEI of Central Macedonia
Greece

Giorgos Kariotis
TEI of Central Macedonia
Greece

II. Research Methodology and System Architecture

The overall objective of the research work is to devise, implement and test the necessary algorithms in order to provide online real-time event detection from social network streams and to offer these algorithms for access by multiple concurrent users and applications in the form of a framework of bundled web services.

The framework is able to combine information (posts) from one or more social networking sites and to process them in real-time as they become available. The processing is based on identifying the spatial and temporal metadata information that is associated with each post and based on that to group the posts respectively. The closer in space and time the posts are, the more likely to refer to the same event. We use spatial ontologies [11, 12, 13] to correctly identify space and spatial operators (e.g. near, in, close) to better use the spatial dimension of the information. The same is served by the use of a temporal ontology [11, 13] and the corresponding operators that allow us to classify the events in short, medium, large, etc. Furthermore we at the same time analyze the text from the post and the corresponding textual metadata (e.g. title) of the post (text analysis) and thus decide whether two or more posts define a new event (e.g. when the same but new subject is referenced) or add information to an existing event (e.g. when the same but existing subject is referenced). Then, as more posts become available, we are able to classify them under either an existing event or create new events as previously (taking the form of clusters of posts, which will be formed and changed dynamically). The definition of the number of posts required to detect an event, the accuracy with which it is detected and the confidence interval for each detection are three of the targets of this research. This is bundled in a robust engine that will be responsible for the metadata processing (both spatial and temporal) and the text mining parts.

The framework provides a services API (either REST [15] or SOA [16, 17]) that can be used by programmers in order to build new applications on top. The API allows the user to define the social networks that is interested in and provide the necessary credentials and then silently observes the posts as they arrive until an event is detected; then the user will be notified of this event and if interested can follow up, otherwise will wait until the next event. The user does not need to register the events of interest, since in our ever-changing world it is rather impossible to know them beforehand. The framework provides the results in an way that can be easily combine with other information (e.g. in mashups) and in one or more of the established semantic web languages like XML, RDF, GML [4, 5, 6, 7]. A schematic representation of the proposed technical architecture is shown in Figure 1.

Finally we are also building a proof-of-concept application that uses the services engine and API to provide a useful example to the end user; we provide a GUI-based application that will demonstrate the capabilities of the framework. Moreover we make every effort that the framework will be extensible enough in order to allow for future extensions like the ability to semantically relate two events and to decide whether an event was the cause for another event that followed.

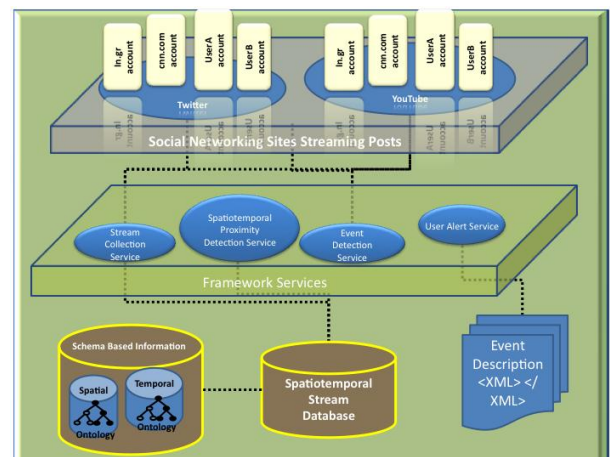


Fig. 1. System Architecture

III. Usage Example

Let's assume that a user is interested in finding events that are reported in posts in Twitter social networking site. Additionally let's assume that he has access to a set of posts originating from accounts including but not limited to cnn.com, in.gr, UserA and UserB, which for simplicity we accept that are the same in both sites (please note that not everyone has access to all the information published to a social network; thus one has access to events that are originated to posts coming either from users that have made their posts public or from users that have authorized our access to their posts). After the user/application declares that to the framework then we wait until there is an adequate number of posts that relate to a specific event e.g. an earthquake or a fire, which is identified dynamically. If the created cluster becomes robust enough then we can declare that an event has occurred providing the representative of the cluster as a description and the saved posts (not all posts can be saved) as additional information. This event might have initially been posted by users and then be picked up by the media, like cnn.com. Users might also be able to upload pictures or videos through their mobile phones and these can be added to the available reports by the media. The flow of information in this case is depicted in Figure 2.

In summary, we provide a comprehensive framework (and an extensible programmable component) that would allow event detection in social networks going beyond the traditional text based approach by exploiting the social, temporal, spatial and of course textual dimensions of the social networks.

