

# "Making Sense of it All": An Attempt to Aid Journalists in Analysing and Filtering User Generated Content

Sotiris Diplaris, Symeon Papadopoulos,  
Ioannis Kompatsiaris  
Informatics and Telematics Institute  
6<sup>th</sup> Km Charilaou-Thermi road, 57001, Thessaloniki, Greece  
+30 2311 257 778  
{diplaris, papadop, ikom}@iti.gr

Nicolaus Heise, Jochen Spangenberg  
Deutsche Welle  
Strategy, Marketing & Distribution / New Media  
Voltastr. 6, 13355 Berlin, Germany  
+49 30 4646 5604  
{nicolaus.heise,  
jochen.spangenberg}@dw.de

Nic Newman  
School of Arts - Journalism,  
City University London  
Northampton Square,  
London EC1V 0HB  
nic.newman@gmail.com

Hakim Hacid  
Bell Labs France, Alcatel-Lucent  
Route de Villejust, 91620 Nozay,  
+33 (0) 1 3077 1808  
hakim.hacid@alcatel-lucent.com

## ABSTRACT

This position paper explores how journalists can embrace new ways of content provision and authoring, by aggregating and analyzing content gathered from Social Media. Current challenges in the news media industry are reviewed and a new system for capturing emerging knowledge from Social Media is described. Novel features that assist professional journalists in processing sheer amounts of Social Media information are presented with a reference to the technical requirements of the system. First implementation steps are also discussed, particularly focusing in event detection and user influence identification.

## Categories and Subject Descriptors

D.2.1 [Software Engineering]: Requirements/Specifications.

## General Terms

Algorithms, Design, Reliability, Verification.

## Keywords

social indexing, social media, sensor mining, news, journalism

## 1. INTRODUCTION

The emergence of the web 2.0 and, subsequently, Social Networks, has fundamentally changed the way in which information is gathered and provided. It also had vast impacts on the relationships of traditional information providers with their audiences. From a previously mainly unidirectional relationship, there has been a shift to a culture of exchange and sharing. A selected few, e.g. established outlets such as news agencies or media organisations, are no longer the only ones who decide what is reported when and how. Their monopoly on the means of information production and distribution has fallen, or is in the process of being undermined (the extent to which this happens

being dependent on specific market conditions and respective regulations).

This changed relationship, as well as a new way of reporting about events as they occur, could be witnessed very strikingly during the Tsunami that affected large parts of South-East Asia in late 2004, and the London bombings of July 2005 [1]. Within hours of the events occurring, those at the scene were producing photos, videos and texts. This was quickly passed on and shared, and subsequently distributed to large audiences via traditional media channels and organisations.<sup>1</sup> Other events were to follow, e.g. the Mumbai attacks in November 2008<sup>2</sup> and the emergency landing of a passenger plane on New York's Hudson river in January 2009.<sup>3</sup> These are just a few selected occasions in which content posted on Social Networks set the news agenda, at least in the initial moments of events, and proved to be significant components in the reporting process that followed.

<sup>1</sup> The BBC's Head of News in 2005, Richard Sambrook, gives the following account of events: "Within six hours we received more than 1,000 photographs, 20 pieces of amateur video, 4,000 text messages, and 20,000 e-mails. People were participating in our coverage in a way we had never seen before. By the next day, our main evening TV newscast began with a package edited entirely from video sent in by viewers." Richard Sambrook "Citizen Journalism and the BBC". The Nieman Foundation for Journalism at Harvard. See <http://www.nieman.harvard.edu/reportsitem.aspx?id=100542>.

<sup>2</sup> Information about the attack as it unfolded was quickly shared via Twitter. Photographer Vinukumar Ranganathan uploaded more than 100 photos onto Flickr (see <http://www.flickr.com/photos/soumik/3062552427/in/photostream/>) which were widely shared thereafter.

<sup>3</sup> One of the first mentions of the crash was a Tweet by Janis Krums who was aboard a passenger ferry used to rescue stranded passengers from the floating plane. Krums uploaded a photo he took of the floating plane to TwitPic from his iPhone. See <https://twitter.com/#!/jkrums/status/1121915133>. That photo subsequently made it to a vast number of news sites.

