ACL 2018

Named Entities
Proceedings of the Seventh Workshop

Nancy Chen, Rafael E. Banchs, Xiangyu Duan, Min Zhang, Haizhou Li

July 20, 2018
Melbourne, Australia

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 978-1-948087-37-7

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; and the fourth one, NEWS 2012, was held as an ACL 2012 workshop in Jeju, Korea. The fifth one, NEWS 2015, was held as an ACL-IJCNLP 2015 workshop in Beijing, China. The sixth one, NEWS2016, was held as an ACL 2016 workshop in Berlin, Germany. The current edition, NEWS2018, was held as an ACL 2018 workshop in Melbourne, Australia.

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, 7 research papers were submitted, each paper was reviewed by at least 2 reviewers from the program committee. The 7 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 2018 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, NEWS 2015, and NEWS 2016. In NEWS 2018, by leveraging on the previous success of NEWS workshop series, the Shared Task featured 19 tasks on proper name transliteration, including 13 different languages and two different Japanese scripts. A total of 6 teams from 8 different institutions participated in the evaluation, submitting 424 runs, involving different transliteration methodologies.

We hope that NEWS 2018 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 (Japan), Institute for Infocomm Research (Singapore), National University of Singapore (NUS), Artificial Intelligence Laboratory at the Ho Chi Minh City University of Science (AILab, VNU-HCMUS, Vietnam), Microsoft Research India, the Computer Science & Engineering Department of Jadavpur University (India), the National Electronics and Computer Technology Center (NECTEC, Thailand) and Sarvnaz Karim (RMIT, Australia) for providing the corpora and technical support for the shared task. Without those, the Shared Task would not be possible. In addition, we want to thank Grandee Lee and Snigdha Singhania for their help and support with CodaLab and the baseline systems, respectively. Finally, we thank all the program committee members for reviewing the submissions in spite of the tight schedule.

Organizers:

Nancy Chen, Singapore University of Technology and Design
Rafael E. Banchs, Nanyang Technological University, Singapore
Xiangyu Duan, Soochow University
Min Zhang, Soochow University, China
Haizhou Li, Institute for Infocomm Research, Singapore

Program Committee:

Rafael E. Banchs, Nanyang Technological University
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

Conference Program

Friday, 20 July, 2018

8:30–8:40 **Opening Remarks**

8:40–9:00 *Automatic Extraction of Entities and Relation from Legal Documents*
Judith Jeyafreeda Andrew

9:00–9:20 *Connecting Distant Entities with Induction through Conditional Random Fields for Named Entity Recognition: Precursor-Induced CRF*
Wangjin Lee and Jinwook Choi

9:20–9:40 *A Sequence Learning Method for Domain-Specific Entity Linking*
Emrah Inan and Oguz Dikenelli

9:40–10:00 *Attention-based Semantic Priming for Slot-filling*
Jiewen Wu, Rafael E. Banchs, Luis Fernando D'Haro, Pavitra Krishnaswamy and Nancy Chen

10:00–10:20 *Named Entity Recognition for Hindi-English Code-Mixed Social Media Text*
Vinay Singh, Deepanshu Vijay, Syed Sarfaraz Akhtar and Manish Shrivastava

10:30–11:00 ***Coffee Break***

11:00–11:20 *Forms of Anaphoric Reference to Organisational Named Entities: Hoping to widen appeal, they diversified*
Christian Hardmeier, Luca Bevacqua, Sharid Loáiciga and Hannah Rohde

11:20–11:40 *Named-Entity Tagging and Domain adaptation for Better Customized Translation*
Zhongwei Li, Xuancong Wang, AiTi Aw, Eng Siong Chng and Haizhou Li

12:00–14:00 ***Lunch***

14:10–14:20 *NEWS 2018 Whitepaper*
Nancy Chen, Xiangyu Duan, Min Zhang, Rafael E. Banchs and Haizhou Li

14:20–14:40 *Report of NEWS 2018 Named Entity Transliteration Shared Task*
Nancy Chen, Rafael E. Banchs, Min Zhang, Xiangyu Duan and Haizhou Li

Friday, 20 July, 2018 (continued)

14:40–15:00 *Statistical Machine Transliteration Baselines for NEWS 2018*
Snigdha Singhania, Minh Nguyen, Gia H Ngo and Nancy Chen

15:00–15:20 *A Deep Learning Based Approach to Transliteration*
Soumyadeep Kundu, Sayantan Paul and Santanu Pal

15:30–16:00 ***Coffee Break***

16:00–16:20 *Comparison of Assorted Models for Transliteration*
Saeed Najafi, Bradley Hauer, Rashed Rubby Riyadh, Leyuan Yu and Grzegorz Kondrak

16:20–16:40 *Neural Machine Translation Techniques for Named Entity Transliteration*
Roman Grundkiewicz and Kenneth Heafield

16:40–17:00 *Low-Resource Machine Transliteration Using Recurrent Neural Networks of Asian Languages*
Ngoc Tan Le and Fatiha Sadat