Overview of the Third International Workshop on Search and Mining User-generated Contents

Iván Cantador
Univ. Autónoma de Madrid
Madrid, Spain
ivan.cantador@uam.es

José Carlos Cortizo
BrainSins
Alcorcón, Madrid, Spain
jccortizo@brainsins.com

Francisco Carrero
BrainSins
Alcorcón, Madrid, Spain
fcarrero@brainsins.com

Jose Antonio Troyano
Universidad de Sevilla
Sevilla, Spain
troyano@us.es

Paolo Rosso
Univ. Politécnica de Valencia
Valencia, Spain
prosso@dsic.upv.es

Markus Schedl
Johannes Kepler University
Linz, Austria
markus.schedl@jku.at

ABSTRACT

In this paper, we provide an overview of the 3rd International Workshop on Search and Mining User-generated Contents, held in conjunction with the 20th ACM International Conference on Information and Knowledge Management. We present the motivation and goals of the workshop, and some statistics and details about accepted papers and keynotes.

Categories and Subject Descriptors

H.3 [**Information Systems**]: Information Storage and Retrieval; H.4 [**Information Systems**]: Information Systems Applications; I.2 [**Computing Methodologies**]: Artificial Intelligence; I.7 [**Computing Methodologies**]: Document and Text Processing

General Terms

Documentation, Experimentation, Algorithms

Keywords

user-generated contents, social media, data mining, text mining, opinion mining, information retrieval

1. MOTIVATION AND GOALS

Nowadays, the huge amount and variety of user generated contents available in the Web open a wide range of opportunities to enhance information retrieval and e-commerce applications. Opinions and reviews about products, annotations and bookmarks on multimedia resources, and friend relations in social networks are just a few examples of personal information sources to be exploited in order to both improve the user's experience and increase the companies' revenues in online search and commerce activities.

As in its previous editions, the 3rd Workshop on Search and Mining User-generated Contents (SMUC 2011) aimed at becoming a multidisciplinary forum for researchers and practitioners that work on knowledge extraction, management and exploitation in Social Media, and belong to different,

but complementary fields such as Web (content/structure/usage) mining, information retrieval, opinion mining and sentiment analysis, user modeling, personalization and recommendation, and multimedia processing and retrieval.

In SMUC 2011 workshop, we identify four main research themes into which the above research problems can be categorized: Searching in Social Media, Mining Social Media, Opinion Mining and Sentiment Analysis, and Multimedia Processing and Retrieval. For all of them, we are interested in developing and testing intelligent systems and applications, involving innovative research from the fields of user modeling, personalization, recommendation, information visualization, and business intelligence, to name a few. Different research lines, backgrounds, perspectives and degrees of expertise were presented at the workshop, and thus very interesting multidisciplinary discussions, collaborations and work synergies between the workshop attendees were expected as main outcomes of the event.

2. ACCEPTED PAPERS

The SMUC 2011 workshop received a total of 17 submissions, 12 from Europe, 3 from Asia and 2 from North America. Out of these, 11 (65%) were accepted as full papers.

Categorizing the submissions according to the main topics of the workshop yields the following distribution: seven papers focus on data/text mining, five papers focus on search, information filtering and recommender systems, three papers are related to multimedia mining, and two papers address opinion mining topics.

Another categorization of the papers can be considered based on the different information sources used for experimentation. Twitter represents by far the most popular information source among contributions as five of the accepted papers rely on the microblogging service. Another two papers use multimedia sources like Flickr, TV and YouTube, and the remaining papers make use of Wikipedia, product reviews web sites, SMS, and other sources.

The following are the papers presented at SMUC 2011 workshop:

- "I'm Eating a Sandwich in Glasgow: Modeling Locations with Tweets" [4] describes an interesting approach, tested on a large dataset, for locating tweets based on language modeling.

- "Mining Tag Similarity in Folksonomies" [7] proposes a hybrid (linguistic/semantic) approach to determining similarity amongst the user-defined tags of a folksonomy.

- "On the Generation of Rich Content Metadata from Social Media" [3] addresses the problem of analyzing content quality as a means to enhance the retrieval and filtering of tweets related to TV shows.

- "Predicting Age and Gender in Online Social Networks" [5] describes a text categorization approach for the identification of age and gender in social media texts.

- "A Comparative Evaluation of Personality Estimation Algorithms for the TWIN Recommender System" [6] presents a comparative evaluation of personality estimation algorithms for the TWIN ("Tell me What I Need") recommender system.