Nowadays, it is no longer unusual that breaking news events are first reported on Social Networks. Technology and available infrastructure have made this possible (connected people who are "always on", multimedia capable devices such as smartphones, high transmission speeds "on the go", improved usability etc). As a result, more and more content is provided, posted or shared on Social Networks such as Twitter, YouTube, Facebook, MySpace or Flickr, to name but a few. These "pipes" are filled by individuals with particular agendas or "ordinary people". The latter are often referred to as "citizen journalists" (although this phrase is somewhat misleading as the mere fact of making content available to a larger number of recipients does not necessarily make the producer a journalist). Nevertheless, content published on Social Networks has increasingly made it into the channels and services of and become a part of the reporting process for traditional information providers such as news organisations. New and affordable publishing and distribution tools (e.g. Social Networks such as the aforementioned, blogs, or services such as Storify) have made this possible. In reverse, all this has had and is still having significant impacts on the way information is produced and distributed nowadays. It is no longer in the hands of a "selected few" to decide what is being published when and how, and how information is passed on and shared. Instead, Social Networks have become an integral part of both the information gathering process as well as the communication mix for all kinds of aims and ambitions, for example (political) campaigning, awareness-raising, marketing, and - and this is the focus of this paper - news and information gathering and subsequent provision in the journalistic domain.

The challenge for traditional information providers is to use these new content authoring and provision methods and channels offered by Social Media to their fullest advantage: to embrace it instead of feel threatened by it. And use it in both the process of information gathering as well as the distribution of information. This paper deals primarily with the use of content from Social Networks for the information gathering process and subsequent reporting by professional journalists. It does not focus on the use of Social Networks for the spreading and distribution of information (i.e. a news provider using Social Networks to disseminate its information).

## 1.1 Challenging issues in journalism when relying on social networks

Although social networks can be seen as an opportunity for journalists thanks to, e.g. the huge amount of generated content and opinionated content, they also represent big challenges. In fact, to get the most out of content residing in Social Networks, a number of challenges still exist or are as yet unsolved. There are still large areas that require solutions or improvements. From an information provider's perspective, this includes - but is not limited to - the following aspects (focusing on information gathering):

- (a) *Verification*: ensure that the content posted in Social Networks is accurate / true;
- (b) *Filtering*: according to particular needs / interests;
- (c) *Sensing*: discover trending topics and what is "up and coming" in order to guide further investigation;
- (d) *Analysis*: analyze particular trends and tendencies according to specific questions;

- (e) *Visualisation*: present search results in an attractive, easy to understand way;
- (f) *Cross-platform issues*: enable searches across different Social Media platforms;
- (g) *Speed*: time is money, all processes need to happen quickly and efficiently, without being at the expense of accuracy;
- (h) *Legal*: copyright/ownership rules need to be adhered to and solved in a timely and user-friendly way;
- (i) *Attribution*: content needs to be attributed to sources, without compromising contributors' privacy, and guaranteeing their protection;
- (j) *Business*: transactions (e.g. of posted content) must be ensured in a safe and fair manner that is legally binding;
- (k) *Linguistics*: searches should work across different languages;
- (l) *Usability*: tools and interfaces should be intuitive and easy to use.

## 1.2 Our solution

SocialSensor<sup>4</sup> is a 3-year FP7 European Integrated Project aiming to tackle some of the challenges outlined above and offer solutions as well as improvements. In the project framework, new techniques for analysis, aggregation and real-time search of user-generated content will be developed in order to extract useful information and make it available for use in different applications. Innovative solutions from the fields of information extraction and retrieval, social network analysis, user modeling, semantic web services, and media adaptation, delivery and presentation, will compose a software platform that crawls and analyses multimedia UGC (User Generated Content) from the social web, combines it with professional content, and makes it searchable for professional users, but also recommends, delivers and presents it to media consumers depending on their context and their personal profile. To achieve this, crucial issues have to be tackled, such as the sheer data volume, its heterogeneity and low quality, to name but a few of the research challenges.

