ACL 2016

Proceedings of
NEWS 2016
The Sixth Named Entities Workshop

Xiangyu Duan, Rafael E. Banchs, Min Zhang, Haizhou Li, A. Kumara (Editors)

August 12, 2016
Berlin, Germany

©2016 The Association for Computational Linguistics

Order copies of this and other ACL proceedings from:

Association for Computational Linguistics (ACL)
209 N. Eighth Street
Stroudsburg, PA 18360
USA
Tel: +1-570-476-8006
Fax: +1-570-476-0860
acl@aclweb.org

ISBN 978-1-945626-16-6

Introduction

The workshop series, Named Entities WorkShop (NEWS), focus on research on all aspects of the Named Entities, such as, identifying and analyzing named entities, mining, translating and transliterating named entities, etc. The first of the NEWS workshops (NEWS 2009) was held as a part of ACL-IJCNLP 2009 conference in Singapore; the second one, NEWS 2010, was held as an ACL 2010 workshop in Uppsala, Sweden; the third one, NEWS 2011, was held as an IJCNLP 2011 workshop in Chiang Mai, Thailand; the fourth one, NEWS 2012, was held as an ACL 2012 workshop in Jeju, Korea; and the fifth one, NEWS 2015, was held as ACL2015 workshop in Bejing, China. The current edition, NEWS 2016, was held as an ACL 2016 workshop in Berlin, Germany.

The purpose of the NEWS workshop series is to bring together researchers across the world interested in identification, analysis, extraction, mining and transformation of named entities in monolingual or multilingual natural language text corpora. The workshop scope includes many interesting specific research areas pertaining to the named entities, such as, orthographic and phonetic characteristics, corpus analysis, unsupervised and supervised named entities extraction in monolingual or multilingual corpus, transliteration modeling, and evaluation methodologies, to name a few. For this year edition, 6 research papers were submitted, each paper was reviewed by at least 2 reviewers from the program committee. The 6 papers were all chosen for publication, covering named entity recognition and machine transliteration, which applied various new trend methods such as deep neural networks and graph-based semi-supervised learning.

Following the tradition of the NEWS workshop series, NEWS 2016 continued the machine transliteration shared task this year as well. The shared task was first introduced in NEWS 2009 and continued in NEWS 2010, NEWS 2011, NEWS 2012, and NEWS 2015. In NEWS 2016, by leveraging on the previous success of NEWS workshop series, we released the hand-crafted parallel named entities corpora to include 14 different language pairs from 12 language families, and made them available as the common dataset for the shared task. In total, 5 international teams participated from around the globe. The approaches ranged from traditional learning methods (such as, Phrasal SMT-based, Conditional Random Fields, etc.) to somewhat new approaches (such as, neural network transduction, integration of transliteration mining, hybrid system combination). A concrete study and targeted process between two languages often generate better performances. A report of the shared task that summarizes all submissions and the original whitepaper are also included in the proceedings, and will be presented in the workshop. The participants in the shared task were asked to submit short system papers (4 content pages each) describing their approaches, and each of such papers was reviewed by at least two members of the program committee to help improve the quality.

We hope that NEWS 2016 would provide an exciting and productive forum for researchers working in this research area, and the NEWS-released data continues to serve as a standard dataset for machine transliteration generation and mining. We wish to thank all the researchers for their research submission and the enthusiastic participation in the transliteration shared tasks. We wish to express our gratitude to CJK Institute, Institute for Infocomm Research, Microsoft Research India, Thailand National Electronics and Computer Technology Centre and The Royal Melbourne Institute of Technology (RMIT)/Sarvnaz Karimi for preparing the data released as a part of the shared tasks. Finally, we thank all the program committee members for reviewing the submissions in spite of the tight schedule.

Workshop Organizers:
Xiangyu Duan, Soochow University, China
Rafael E Banchs, Institute for Infocomm Research, Singapore
Min Zhang, Soochow University, China
Haizhou Li, Institute for Infocomm Research, Singapore
A Kumaran, Microsoft Research, India

August 12, 2016
Berlin, Germany

Organizers:

Xiangyu Duan, Soochow University
Rafael E Banchs, Institute for Infocomm Research, Singapore
Min Zhang, Soochow University, China
Haizhou Li, Institute for Infocomm Research, Singapore
A Kumaran, Microsoft Research, India

Program Committee:

Rafael E. Banchs, Institute for Infocomm Research
Sivaji Bandyopadhyay, Jadavpur University
Marta R. Costa-jussà, Instituto Politécnico Nacional
Xiangyu Duan, Soochow University
Guohong Fu, Heilongjiang University
Sarvnaz Karimi, CSIRO
Mitesh M. Khapra, IBM Research India
Grzegorz Kondrak, University of Alberta
Jong-Hoon Oh, NICT
Richard Sproat, Google
Keh-Yih Su, Institute of Information Science, Academia Sinica
Raghavendra Udupa, Microsoft Research India
Chai Wutiwiwatchai, Intelligent Informatics Research Unit, National Electronics and Computer Technology Center
Deyi Xiong, Soochow University
Muyun Yang, Harbin Institute of Technology
Min Zhang, Soochow University

