

## Preface

Welcome to the proceedings of the 15th China National Conference on Computational Linguistics (15th CCL) and the Fourth International Symposium on Natural Language Processing Based on Naturally Annotated Big Data (4th NLP-NABD). The conference and symposium were hosted by Ludong University located in Yantai City, Shandong Province, China.

CCL is an annual conference (bi-annual before 2013) that started in 1991. It is the flagship conference of the Chinese Information Processing Society of China (CIPS), which is the largest NLP scholar and expert community in China. CCL is a premier nationwide 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 15th CCL, the Fourth 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 2016 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 and the Major Project of the National Social Science Foundation of China under grant no. 13&ZD190.

The Program Committee selected 101 papers (64 Chinese papers and 37 English papers) out of 307 submissions from China, Hong Kong (region), Singapore, and USA for publication. The total accepted rate is 32.9 %. The 37 English papers cover the following topics:

- Semantics (3)
- Machine translation (5)
- Multilinguality in NLP (6)
- Knowledge graph and information extraction (5)

- Linguistic resource annotation and evaluation (2)
- Information retrieval and question answering (2)
- Text classification and summarization (5)
- Social computing and sentiment analysis (5)
- NLP applications (4)

The final program for the 15th CCL and the fourth 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 to make the conference successful, and we hope all the participants enjoyed a remarkable visit to Yantai, a historical and beautiful seaside city in East China.

August 2016

Maosong Sun  
Tat-Seng Chua  
Xuanjing Huang  
Hongfei Lin  
Jian-Yun Nie  
Jun Zhao

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

15th China National Conference, CCL 2016, and 4th  
International Symposium, NLP-NABD 2016, Yantai,  
China, October 15-16, 2016, Proceedings

Sun, M.; Huang, X.; Lin, H.; Liu, Z.; Liu, Y. (Eds.)

2016, XVIII, 460 p. 139 illus., Softcover

ISBN: 978-3-319-47673-5