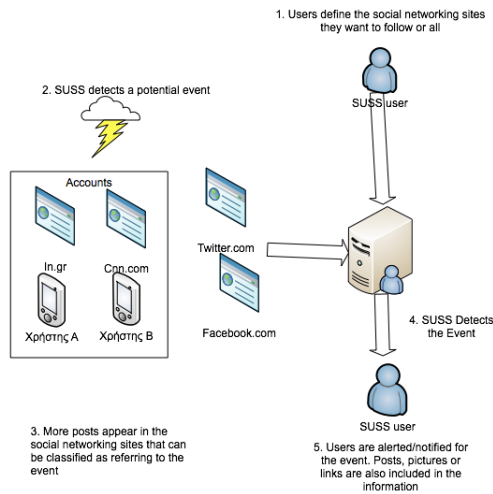


Figure 2. Usage example and information flow

IV. Conclusions

We propose a service-based system (thus both reusable and extensible) that allows for real-time detection of events happening in the world at large and reported in online social networks. We use both time and space to precisely identify an event and we are gathering the complete set of tweets that correspond to this event, so as to report them with the event. In this paper we describe the architecture of the system and the main concept of this work. We plan in the near future to present the results of the framework and the applications built on top of it.

Acknowledgment

This research is implemented through the Operational Program "Education and Lifelong Learning" and is co-financed by the European Union (European Social Fund) and Greek national funds.



References

- [1] M. Balazinska, A. Deshpande, M.J. Franklin, P.B. Gibbons, J. Gray, S. Nath, M. Hansen, M. Liebhold, A. Szalay, V. Tao, (2007) Data Management in the Worldwide Sensor Web, Pervasive Computing, IEEE, April-June 2007, Volume: 6, Issue: 2, pp. 30-40, ISSN: 1536-1268
- [2] Mike Botts, George Percivall, Carl Reed and John Davidson, (2008) OGC® Sensor Web Enablement: Overview and High Level Architecture, Book Series: Lecture Notes in Computer Science, Book: GeoSensor Networks, Volume 4540/2008, Publisher: Springer Berlin / Heidelberg, pp. 175-190, ISBN 978-3-540-79995-5
- [3] Tom Gruber, (2008) Collective knowledge systems: Where the Social Web meets the Semantic Web, Web Semantics: Science, Services and

Agents on the World Wide Web, Volume 6, Issue 1, Semantic Web and Web 2.0, February 2008, Pages 4-13, ISSN 1570-8268

- [4] D. Brickley, R.V. Guha and B. McBride, (2004) RDF Vocabulary Description Language 1.0: RDF Schema, W3C Recommendation 10 February 2004.
- [5] G. Klyne, J.J. Carroll and B. McBride, (2004) Resource Description Framework (RDF): Concepts and Abstract Syntax. W3C Recommendation. 10 February 2004.
- [6] P. V. Biron, and A. Malhotra, (2001) XML SchemaPart 2: Datatypes. W3C Recommendation 02 May 2001.
- [7] H.S.Thompson, D.Beech, M. Maloney, and N. Mendelsohn, (2001) XML Schema Part 1: Structures.W3C Recommendation 2 May 2001.
- [8] E. Prud'hommeaux and A. Seaborne, (2007) "SPARQL Query Language for RDF," W3C working draft, 26 Mar. 2007; www.w3.org/TR/rdf-sparql-query/.
- [9] J. Hendler, (2007) "The Dark Side of the Semantic Web," IEEE Intelligent Systems, vol. 22, no. 1, 2007, pp. 2-4.
- [10] T. Berners-Lee, J. Hendler, and O. Lassila, (2001) "The Semantic Web," Scientific American, May 2001.
- [11] Pierre Grenon and Barry Smith, (2004) SNAP and SPAN: Towards Dynamic Spatial Ontology, Spatial Cognition & Computation, 4:1, 69-104
- [12] S.Spaccapietra, N.Cullot, C. Parent and C. Vangenot, (2004). On spatial ontologies. In Proceedings of the VI Brazilian Symposium on Geoinformatica (GEOINFO 2004), Campos do Jordao, Brazil.
- [13] J. R. Hobbs, and F. Pan, (2004). An ontology of time for the semantic web. ACM Transactions on Asian Language Information Processing (TALIP) 3, 1 (Mar. 2004), 66-85.
- [14] W. Kuhn, (2002). Modeling the semantics of geographic categories through conceptual integration. In Egenhofer, M. J. and Mark, D. M., editors, GIScience, volume 2478 of Lecture Notes in Computer Science, pages 108-118. Springer.
- [15] S. Tyagi, (2006) RESTful Web Services, available at <http://java.sun.com/developer/technicalArticles/WebServices/restful/>
- [16] E. Ort, (2005) Service-Oriented Architecture and Web Services: Concepts, Technologies, and Tools, available at <http://java.sun.com/developer/technicalArticles/WebServices/soa2/>
- [17] M. N. Huhns, M. P. Singh, "Service-Oriented Computing: Key Concepts and Principles," IEEE Internet Computing, vol. 9, no. 1, pp. 75-81, Jan./Feb. 2005
- [18] M. Doerr, "The CIDOC CRM - an Ontological Approach to Semantic Interoperability of Metadata", AI Magazine, Special Issue on Ontologies, Nov. 2002
- [19] Nielsen (2010), Social Networks/Blogs Now Account for One in Every Four and a Half Minutes Online, http://blog.nielsen.com/nielsenwire/online_mobile/social-media-accounts-for-22-percent-of-time-online/, retrieved 2010-10-29