NAACL HLT 2016

Workshop on Multilingual and Cross-lingual Methods in NLP

Proceedings of the Workshop

June 17, 2016
San Diego, California, USA

We thank our sponsor Google Inc. for a generous support.

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 978-1-941643-87-7

Introduction

The goal of this workshop is to expand the current area of cross-lingual learning to include more NLP problems, encourage approaches that explore low-resource scenarios, and improve upon existing approaches to multilinguality.

State-of-the-art NLP tools such as text parsing, speech recognition and synthesis, text and speech translation, semantic analysis and inference, rely on availability of language-specific data resources that exist only for a few resource-rich languages. To make NLP tools available in more languages, techniques have been developed for projecting such resources from resource-rich languages using parallel (translated) data as a bridge for cross-lingual NLP applications. The limiting reagent in these methods is parallel data or bilingual lexicons. While small parallel corpora do exist for many languages, suitably large parallel corpora are expensive, and these typically exist only for English and a few other geopolitically or economically important language pairs. Given this state of affairs, there is an urgent need for new cross-lingual methods, language-independent multilingual methods, and methods for establishing lexical links across languages that do not necessarily rely on large-scale parallel corpora. Without new strategies, most of the 7,000+ languages in the world—many with millions of speakers—will remain resource-poor from the standpoint of NLP.

This workshop features submissions from a diverse range of multilingual NLP problems, and invited talks from leading researchers working on multilingual NLP. We would like to thank the members of the program committee for their diligent work — the reviews were all very thorough, and detailed, which helped the authors improve their papers.

Organizers:

Dipanjan Das, Google Inc., USA
Chris Dyer, Google DeepMind, UK
Manaal Faruqui, Carnegie Mellon University, USA
Yulia Tsvetkov, Carnegie Mellon University, USA

Program Committee:

Waleed Ammar, Carnegie Mellon University, USA
Miguel Ballesteros, Pompeu Fabra, Spain
Mohit Bansal, Toyota Technological Institute at Chicago, USA
Phil Blunsom, Google DeepMind / Oxford, UK
Jan Botha, Google Inc., UK
Chris Callison-Burch, University of Pennsylvania, USA
Marine Carpuat, University of Maryland, USA
David Chiang, Notre Dame University, USA
Shay Cohen, University of Edinburgh, UK
Ryan Cotterell, Johns Hopkins University, USA
Rajarshi Das, University of Massachusetts, USA
Mona Diab, George Washington University, USA
Nadir Durrani, Qatar Computing Research Institute, Qatar
Kevin Gimpel, Toyota Technological Institute at Chicago, USA
Jiang Guo, Harbin Institute of Technology, China
David Hall, Semantic Machines, USA
Karl Moritz Hermann, Google DeepMind, UK
Dirk Hovy, University of Copenhagen, Denmark
Jagadeesh Jagarlamudi, Google Inc., USA
David Jurgens, Stanford University, USA
Young-Bum Kim, Microsoft Research, USA
Lingpeng Kong, Carnegie Mellon University, USA
Mirella Lapata, University of Edinburgh, UK
Omer Levy, Bar-Ilan University, Israel
Minh Thang Luong, Stanford University, USA
Ryan McDonald, Google Inc., UK
Gerard de Melo, Tsinghua University, China
Karthik Narasimhan, Massachusetts Institute of Technology, USA
Tahira Naseem, IBM Research, USA
Avneesh Saluja, AirBnb, USA
Anoop Sarkar, Simon Fraser University, Canada
Anders Søgaard, University of Copenhagen, Denmark
Oscar Tackstrom, Google Inc., USA
Jason Utt, University of Stuttgart, Germany

Shuly Wintner, University of Haifa, Israel
Dani Yogatama, Baidu, USA
Daniel Zeman, Charles University in Prague, Czech Republic

Invited Speakers:

Kyunghyun Cho, New York University, USA
Chris Dyer, Google DeepMind, UK
Dan Garrette, University of Washington, USA
Kevin Knight, University of Southern California, USA
Nathan Schneider, Georgetown University, USA
Ivan Titov, University of Amsterdam, Netherlands
David Yarowsky, Johns Hopkins University, USA

Table of Contents

Workshop Program

9:15–9:30	*Opening Remarks* Yulia Tsvetkov
9:30–10:10	*Evaluation by Compression* Invited Talk by Kevin Knight
10:10–10:50	*Multi-way, Multilingual Neural Machine Translation* Invited Talk by Kyunghyun Cho
10:50–11:10	*Coffee Break*
11:10–11:50	*The Case for a Coarse-grained Multilingual Representation of Case and Adposition Semantics* Invited Talk by Nathan Schneider
11:50–12:30	*To be decided* Invited Talk by Chris Dyer
12:30–1:30	*Lunch and Setting Posters*
1:30–1:50	*Comparing Fifty Natural Languages and Twelve Genetic Languages Using Word Embedding Language Divergence (WELD) as a Quantitative Measure of Language Distance* Ehsaneddin Asgari and Mohammad R.K. Mofrad
2:00–3:30	*Posters and Coffee*
3:30–4:10	*Cross-lingual and Unsupervised Learning of Semantic Representations* Invited Talk by Ivan Titov
4:10–4:50	*Unsupervised Modeling of Code-Switching and Orthographic Variation, and its Application to the Study of Digital Humanities* Invited Talk by Dan Garrette
4:50–5:30	*Cross-lingual Learning of Universalized Morphosemantics* Invited Talk by David Yarowsky
5:30–5:45	*Best Paper & Poster Awards*