The resulting multimedia search system will be showcased and evaluated in the news domain, among others. The news use case targets two end user groups: (a) news professionals who are interested in leveraging UGC in their work, (b) casual online and mobile news readers. With respect to professional usage, different scenarios will be supported, such as the discovery of emerging trends and topics, aggregation of UGC with professional content, analysis of massive amounts of social data for new insights and profiling of news portal users, and recommendation of relevant content. Casual news readers will benefit from innovative features, such as real-time discovery of news items, proactive delivery (push) of relevant content to users based on their context, and socialization of users with other news reader through ad hoc social networking.

In this position paper, we present a perspective from the news professionals' side on the user requirements of the envisioned social media analysis system, by also describing tools and methodologies under development that will assist in realizing this vision.

---

<sup>4</sup> <http://www.socialsensor.eu>

## 2. SOCIAL MEDIA MADE FOR JOURNALISM

Social Media like Twitter, Facebook, YouTube, LinkedIn and others have become a hugely important source of information for journalists and media organizations [2-3]. Nevertheless, journalists have to deal with considerable difficulties and challenges when using Social Media. They need to monitor a variety of different platforms with different interfaces, log-in mechanisms and visualisation methods at the same time. They are overwhelmed by a large amount of tweets, postings, images, videos and other content (often in different languages) that are practically impossible to process in real time. They need to take into account that despite a growing number of subscribers of social networks, opinion gathered from Social Media is far from being representative. And with each piece of important information they have identified they are challenged with the difficult task of verifying this information as quickly as possible from other sources. Without having any means to assess the likelihood that a piece of information or a content item is true or that a Social Media user is trustworthy, the process of verification - one of the basic principles of professional journalism - takes a long time and foils the advantages of information gathering from Social Media.

Considering the current situation, we have identified the following core journalistic needs when dealing with Social Media (Table 1).

**Table 1. Key journalistic needs.**

Requirement	Description
Trend and Sentiment detection	Journalists need ways of tracking trends and sentiment in a specific moment and over time.
Real-time alerts	Journalists need to be alerted in real time about breaking news and other new developments on issues they are working on.
Trustworthiness	Journalists need to have access to eyewitnesses or other trustworthy informants on breaking news.
Responsiveness	Journalists need to quickly find answers to specific questions they have about a story they are working on.
Access to contributors	Journalists need to have access to individuals and specific groups (e.g. key influencers in Social Media).
Verification	Journalists need tools that support them in the verification process.

Given that the working environment for journalists typically provides good connectivity, these core journalistic needs should be addressed by a specialised system for news professionals with advanced analysis capabilities.

What SocialSensor aspires to achieve in terms of its News use case is a **single tool that quickly surfaces trusted material from Social Media - and does so with context**. The following passage elaborates further on the motto from above:

- "a single tool": one platform, one interface
- "quickly": fast, in real time
- "surfaces": automatic discovery and clustering of information

- "trusted": automatically verified or automatic support in the verification process
- "material": any material (text, image, audio, video = multimedia)
- "Social Media": across all relevant Social Media platforms (or a combination of platforms relevant to journalists)
- "with context": personal background, sentiment, location of users, comparison with mainstream media sources

Therefore, such a tool should be able to:

- identify and visualise events and trends across Social Media sources in real time,
- identify key influencers and opinion formers around any event, and
- support journalists in verifying user generated content (text, images, video and audio) from Social Media sources

There are many tools available today that allow the social web to be filtered and navigated. The envisioned news tool will be different primarily because it is set up to support serious journalism. The sources, algorithms and interfaces are designed for serious news. The trends and content are surfaced with the intelligence and knowledge that helps influence the news agenda. An ideal tool should enable journalists to see public opinion ‘in the raw’ as it develops around subjects, people and events. This means lists - but crucially will also display sentiment around breaking and running stories that is statistically valid and immediately useable.

Beyond that, such a tool will help journalists surface the best user generated content coming from many sources. Unlike other tools, it will operate cross-network, cross-media and cross-language.

## 3. JOURNALISTIC REQUIREMENTS

Attempting to elaborate and specialize on the above principles for creating a journalistic tool that encapsulates semantics and knowledge from social media, we have concluded with a set of desired features for the SocialSensor journalistic system.