- "Mining Twitter Tweets to Tag Flickr Photos and YouTube Videos" [9] presents an analysis of the exploitation of hashtags and tweets contents for the automatic annotation of web resources, such as Flickr photos and Youtube videos.

- "ThemeCrowds: Multiresolution Summaries of Twitter Usage" [1] presents an interesting visualization interface that shows users and tags on tweets using hierarchical clustering for the generation of treemaps and antichains to visualize information incrementally.

- "Detection of Near-duplicate User Generated Contents: The SMS Spam Collection" [10] deals with SMS spam filtering, and examines the use of existing plagiarism detection tools to detect near-duplicates in SMS messages.

- "Characterizing Wikipedia Pages Using Edit Network Motif Profiles" [11] attempts to characterize the quality (class) of Wikipedia pages, and determine their authoritativeness based on patterns of page edit networks.

- "Improved Answer Ranking in Social Question-Answering Portals" [2] deals with the question of how to identify good answers for user questions in community Question Answering portals, improving answer ranking by learning from preference pairs of marked questions.

- "Trend-based and Reputation-Versed Personalized News Network" [8] describes an approach to filtering Twitter and Facebook messages using trend mining with respect to the reputation of sources.

3. KEYNOTES

Three invited speakers presented very interesting talks about main topics of SMUC 2011 workshop: information retrieval, opinion mining, and social network and community analysis in Social Media.

- "Analysis of communities in Social Media", by Martin Atzmüller, from University of Kassel, Germany.

- "Information Retrieval techniques for Social Media", by Joemon M. Jose, from University of Glasgow, UK.

- "The challenge of understanding the flow of sentiments in Social Media", by David E. Losada, from Universidad de Santiago de Compostela, Spain.

4. ACKNOWLEDGMENTS

5. REFERENCES

[1] D. Archambault, D. Greene, P. Cunningham, and N. Hurley. ThemeCrowds: Multiresolution Summaries of Twitter Usage. In *Proc. of the 3rd Workshop on Search and Mining User-generated Contents*, Glasgow, UK, October 2011.

[2] F. Hieber and S. Riezler. Improved Answer Ranking in Social Question-Answering Portals. In *Proc. of the 3rd Workshop on Search and Mining User-generated Contents*, Glasgow, UK, October 2011.

[3] G. Inches, A. Basso, and F. Crestani. On The Generation of Rich Content Metadata from Social Media. In *Proc. of the 3rd Workshop on Search and Mining User-generated Contents*, Glasgow, UK, October 2011.

[4] S. Kinsella, V. Murdock, and N. O'Hare. "I'm Eating a Sandwich in Glasgow": Modeling Locations with Tweets. In *Proc. of the 3rd Workshop on Search and Mining User-generated Contents*, Glasgow, UK, October 2011.

[5] C. Peersman, W. Daelemans, and L. V. Vaerenbergh. Predicting Age and Gender in Online Social Networks. In *Proc. of the 3rd Workshop on Search and Mining User-generated Contents*, Glasgow, UK, October 2011.

[6] A. Roshchina, J. Cardiff, and P. Rosso. A Comparative Evaluation of Personality Estimation Algorithms for the TWIN Recommender System. In *Proc. of the 3rd Workshop on Search and Mining User-generated Contents*, Glasgow, UK, October 2011.

[7] G. Solskinnsbakk and J. A. Gulla. Mining Tag Similarity in Folksonomies. In *Proc. of the 3rd Workshop on Search and Mining User-generated Contents*, Glasgow, UK, October 2011.

[8] O. Streibel and R. Alnemr. Trend-based and Reputation-Versed Personalized News Network. In *Proc. of the 3rd Workshop on Search and Mining User-generated Contents*, Glasgow, UK, October 2011.

[9] A. Sureka and D. Correa. Mining Twitter Tweets to Tag Flickr Photos and YouTube Videos. In *Proc. of the 3rd Workshop on Search and Mining User-generated Contents*, Glasgow, UK, October 2011.

[10] E. Vallés and P. Rosso. Detection of Near-duplicate User Generated Contents: The SMS Spam Collection. In *Proc. of the 3rd Workshop on Search and Mining User-generated Contents*, Glasgow, UK, October 2011.

[11] G. Wu, M. Harrigan, and P. Cunningham. Characterizing Wikipedia Pages Using Edit Network Motif Profiles. In *Proc. of the 3rd Workshop on Search and Mining User-generated Contents*, Glasgow, UK, October 2011.