Table of Contents

Leveraging Entity Linking and Related Language Projection to Improve Name Transliteration
Ying Lin, Xiaoman Pan, Aliya Deri, Heng Ji and Kevin Knight 1

Multi-source named entity typing for social media
Reuth Vexler and Einat Minkov 11

Evaluating and Combining Name Entity Recognition Systems
Ridong Jiang, Rafael E. Banchs and Haizhou Li 21

German NER with a Multilingual Rule Based Information Extraction System: Analysis and Issues
Anna Druzhkina, Alexey Leontyev and Maria Stepanova 28

Spanish NER with Word Representations and Conditional Random Fields
Jenny Linet Copara Zea, Jose Eduardo Ochoa Luna, Camilo Thorne and Goran Glavaš 34

Constructing a Japanese Basic Named Entity Corpus of Various Genres
Tomoya Iwakura, Kanako Komiya and Ryuichi Tachibana 41

Linguistic Issues in the Machine Transliteration of Chinese, Japanese and Arabic Names
Jack Halpern 47

Whitepaper of NEWS 2016 Shared Task on Machine Transliteration
Xiangyu Duan, Min Zhang, Haizhou Li, Rafael Banchs and A. Kumaran 49

Report of NEWS 2016 Machine Transliteration Shared Task
Xiangyu Duan, Rafael Banchs, Min Zhang, Haizhou Li and A. Kumaran 58

Applying Neural Networks to English-Chinese Named Entity Transliteration
Yan Shao and Joakim Nivre 73

Target-Bidirectional Neural Models for Machine Transliteration
Andrew Finch, Lemao Liu, Xiaolin Wang and Eiichiro Sumita 78

Regulating Orthography-Phonology Relationship for English to Thai Transliteration
Binh Minh Nguyen, Hoang Gia Ngo and Nancy F. Chen 83

Moses-based official baseline for NEWS 2016
Marta R. Costa-jussà 88

Conference Program

9:00–9:10 **Welcome Message and Workshop Presentation by the organizing team**

9:10–10:30 **Research Papers I**

9:10–9:40 *Leveraging Entity Linking and Related Language Projection to Improve Name Transliteration*
Ying Lin, Xiaoman Pan, Aliya Deri, Heng Ji and Kevin Knight

9:40–10:10 *Multi-source named entity typing for social media*
Reuth Vexler and Einat Minkov

10:10–10:30 *Evaluating and Combining Name Entity Recognition Systems*
Ridong Jiang, Rafael E. Banchs and Haizhou Li

10:30–11:00 **Coffee Break**

11:00–12:30 **Research Papers II**

11:00–11:30 *German NER with a Multilingual Rule Based Information Extraction System: Analysis and Issues*
Anna Druzhkina, Alexey Leontyev and Maria Stepanova

11:30–12:00 *Spanish NER with Word Representations and Conditional Random Fields*
Jenny Linet Copara Zea, Jose Eduardo Ochoa Luna, Camilo Thorne and Goran Glavaš

12:00–12:30 *Constructing a Japanese Basic Named Entity Corpus of Various Genres*
Tomoya Iwakura, Kanako Komiya and Ryuichi Tachibana

12:30–14:30 Lunch Break

14:30–15:30 Keynote Speech

14:30–15:30 *Linguistic Issues in the Machine Transliteration of Chinese, Japanese and Arabic Names*
Jack Halpern

15:30–16:00 Coffee Break

16:00–17:20 Shared Task on Name Entity Transliteration

16:00–16:10 *Whitepaper of NEWS 2016 Shared Task on Machine Transliteration*
Xiangyu Duan, Min Zhang, Haizhou Li, Rafael Banchs and A. Kumaran

16:00–16:10 *Report of NEWS 2016 Machine Transliteration Shared Task*
Xiangyu Duan, Rafael Banchs, Min Zhang, Haizhou Li and A. Kumaran

16:10–16:30 *Applying Neural Networks to English-Chinese Named Entity Transliteration*
Yan Shao and Joakim Nivre

16:30–16:50 *Target-Bidirectional Neural Models for Machine Transliteration*
Andrew Finch, Lemao Liu, Xiaolin Wang and Eiichiro Sumita

16:50–17:10 *Regulating Orthography-Phonology Relationship for English to Thai Transliteration*
Binh Minh Nguyen, Hoang Gia Ngo and Nancy F. Chen

17:10–17:20 *Moses-based official baseline for NEWS 2016*
Marta R. Costa-jussà

17:20–17:30 Closing Remarks