Reports on the Workshops Held at the Sixth International AAAI Conference on Weblogs and Social Media

Daniel Archambault, Ruben Bouwmeester, Cosmin Cabulea, Elizabeth M. Daly, Giusy Di Lorenzo, Maarten de Rijke, Martin Harrigan, Eser Kandogan, Michael Muller, Mor Naaman, Daniele Quercia, Damiano Spina, Markus Strohmaier, Arkaitz Zubiaga

■ *The workshops day at the Sixth International Conference on Weblogs and Social Media included three full-day workshops (Real-Time Analysis and Mining of Social Streams, Social Media Visualization, and When the City Meets the Citizen) and one half-day workshop (The Potential of Social Media Tools and Data for Journalists in the News Media Industry). Workshops were held on June 4, 2012, the day prior to the first day of the technical conference, in Dublin, Ireland.*

The Potential of Social Media Tools and Data for Journalists in the News Media Industry

In 2012 the International Conference on Weblogs and Social Media (ICWSM 2012), held in Dublin, Ireland, focused primarily on cutting-edge research in social media. Most notably, the increasing influence of user-generated content in the newsroom was discussed.

As social media are becoming more and more relevant as a source of information, traditional media organizations are faced with new challenges. Aanyone equipped with a smartphone can now capture and publish events as they unfold. In many cases established news agencies are not the first point of call for information anymore. In order to stay competitive media organizations are increasingly depending on content from social networks to cover and present all perspectives of an event. However, they face one crucial question when it comes to using content from these networks: "Is the source reliable?" At this workshop, we discussed existing approaches and ways in which some of the prevailing challenges are encountered when developing new methodologies.

The keynote speaker, Katrin Weller, questioned whether Twitter is actually a social network. No, she states. Status updates on Twitter are covering all sort of topics — from popular culture to neuroscience, from intimate, personal details to major press releases of world-leading companies. In their totality, they can be considered as a giant but completely unstructured and unorganized knowledge base of what is going on. On the one hand this enables browsing and discovering new interesting pieces of information (serendipity effects). On the other hand it may pose enormous challenges to people looking for particular information. How can we access this rich resource of social content, which

includes facts as well as jokes, observations, personal reports, emotions, and opinions of more than 140 million users? According to studies, Twitter is actually used for information dissemination and consumption instead of interaction. Research also shows that 47 percent of journalists are already using Twitter for sourcing new story angles. However, the trustworthiness of information remains questionable.

Two EC cofunded projects, Arcomem and Socios, were also presented at the workshop. One of the use cases in Socios is a web-based application for journalists that aims to make journalistic research in social media fast and easy. The application allows easier access to content across social media platforms (Twitter, Facebook, Flickr, YouTube, DailyMotion, and MySpace) by aggregating content related to a topical search query. It also provides journalists with a range of added value functionalities (for example, social filtering, media item ranking, transaction negotiation, or event detection).
Arcomem aims to harvest, refine, analyze, contextualize, and preserve data from social networks on specific events for future use.

The primary lesson of the workshop came from Jochen Spangenberg (Deutsche Welle), host of the workshop: "Be the first, but first get it right." Because of social media we will, in time, have more citizen journalists. We will have more sources for news. But we should realize that the basic rules of journalism, such as confirming and fact checking, will always stay the same.

The chair of this workshop was Jochen Spangenberg. The papers of the workshop were published as AAAI Technical Report WS-12-01. This report was authored by Ruben Bouwmeester and Cosmin Cabulea.

Real-Time Analysis and Mining of Social Streams

The recent increase of real-time data provided by users on social networking services has leveraged an important gain in the real-time processing of social streams. Processing the streams in real time can help enhance search engines, news media, and many other systems by feeding them with fresh knowledge about current affairs. Performing such analysis in real time is of the utmost importance for early reporting of breaking news, events, trends, and any other knowledge related to current affairs. However, analyzing social streams in real time makes the task more challenging as it requires making decisions without a clue of what will be next in the stream.