Journalists need to be provided with richer and more comprehensive information on chosen topics on a more representative basis than current tools are able to provide. As journalists aim for a single view of all social media activity around specific stories and events, an innovative journalistic tool needs to be able to crawl content from ideally all major social networks (at least Twitter, Facebook, Google+, Youtube and Flickr). But to gather information from social networks only is not sufficient. All newsworthy content should also be put in context with traditional mainstream media like television and newspapers. Not only in order to get a consolidated view of emerging events but also to see how the sources interrelate, to better understand the language used in social media and to harmonize it with language used in traditional mainstream media.

On the basis of this crawl, a journalistic tool would ideally automatically create dynamic news topic lists that are updated in real time. This would allow journalists to see developing trends as they emerge, to be alerted to big breaking news stories or simply to find out how strongly a story is playing with ordinary people. From a technical point of view, this would require a system with

real time indexing capabilities, along with news items clustering schemes around automatically defined clusters (topics not manually pre-defined as news categories). The latter strategy also involves the proper naming of the emerging clusters by injecting words and context into the news list algorithm so as to help determine what clusters mean.

Given the large amount of crawled information and data, the next step is to filter and display this information in a user-friendly manner using a variety of intelligent techniques. Journalists should be enabled to refine crawled information according to their specific needs (e.g. by persons, countries, organizations). This includes a system of adaptable clustering methods based on words and context. Visualization techniques offered by the tool should support journalists in filtering and analyzing search results by location, time, type, source, contributor etc. which would also allow them to present information and data to their audience, readers and customers in a more attractive and understandable fashion.

A journalist is not only interested in pure facts but also in the sentiment developing around an event. Sentiment analysis is therefore another important feature for a journalistic tool. The system should not only analyze the popularity of media items, URL's and user generated content but also gauge public opinion and trends in general. Equally important is the identification of key influencers. They are a reliable indicator for evolving trends and it is often crucial to follow these key influencers or trendsetters in order to be informed about new developments in real time. Key influencers and opinion formers should be automatically identified and listed around any news topics. To this end, new algorithms on influence identification around topics have to be developed. They need to be able to provide lists of influencers per topic and allow filtering by country, time etc.

But the system should not only identify these key influencers but also provide the journalist where possible with their contact details like websites, email-accounts or even telephone numbers. The automatic discovery of this information requires to identify sources of information linking Social Media to publicly available emails, phone numbers, webpages etc. An intuitive interface containing a list of "most useful people", and search capabilities is then needed to easily link this information together in a usable way. This would enable journalists to directly interact with these key influencers, to interview them, to commission content for mainstream media outlet and to investigate whether certain information is accurate.

The latter refers to one of the main principles of professional journalism: verification. Each information researched by a journalist needs to be confirmed by at least two sources before being published. But with a lot of - especially user generated -content gathered from social media accuracy is a big issue and the verification process is difficult and often not satisfactory. A journalistic tool which supports journalists in this verification process by checking the authenticity of media, checking location and time of media, checking similarities and differences of media items and assessing the reliability of contributors by calculating trust scores would be a huge benefit for journalists. This includes the automatic tampering of photos and the constant search for metadata regarding specific media items. Also, similarity to other content items (i.e. visual similarity with other images and videos, or textual similarity with other comments) has to be automatically cross-checked, building on what exists already (e.g. Tineye - see [www.tineye.com](http://www.tineye.com)).

A contributor's reliability can be assessed based on hidden context and personal details. By this, journalists can have a first indication of whether a contributor is worth pursuing (or not). In deriving user reliability information, trust and reputation come into play. A trust score has to be calculated for all users based on their reputation (and by also assessing their network of followers) and the history of their contributions.

The picture would be complete if the system was able to crawl in multiple languages, offered an immediate translation tool and introduced elements of a digital rights management enabling journalists to efficiently acquire the rights to user generated content. A challenge is thus to integrate existing translation technologies into the system, taking into account journalistic workflows and respective procedures.

Finally, the discovered feedback from Social Media sources needs to be aggregated to mainstream news. Journalists will be enabled to see social feedback to a specific mainstream web story to help determine their next steps and angles. This procedure would need a track back mechanism against any URL already indexed to the index of Social Media sources.

