

## Preface

Welcome to the proceedings of the 14th China National Conference on Computational Linguistics (14th CCL) and the Third International Symposium on Natural Language Processing Based on Naturally Annotated Big Data (3rd NLP-NABD). The conference was hosted by Guangdong University of Foreign Studies.

CCL is an annual conference (bi-annual before 2013) that started in 1991. It is the flagship conference of the Chinese Information Processing Society (CIPS), which is the largest NLP scholar and expert community in China. CCL is a premier nation-wide forum for disseminating new scholarly and technological work in computational linguistics, with a major emphasis on computer processing of the languages in China such as Mandarin, Tibetan, Mongolian, and Uyghur.

Affiliated with the 14th CCL, the Third International Symposium on Natural Language Processing Based on Naturally Annotated Big Data (NLP-NABD) covered all the NLP topics, with particular focus on methodologies and techniques relating to naturally annotated big data. In contrast to manually annotated data such as treebanks that are constructed for specific NLP tasks, naturally annotated data come into existence through users' normal activities, such as writing, conversation, and interactions on the Web. Although the original purposes of these data typically were unrelated to NLP, they can nonetheless be purposefully exploited by computational linguists to acquire linguistic knowledge. For example, punctuation marks in Chinese text can help word boundaries identification, social tags in social media can provide signals for keyword extraction, and categories listed in Wikipedia can benefit text classification. The natural annotation can be explicit, as in the aforementioned examples, or implicit, as in Hearst patterns (e.g., "Beijing and other cities" implies "Beijing is a city"). This symposium focuses on numerous research challenges ranging from very-large-scale unsupervised/semi-supervised machine learning (deep learning, for instance) of naturally annotated big data to integration of the learned resources and models with existing handcrafted "core" resources and "core" language computing models. NLP-NABD 2015 was supported by the National Key Basic Research Program of China (i.e., "973" Program) "Theory and Methods for Cyber-Physical-Human Space-Oriented Web Chinese Information Processing" under grant no. 2014CB340500.

The Program Committee selected 115 papers (81 Chinese papers and 34 English papers) out of 283 submissions from China, Hong Kong (region), Singapore, and the USA for publication. The 34 English papers cover the following topics:

- Lexical semantics and ontologies (4)
- Semantics (4)
- Sentiment analysis, opinion mining and text classification (2)
- Machine translation (3)
- Multilinguality in NLP (4)
- Machine learning method for NLP (3)
- Knowledge graph and information extraction (3)

- Discourse, coreference, and pragmatics (1)
- Information retrieval and question answering (3)
- Social computing (2)
- NLP applications (5)

The final program for the 14th CCL and the Third NLP-NABD was the result of a great deal of work by many dedicated colleagues. We want to thank, first of all, the authors who submitted their papers, and thus contributed to the creation of the high-quality program that allowed us to look forward to an exciting joint conference. We are deeply indebted to all the Program Committee members for providing high-quality and insightful reviews under a tight schedule. We are extremely grateful to the sponsors of the conference. Finally, we extend a special word of thanks to all the colleagues of the Organizing Committee and secretariat for their hard work in organizing the conference, and to Springer for their assistance in publishing the proceedings in due time.

We thank the Program and Organizing Committees for helping make the conference successful and we hope the participants enjoy a memorable visit to Guangzhou.

November 2015

*CCL Program Committee Chairs:*

Maosong Sun

Keh-Yih Su

*CCL Program Committee Co-Chairs:*

Jun Zhao

Min Zhang

Heng Ji

*NLP-NABD Program Committee Chairs:*

Maosong Sun

Randy Goebel

Ting Liu

Chinese Computational Linguistics and Natural  
Language Processing Based on Naturally Annotated Big  
Data

14th China National Conference, CCL 2015 and Third  
International Symposium, NLP-NABD 2015, Guangzhou,  
China, November 13-14, 2015, Proceedings

Sun, M.; Liu, Z.; Zhang, M.; Liu, Y. (Eds.)

2015, XVI, 426 p. 97 illus. in color., Softcover

ISBN: 978-3-319-25815-7