

INTERNATIONAL WORKSHOP
EVENTS IN EMERGING TEXT TYPES (eETTs)

held in conjunction with the International Conference
RANLP - 2009, 14-16 September 2009, Borovets, Bulgaria

PROCEEDINGS

Edited by
Constantin Orăsan, Laura Hasler and Corina Forăscu

Borovets, Bulgaria

17 September 2009

International Workshop

EVENTS IN EMERGING TEXT TYPES (eETTs)

PROCEEDINGS

Borovets, Bulgaria

17 September 2009

ISBN 978-954-452-011-3

Designed and Printed by INCOMA Ltd.
Shoumen, Bulgaria

Foreword

The proliferation of the Internet has revolutionised the way information is disseminated and presented. Blogs no longer just relay and comment on news stories but also influence what is talked about in the news. Such changes have not gone unnoticed by the computational linguistics research community, which is increasingly processing or exploiting blogs in an attempt to keep track of what is going on and mine information. This workshop focuses on how events can be identified and how information related to event processing (e.g. NP coreference, temporal processing) can be extracted from blogs and other online sources. Emphasis is on how existing methods for event processing need to be adapted in order to process this medium, and on linguistic differences in the reporting of events in blogs and more traditional news texts.

Event detection and processing is not a new topic in computational linguistics, but until now it has focused mainly on processing of newswire. The TimeBank corpus (Pustejovsky et. al. 2003), the AQUAINT TimeML corpus, and the NP4E corpus (Hasler, Orasan and Naumann 2006) exclusively contain newswire, which may make them inappropriate for the development of methods which need to process other text types. Moreover, the informal style and structure of most blog entries makes event detection in these documents a difficult task. This workshop gives researchers the opportunity to present efforts to develop resources related to event identification and processing using blog entries, including annotation guidelines and linguistic analyses of such resources.

We would like to thank the organisers of the 7th International Conference on Recent Advances in Natural Language Processing, RANLP 2009, for hosting this workshop as one of their satellite events. We would also like to thank the researchers who submitted papers and the members of Programme Committee for their help in the reviewing process.

We hope you enjoy the workshop,

The organising committee

Organisers and Sponsors

Events in Emerging Text Types (eETTs) is organised by:

Constantin Orăsan, University of Wolverhampton, UK

Laura Hasler, University of Wolverhampton, UK

Corina Forăscu, “A.I. Cuza” University of Iași, Romania

PROGRAMME COMMITTEE

Alexandra Balahur, Alicante University, Spain
Carmen Banea, University of North Texas, USA
Sabin-Corneliu Buraga, “A.I. Cuza” University of Iași, Romania
Di Cai, University of Wolverhampton, UK
Andras Csomai, Google
Iustin Dornescu, University of Wolverhampton, UK
Erhard Hinrichs, Tübingen University, Germany
Véronique Hoste, University College Gent, Belgium
Radu Ion, Research Institute for AI, Romanian Academy
Rafael Muñoz, University of Alicante, Spain
Vivi Năstase, EML Research, Germany
Nicolas Nicolov, Umbria Inc, USA
Georgios Paltoglou, University of Wolverhampton, UK
Livia Polanyi, Powerset/Microsoft, USA
Irina Prodanof, ILC-CNR, Pisa, Italy
Mark Rogers, Market Sentinel, UK
Doina Tătar, “Babes-Bolyai” University, Romania

Table of Contents

<i>Summarizing Blog Entries versus News Texts</i> Shamima Mithun and Leila Kosseim	1
<i>Event Detection in Blogs using Temporal Random Indexing</i> David Jurgens and Keith Stevens	9
<i>A Pairwise Event Coreference Model, Feature Impact and Evaluation for Event Coreference Resolution</i> Zheng Chen, Heng Ji and Robert Haralick	17
<i>Summarizing Threads in Blogs Using Opinion Polarity</i> Alexandra Balahur, Elena Lloret, Ester Boldrini, Andrés Montoyo, Manuel Palomar and Patricio Martínez-Barco	23
<i>Catching the news: two key cases from today</i> Ruslana Margova and Irina Temnikova	32
<i>Detecting Opinion Sentences Specific to Product Features in Customer Reviews using Typed Dependency Relations</i> Ashequl Qadir	38

Workshop Program

Friday, September 17, 2009

Session 1

- 9:20–9:30 Welcome and opening remarks
- 9:30–10:30 Invited Talk by Mark Rogers: *The value of language processing to commercial clients*
- 10:30–11:00 *A Pairwise Event Coreference Model, Feature Impact and Evaluation for Event Coreference Resolution*
Zheng Chen, Heng Ji and Robert Haralick

Session 2

- 11:30–12:00 *Catching the news: two key cases from today*
Ruslana Margova and Irina Temnikova
- 12:00–12:30 *Summarizing Blog Entries versus News Texts*
Shamima Mithun and Leila Kosseim

Session 3

- 14:00–15:00 Invited Talk by Mike Thelwall: *Detecting public news interests from blogs and MySpace*
- 15:00–15:30 *Summarizing Threads in Blogs Using Opinion Polarity*
Alexandra Balahur, Elena Lloret, Ester Boldrini, Andrés Montoyo, Manuel Palomar and Patricio Martínez-Barco

Session 4

- 16:00–16:30 *Detecting Opinion Sentences Specific to Product Features in Customer Reviews using Typed Dependency Relations*
Ashequl Qadir
- 16:30–17:00 *Event Detection in Blogs using Temporal Random Indexing*
David Jurgens and Keith Stevens
- 17:00–17:30 Round table/Closing session