#### 4. TOWARDS LEVERAGING SOCIAL SENSORS

First steps towards realizing the envisioned system are already taking place and a main goal to this direction is to automatically understand information streams coming from Social Media.

To technically address the heterogeneity of different content sources, a novel paradigm for media content organization is needed, in order to bring together diverse web resources that refer to the same entity of interest. SocialSensor attempts to bring new mining techniques for intelligently merging the content coming from different sources in one fundamental object, labelled "DySCO", which is fed through an analysis of the large, heterogeneous, and continuously evolving data.

Figure 1 depicts the typical lifecycle of a DySCO. There are two main stages involved:

- (a) creation and maintenance, and
- (b) search, delivery and presentation.

Our vision is that the social dimension, if correctly used, can reinforce the mining process through:

- (a) the availability of more information in the network and
- (b) the fact that social interactions represent mainly human interactions providing an implicit understanding of users, which is a key lacking dimension in most existing mining strategies.

Currently, online content is indexed and searched at an atomic level, i.e. each content item is processed and indexed independently of the rest of the collection. SocialSensor will attempt to extend this paradigm by performing indexing and search over composite objects relating to a common topic of interest. Such composite objects are called DySCOs. The benefit of using DySCOs over single items is that it will be possible to extract aggregate knowledge/inferences by analysing them as a collection. In addition, performing the indexing at a collection-level could enable richer representation of contextual information with respect to content, i.e. the indexing mechanism will be able to access contextual information about content items.

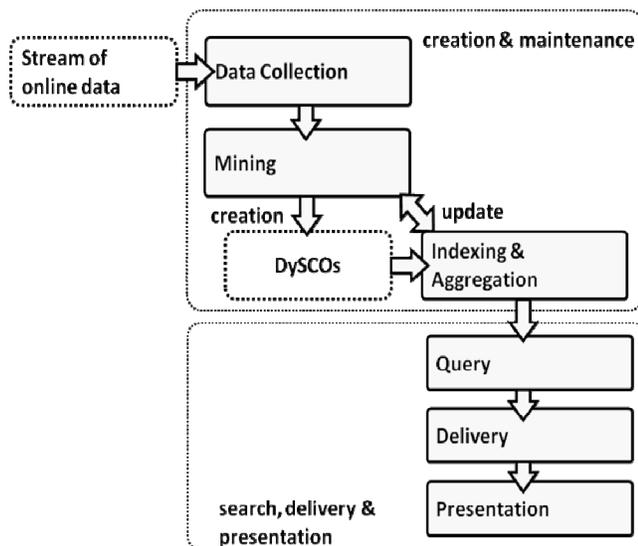


Figure 1. DySCO lifecycle.

In this sense DySCOs can be defined as composite objects centered around a particular topic of interest (news topic) that encode contextual and inferred information about collections of content (news) items that are detected to be related to the given topic of interest.

To realize this, SocialSensor will improve statistical models to make them support the new constraints imposed by the social Web ecosystem. These models and techniques include (extensions of) topical model [4], frequent item sets mining [5], clustering [6], dynamics of social networks and community detection [7], natural language processing techniques [8], and relevance and ranking modeling [9].

In the following sections we present two research directions, encapsulating the above concepts, which try to provide some of the features discussed in Section 3, i.e. event detection from image collections, and influence detection based on social interaction analysis.

#### 4.1 Event Detection

The current paradigms of browsing through collections of user contributed photos are ineffective for the exploration of very large photo collections due to the fact that they provide neither a high-level overview of the collection nor a structured means for exploring it.

Different techniques for automatically detecting events in social multimedia collections have been recently developed, exploiting geotagging information [10], or applying filtering, grouping and expansion rules on the collection [11]. Other methods try to detect events in text streams from the social media appeared recently. Some promising work can be found in [12], [13] and [14].

A new approach towards the automatic discovery of events and within image collections can assist journalists in discovering emerging clusters of trending events or topics in standard photo collections from Social Media.

