

Social Data and Multimedia Analytics for News and Events Applications

Ioannis Kompatsiaris
Information Technologies Institute
6th Km Charilaou-Thermi road
GR57001, Thessaloniki, Greece
+30 2311257774
ikom@iti.gr

Sotiris Diplaris
Information Technologies Institute
6th Km Charilaou-Thermi road
GR57001, Thessaloniki, Greece
+30 2311257778
diplaris@iti.gr

Symeon Papadopoulos
Information Technologies Institute
6th Km Charilaou-Thermi road
GR57001 Thessaloniki, Greece
+30 2311257772
papadop@iti.gr

ABSTRACT

This paper discusses a framework enabling real-time multimedia indexing and search across multiple social media sources. It places particular emphasis on the real-time, social and contextual nature of content and information consumption in order to integrate topic and event detection, mining, search and retrieval, based on aggregation and indexing of shared user-generated multimedia content. User-friendly applications for the News and Events domains have been developed based on these approaches, incorporating novel user-centric media visualisation and browsing methods. The research and development is part of the FP7 EU project SocialSensor.

Categories and Subject Descriptors

I.5.4 [Pattern Recognition]: Applications; H.3.3 [Information Search and Retrieval]: Retrieval models

General Terms

Algorithms, Theory, Experimentation

Keywords

social media, multimedia analysis, multimedia search, analytics

1. INTRODUCTION

Social networks have become an integral part of modern life driving more and faster communication than ever before. Media content is created and published online at unprecedented rates by both regular users and professional organisations. At the same time the wide availability of smartphones has enabled the creation and instant sharing of media content at the time and place of an event, creating an “online reflection” of what happens in the real world. However, the fast pace, the huge volume and the unpredictable nature of user-contributed content make it extremely challenging to obtain informative views of evolving news stories and events in real time and to quickly surface relevant media content.

Especially with regard to the news industry the challenge is to use and embrace new content authoring and provision methods, and the channels offered by social media and mobile technologies, to

their fullest advantage, for both information gathering and information distribution. A key challenge in this respect is to develop appropriate tools for quickly surfacing trends, sentiments and discussions in social media, in relevant and useful ways.

Another domain where social media analytics are gaining great importance is the organization of large events, such as festivals and expos. As attendants of such events need to organise their visits, new methods providing context-aware information are becoming necessary in order to enhance user experience. Event organisers can also benefit from social media sensing applications that can help them capture the pulse of their events and gain valuable insights into their impact on visitors. To this end, they need tools that help them make sense of the large amounts of online messages and shared content.

This short paper presents a framework for social media analytics that have been developed in the context of SocialSensor, a 3-year FP7 European Integrated Project¹ aiming to tackle some of the challenges outlined above and to offer solutions as well as improvements in the industry domains of professional news and event organisation. In the first domain, we present a system for crawling, indexing and browsing social media content with the goal of facilitating the discovery of newsworthy and interesting social multimedia. In the second, we present the EventSense application that helps event organisers extract insights from large events by mining large amounts of online messages shared through OSNs. Offering real-time social indexing capabilities for both of these use cases is expected to have a transformational impact on both sectors.

The subsequent chapters outline some of the research applications developed by SocialSensor, for sensor mining and social search, as well as results from their evaluation.

2. SOCIAL MULTIMEDIA SEARCH

The first application facilitates the targeted collection, indexing and browsing of shared media content through a hybrid crawling strategy, comprising both a stream-based and a query-driven approach. In addition, it integrates very efficient and scalable image indexing and clustering implementations.

The crawler is responsible for the collection of data and content from online sources in the form of Items (posts made in a social platform, e.g. tweets), WebPages (URLs embedded in collected Items) and MediaItems (images/videos embedded in Items or WebPages), given a set of crawling specifications (arguments

(c) 2014, Copyright is with the authors. Published in the Workshop Proceedings of the EDBT/ICDT 2014 Joint Conference (March 28, 2014, Athens, Greece) on CEUR-WS.org (ISSN 1613-0073). Distribution of this paper is permitted under the terms of the Creative Commons license CC-by-nc-nd 4.0.

¹ <http://www.socialsensor.eu>

specifying what to crawl, e.g. a hashtag on Twitter) as input. The proposed crawling and indexing framework, depicted in Figure 1, comprises the following components: (a) Item collection, comprising stream-based (stream-manager) and query-driven (search-manager) components, (b) structured data repositories based on mongoDB and Solr, (c) fetching and indexing components, including WebPage fetching, article extraction, MediaItem detection and extraction, feature extraction and visual indexing [1], (d) aggregation components, including geo, visual clustering and analytics. Several of the system components are available as open-source projects on GitHub², intended for use in the professional journalism domain [2].