The first International Workshop on Real-Time Analysis and Mining of Social Streams brought together experts in the real-time analysis and mining of social streams and enabled the further development of discussion and exchange of knowledge around these tasks. Given the novelty of the research field, the workshop also aimed to encourage attendees to build a discussion forum to share on the current state of the research field, as well as to propose solutions for the shortcomings.

The workshop started with a brilliant keynote speech by Jimmy Lin (professor at the University of Maryland and engineer at Twitter) who has been developing real-time tools for the social media company. Lin presented the main challenges arising from the real-time processing of large-scale data and described how they are dealing with it for improving discovery and search of salient information from Twitter's torrential streams. His descriptions of the architecture used by Twitter to process effectively a huge volume of 320 million tweets every day also helped participants of the workshop understand the difficulties of such a task.

Participants of the Real-Time Analysis and Mining of Social Streams workshop presented research that can be classified into three main topics. First, one of the main issues that need to be considered when processing social media data in real-time is scalability. The stream of information needs to be processed fast as more and more information is coming continuously. As Philipp Cimiano (University of Bielefeld) showed, simple filtering strategies can easily and accurately scale up the classification of social media documents into events. Second, information processing is another important aspect, where making sense of an overwhelming din of data is paramount. Agustín D. Delgado Muñoz (UNED) presented an accurate way of efficiently disambiguating social media mentions of company names. And third, the semantic representation of unstructured data aroused the interest of workshop attendees. The demonstration Confomaton presented by Houda Khrouf (EURECOM) is based on semantic web technologies to aggregate real-time information from different sources to enhance conference experiences.

Participants of the workshop actively contributed to a fruitful discussion, commenting on issues such as availability of resources for real-time research, as well as the next steps of the workshop, including ideas to create a community of researchers interested in the field. Before the workshop finished, workshop attendees also had the opportunity to show the applications they are developing for real-time processing of social media. There, Amit Sheth demonstrated the real-time tool that was developed at Wright State University by exploring tweets sent during the Occupy movements in the United States.

Contributions presented at the workshop paved the way for researchers interested in real-time processing of social streams. As this is a topic of interest to many researchers, multiple ideas were proposed as to future work, considering a wide variety of disciplines including information retrieval, natural language processing, data mining, and the semantic web, among others.

Arkaitz Zubiaga (City University of New York), Damiano Spina (UNED, Spain), Maarten de Rijke (University of Amsterdam), Markus Strohmaier (Graz University of Technology), and Mor Naaman (Rutgers University) served as cochairs of this workshop. The papers of the workshop were published as AAAI Technical Report WS-12-02.

Social Media Visualization

Social media study and analysis bring researchers from many fields into a single setting. Although the tasks of these researchers are varied, data visualization and analytics play an important role. For industry and academics alike, visualization of social media data helps with hypothesis formation and supports the is workshop provided a venue to discuss how we can apply data visualization and analytic techniques to social media data.

The workshop brought together researchers not only in computer science, but members of industry and researchers in the social sciences and humanities. The major themes of the papers presented at the workshop were visualization- and visual-analytics-centered approaches for microblogging, blog, and image data sets. About half of the accepted papers focused on microblogging data while the remainder focused on image and blog data sets. Ben Shneiderman, a world leader in human-computer interaction (HCI) and information visualization, presented an excellent keynote address on social media network analysis, citing some of the work in the field and its future challenges. We also organized an applications panel whereby researchers in the humanities and industry had an opportunity to present interesting problems in their disciplines. Karen Wade (Humanities Institute Ireland), Paul M. Watson (Storyful), and Jimmy Lin (Twitter) provided short presentations and discussion on how their fields can benefit from social media visualization methods. In addition to the paper presentations, we had three poster presentations during the coffee breaks.

