

**3rd International Workshop on Natural  
Language Processing for Informal Text**

# Preface

It is our great pleasure to welcome you to the *3rd International Workshop on Natural Language Processing for Informal Text (NLPIT 2017)*, associated with ICWE 2017. The rapid growth of Internet usage in the last two decades adds new challenges to understand the informal user generated content (UGC) on the Internet. Textual UGC refers to textual posts on social media, blogs, emails, chat conversations, instant messages, forums, reviews, or advertisements that are created by end-users of an online system. A large portion of language used on textual UGC is informal. Informal text is the style of writing that disregards language grammars and uses a mixture of abbreviations and context dependent terms. The straightforward application of state-of-the-art Natural Language Processing approaches on informal text typically results in a significantly degraded performance due to the following reasons: the lack of sentence structure; the lack of enough context required; the seldom entities involved; the noisy sparse contents of users' contributions; and the untrusted facts contained. The NLPIT workshop hopes to bring opportunities and challenges involved in informal text processing under the attention of researchers. In particular, we are interested in discussing informal text modeling, normalization, mining, and understanding in addition to various application areas in which UGC is involved. The workshop is a follow-up of the first NLPIT workshop that was held in conjunction with ICWE 2015: the International Conference on Web Engineering held in Rotterdam, The Netherlands, from 23rd to 26th of July 2015 and the 2nd NLPIT workshop that was held in conjunction with WWW 2016: the International Conference on World Wide Web held in Montreal, Canada, from 11th to 15th of April 2016. The call for papers attracted 6 submissions from 6 different countries. Each paper received at least 3 reviews. The program committee reviewed and accepted 3 papers with an acceptance rate of 50%.

The workshop program started with a keynote presentation given by Roberto Navigli<sup>1</sup> from the Sapienza University of Rome entitled “*Overcoming the Language Barrier with BabelNet and Multilingual Disambiguation of Text*”. The keynote is followed by 3 research presentations. The first paper discussed the characteristics of the Egyptian dialect. The second paper presented a multi-level novel architecture to classify named entities in tweets that uses features extracted from images and text. Finally, the last paper studied the positive impact of online expectation maximization methods for estimating language distributions and improving per-document language identification from those estimates.

---

<http://wwwhome.cs.utwente.nl/~badiehm/nlpit2017/>.

<sup>1</sup> <http://wwwusers.di.uniroma1.it/~navigli/>.

We thank all authors for enabling us to offer an attractive program. We also thank the program committee for reviewing the papers and writing extensive reports with feedback for the authors. Finally, we would like to thank the ICWE 2017 workshop chairs for support in organizing the workshop.

Mena B. Habib  
Florian Kunneman  
Maurice van Keulen

# Organization

## Program Committee

Alexandra Balahur	The European Commission's Joint Research Centre (JRC), Italy
Barbara Plank	University of Copenhagen, Denmark
Chenliang Li	Wuhan University, China
Claudia Hauff	Delft University, The Netherlands
Dolf Trieschnigg	My Data Factory, The Netherlands
Erik Tjong Kim Sang	Meertens Institute, The Netherlands
Gerasimos Spanakis	Maastricht University, The Netherlands
Heba Elfardy	Columbia University, USA
Kevin Gimpel	Toyota Technological Institute, USA
Malvina Nissim	University of Groningen, The Netherlands
Orphee De Clerq	Ghent University, Belgium
Robert Remus	ExB Group, Germany
Sabine Bergler	Concordia University, Canada
Wang Ling	Carnegie Mellon University, USA
Yannis Korkontzelos	Edge Hill University, UK
Zhemín Zhu	Elsevier, The Netherlands