In [6], a Sensor Mining framework exploiting the content dimension of the Social Web was presented for automatically extracting events in large photo collections by means of clustering photos into groups related to a single topic/object of interest, and then classifying the extracted clusters into landmarks and events. The proposed framework consists of the following four analysis steps:

- (a) creation of a graph encoding the similarities between pairs of images,
- (b) clustering of the images by means of community detection on the graph,
- (c) classification of the resulting clusters into landmarks and events,
- (d) post-processing of the clusters.

In the first step, similarities between photos are computed by combining (visual) content- and text-based similarity, by extracting SIFT features and producing co-occurrence similarity graphs respectively. Once the similarity graph is derived, an efficient community detection approach is applied to extract more densely connected sets of nodes. A resulting set of photo clusters is available for further analysis, while a significant number of photos are discarded (i.e. not assigned to any cluster). For each photo cluster, a set of features is extracted that are subsequently provided as input to an SVM or kNN classifier for classifying the cluster as depicting a landmark or an event. The resulting classification accuracy was found to be superior to previous methods [6]. As a last step, event clusters are post-processed in order to derive human-friendly textual descriptions by aggregating the textual metadata of the photos contained in them. Each cluster is assigned a title by computing the most frequent term sequence across the titles of its photos. For each cluster, a convex hull and a time interval are computed based on the geotagging and timestamp information of the individual photos. The aggregate information extracted is helpful in the context of presenting the clusters to journalists interested in exploring the content collection.

#### 4.2 Social Interaction Analysis for Identifying Influencers

Human social interactions result into construction of social interaction networks. Information contained in these social interaction networks is used to enrich DySCO-centered streams with derived social information, such as influence. The influence in a social network is defined as the ability of persons to make their social relatives act similar to them when deciding on something [15]. The work done in this area aims at approximating influence, or a part of it, thanks to features and measures or metrics on a social network.

Different methods have been proposed in the literature [16][17][7] and the influence is used in different contexts such as the prediction of the information diffusion in social networks. For example, [16] proposes to perform a ranking of influential nodes from information diffusion samples without network structure. A clustering is performed on the diffusion sequences in order to form two groups, low and upper group, based on the cardinality of the sequences and Jaccard coefficients comparisons. Nodes included in the upper group are then ranked according to the F-measure that is defined as the weighted harmonic mean of precision and recall. The highest ranked nodes are called "super-mediators" which appear frequently in long diffusion sequences with many active nodes and less frequently in short sequences. Our method separates the computational resources calculation into smaller steps so that a coarse measurement can be quickly provided based on rapidly obtainable information, such as number of friends/followers. Other factors include the creative activity of a content contributor, the number and type of reactions to their interactions, the frequency and reliability of their interactions and their friend's reactions. Third parties can also weigh the results according to their own assigned user priorities. These metrics are refined gradually as additional

information from the social graph is obtained and processed. This resolves the primary challenge to provide useful social metrics on-demand in a timely fashion over a massive amount of dynamic social data. Generic user's influence is advantageously complemented by influence per topic or a set of keywords, derived from content and context of interactions. As the topics extracted from highly dynamic streams might rapidly evolve, we may observe drastic changes in users' influence scores.

A first demo is already available, providing live influence calculations for social applications on request, through a REST API. In order to illustrate the calculation, an influence dashboard is developed which provides specific characteristics of social network users: their general influence, their specific influence per topic, a live effect of their friends and interactions on the influence score and their position within their communities.

## 5. CONCLUSION

In this paper we have outlined and highlighted some of the challenges the (news) media industry and journalists are confronted with today. We have furthermore presented our vision for a system that aims to incorporate emerging knowledge from Social Media. This is to be done with a view to aid professional journalists in sifting through, analysing and filtering the huge volumes of information residing in social networks in order to obtain information of relevance for their journalistic jobs. In order to achieve this, proper context-based analysis techniques have to be developed and applied on content residing in social networks. In order to reveal useful insights and provide useful information for the professional news sector in a timely and reliable fashion.

SocialSensor aspires to provide a tool for professional journalism investing in innovative analysis techniques of social sensors (such as event and influence detection), assisted by effective indexing of real-time social media streams.