One of the challenges identified in this area is building visualization and analytics systems that are able to handle very large data sets and still be interactive. A number of our application panel members cited this as being an area of interest and many of our workshop papers presented interesting approaches to this problem. Future collaboration with researchers in the area of data mining and machine learning could prove fruitful in achieving these scalability goals and provide a tighter integration of computational and visual analytics approaches. Additionally, closer interaction with the user community that requires social media visualization on a daily basis would be beneficial in identifying requirements for products and future directions for research. The workshop provided a venue allowing for an initial point of contact for members of our multidisciplinary field.

This workshop was the first of its kind and provided excellent initial collaboration between researchers and industry. We hope to continue exploring better integration of visualization practices in order to support the tasks of users in the humanities, social sciences, and industry. Participants found that the venue was very useful and would like to see the workshop continue to run at the AAAI ICWSM conference.

Daniel Archambault, Eser Kandogan, and Martin Harrigan served as cochairs of the workshop. The papers of the workshop were published as AAAI Technical Report WS-12-03.

When the City Meets the Citizen

The When the City Meets the Citizen AAAI workshop brought together researchers and practitioners to discuss and explore research challenges and opportunities in applying the pervasive social computing paradigm to understand cities and, crucially, engage their citizens. Social media have played a key role in providing insights into people's activities, opinions, and day-to-day lives. These detailed user-generated information streams offer a unique opportunity for cities to understand and engage their citizens.

The workshop brought together researchers from both industry and academe and included diverse backgrounds of computer science, digital media, geographical information systems, and anthropology. Interaction was a key theme of the workshop where attendees were encouraged to participate through a combination of speed dating, group discussions, and mechanisms to identify overlapping interests among them. Eight papers were presented covering topics that included how social media can be used to understand citizen coordination following an emergency, capture cultural differences in food tastes, and act as a sensor to the city. Another theme included pervasive digital displays and how they can encourage visitors to engage with the city space.

During group discussions a clear challenge was identified on how to balance privacy concerns and still engage users in geolocation-based activities. A combination of users' limited understanding of the implications of sharing data and the limited ability to specify how the information may be used and leveraged were identified as key issues that remain to be addressed.

George MacKerron from UCL's Centre for Advanced Spatial Analysis gave an excellent keynote speech. He presented Mappiness, a mobile phone app that gathers data on how our surroundings and activities affect our happiness. The original goal was to explore how green spaces in urban environments affect city dwellers; however, the possible insights that can be highlighted to policy makers go far beyond the original goal.

Elizabeth Daly, Giusy Di Lorenzo, and Daniele Quercia served as chairs during the workshop sessions. This report was written by Elizabeth M. Daly, Giusy Di Lorenzo, Daniele Quercia, and Michael Muller. The papers of the When the City Meets the Citizen workshop were published as AAAI Technical Report WS-12-04.

Daniel Archambault is a postdoctoral researcher at University College Dublin.

Ruben Bouwmeester is a project manager, innovation projects at Deutsche Welle.

Cosmin Cabulea is a media scientist at Deutsche Welle.

Elizabeth M. Daly is a research scientist at IBM Research – Ireland.

Maarten de Rijke is a professor at the University of Amsterdam.

Giusy Di Lorenzo is a research scientist at IBM Research – Ireland.

Martin Harrigan is a postdoctoral researcher at University College Dublin.

Eser Kandogan is a researcher at IBM Research – Almaden.

Michael Muller is a research staff member at IBM Research – Cambridge, USA.

Mor Naaman is an assistant professor at the Rutgers School of Communication and Information.

Daniele Quercia is a senior researcher at the University of Cambridge.

Damiano Spina is a Ph.D. candidate and research fellow at the NLP & IR Group at UNED (Madrid, Spain)

Markus Strohmaier is an assistant professor at the Faculty of Computer Science at Graz University of Technology.

Arkaitz Zubiaga is a postdoctoral research associate at the City University of New York.