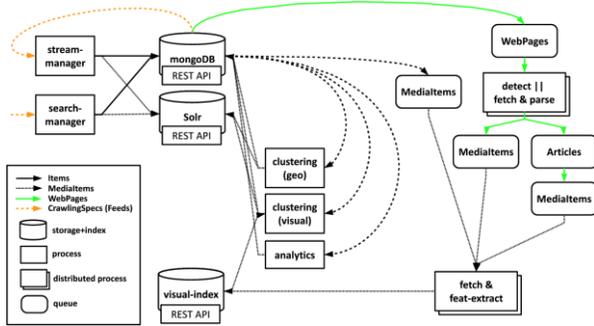


Figure 1. Social multimedia crawling and search.

3. SOCIAL MEDIA SENSING

The EventSense application enables a series of mining operations on social media content around large events, including the automatic association of online messages to entities of interest (e.g. films in the case of a film festival), the discovery of trending topics, and the detection of sentiment (positive, negative, neutral) both at an entity level (e.g. per film) and on aggregate. In addition, the application produces an informative social media summary of the event of interest by automatically selecting and putting together its highlights, e.g. the most discussed entities and topics, the most influential users, the evolution of the discussions' sentiment, and the most shared media and news content.

More specifically, online messages about the event are organized around entities of interest (e.g. films) and topics, and sentiment scores are extracted for each of those, by aggregating the sentiment expressed by individual messages. An overview of the EventSense application is illustrated in Figure 2. For more details please see [3].

4. RESULTS

With respect to the social multimedia crawling and search framework we evaluated its media retrieval capabilities in the context of the #OccupyGezi events. We selected images related to the event from four different sources and used them as visual queries to the NN search component. To assess the search performance, we computed a *coverage* score by dividing the number of queries with at least one correct result with the total number of queries. We note that in most cases coverage ranges from 0.5 to 0.95. Regarding the quality of ranking for similar images, in most cases where at least one correct result was found to be the same or very similar to the query image, the correct

results were ranked above the incorrect ones. It is also noteworthy that in all cases, the response times were sub-second, typically in the range of 100-200msec. More details are available in [2].

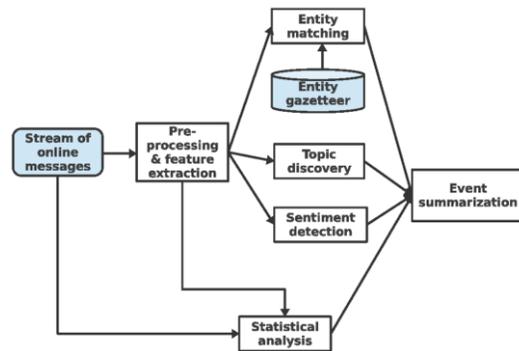


Figure 2. EventSense system.

We conducted an evaluation of EventSense on a Twitter dataset around the 53rd International Film Festival of Thessaloniki taking place between November 2nd and 11th, 2012. The evaluation results provided evidence that real-world event variables, such as film ratings, are correlated with aggregate statistics mined from the stream of online messages. For further details, please see [3].

5. CONCLUSION

We presented two applications for real-time social media content indexing, search and retrieval tailored around the needs of news professionals, as well as organisers and attendees of large events. To this end, innovative analysis techniques of social data and content, assisted by effective indexing of real-time social media streams were developed. Evaluation studies demonstrate the effectiveness of such techniques in collecting diverse content from social networks, in browsing and searching it in multiple ways, and in perceiving the pulse of large-scale events.

6. ACKNOWLEDGMENTS

This work is supported by the SocialSensor FP7 project, partially funded by the EC under contract number 287975.

7. REFERENCES

- [1] Spyromitros-Xioufis, E., Papadopoulos, S., Kompatsiaris, I., Tsoumakas, G., & Vlahavas, I. (2012, May). An empirical study on the combination of surf features with VLAD vectors for image search. In *Image Analysis for Multimedia Interactive Services (WIAMIS), 2012 13th International Workshop on* (pp. 1-4). IEEE.
- [2] Papadopoulos, S., Schinas, E., Mironidis, T., Iliakopoulou, K., Spyromitros-Xioufis, E., Tsampoulatis, I., and Kompatsiaris, I., 2013. *Social Multimedia Crawling and Search*. In *IEEE Computer Society Special Technical Community on Social Networking E-Letter*, vol. 1, no. 3, October 2013.
- [3] Schinas, E., Papadopoulos, S., Diplaris, S., Kompatsiaris, I., Mass, Y., Herzig, J., and Boudakidis, L. 2013. *EventSense: capturing the pulse of large-scale events by mining social media streams*. In *Proc. 17th Panhellenic Conference on Informatics (PCI '13)*. ACM, New York, NY, USA, 17-24.

² <https://github.com/socialsensor>