At the heart of the envisioned system is an alethiometer – a way of "measuring truth". This does not verify content in an absolute sense, but it provides the tools to make it quick and simple for journalist to do so – all in a single interface. It can assess the immediate value of some newsworthy content or flag up the potential unreliability of an individual contributor. It does all this from an interface that allows personal views and filters of this content to ensure that it fits and integrates into the workflows of individual journalists or media organisations. This is to aid them in what they are there for: to provide speedy, accurate and reliable information to audiences about topics as they occur, or determined based on editorial decisions.

## 6. ACKNOWLEDGMENTS

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

## 7. REFERENCES

- [1] A. Stuart. "Citizen Journalism and the Rise of "Mass Self-Communication": Reporting the London Bombings". In *Global Media Journal, Australian Edition*, 1(1), 2007.
- [2] G. Brandstetter, and P. Hörschinger. "Journalism and Social Media". ikp PR and Lobbying GmbH, Vienna, 2010.
- [3] N. Newman. "The rise of social media and its impact on mainstream journalism: A study of how newspapers and broadcasters in the UK and US are responding to a wave of participatory social media, and a historic shift in control towards individual consumers", Reuters Institute for the Study of Journalism, University of Oxford, 2009.
- [4] S. Sizov. "GeoFolk: Latent Spatial Semantics in Web 2.0 Social Media". 3<sup>rd</sup> ACM Int'l Conf Web Search and Data Mining (WSDM), New York, USA, 2010.
- [5] A. Siebes, J. Vreeken, M. van Leeuwen. "Compression based frequent items set mining. Item sets that compress". In *SDM 2006*, pp. 393-404, 2006.
- [6] S. Papadopoulos, C. Zigkolis, Y. Kompatsiaris, A. Vakali. "Cluster-based Landmark and Event Detection on Tagged Photo Collections". *IEEE Multimedia Magazine* 18(1), pp. 52-63, 2011.
- [7] J. Yang, J. Leskovec. "Modeling Information Diffusion in Implicit Networks". *IEEE Int'l Conf. On Data Mining*, 2010.
- [8] D. Nadeau, S. Satoshi. "A survey of named entity recognition and classification". *Linguisticae Investigationes* 30(1), pp. 3-26, 2007.
- [9] V. Carchiolo, A. Longheu, M. Malgeri. "Reliable peers and useful resources: Searching for the best personalised learning path in a trust- and recommendation-aware environment". *Inf. Sci.* 180(10), pp. 1893-1907, 2010.
- [10] M. Brenner and E. Izquierdo. "Mediaeval benchmark: Social event detection in collaborative photo collections". In Larson et al. *Working Notes Proceedings of the MediaEval 2011 Workshop*, Pisa, Italy, Sep 1-2, 2011, volume 807 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2011.
- [11] S. Papadopoulos, C. Zigkolis, Y. Kompatsiaris, and A. Vakali. "Certh @ mediaeval 2011 social event detection task". In Larson et al. *Working Notes Proceedings of the MediaEval 2011 Workshop*, Pisa, Italy, Sep 1-2, 2011, volume 807 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2011.
- [12] T. Sakaki, M. Okazaki, and Y. Matsuo. "Earthquake shakes twitter users: real-time event detection by social sensors". In *Proceedings of the 19th international conference on World wide web*, pages 851–860. ACM, 2010.
- [13] J. Weng and B.S. Lee. *Event detection in twitter*. 2011.
- [14] H. Sayyadi, M. Hurst, and A. Maykov. "Event detection and tracking in social streams". In E. Adar, M. Hurst, T. Finin, N. S. Glance, N. Nicolov, and B.L. Tseng, editors, *ICWSM*. The AAAI Press, 2009.
- [15] S. Wasserman, K. Faust. *Social Network Analysis: Methods and Applications*. Cambridge University Press. 1994.
- [16] K. Saito, M. Kimura, K. Ohara, and H. Motoda. *Discovery of super-mediators of information diffusion in social networks*. In *Proceedings of the 13th international conference on Discovery science, DS'10*, pages 144–158, Berlin, Heidelberg, 2010. Springer-Verlag
- [17] J. Tang, J. Sun, C. Wang, and Z. Yang. *Social influence analysis in large-scale networks*. In *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining, KDD '09*, pages 807–816, New York, NY, USA, 2009. ACM.