

EMNLP 2016

**Workshop on Uphill Battles in Language Processing:
Scaling Early Achievements to Robust Methods**

Workshop Proceedings

November 5, 2016
Austin, Texas, USA



©2016 The Association for Computational Linguistics

Order copies of this and other ACL proceedings from:

Curran Associates
57 Morehouse Lane
Red Hook, New York 12571
USA
Tel: +1-845-758-0400
Fax: +1-845-758-2633
curran@proceedings.com

ISBN 978-1-945626-30-2

Introduction

Welcome to the EMNLP 2016 Workshop on Uphill Battles in Language Processing: Scaling Early Achievements to Robust Methods.

Early researchers in Natural Language Processing had lofty goals, including getting computers to understand stories, engage in natural, cooperative dialogues with people, and translate text and speech fluently and accurately from one human language to another. While there were significant early achievements (including systems such as SHRDLU, LUNAR and COOP), the knowledge they were based on and the techniques they employed could not be scaled up for practical use.

While much of what early researchers set out to achieve has been either forgotten or sidelined in favor of what can be done by exploiting large data sets and processing power, its potential value has not gone away: There is much to be gained from recognizing not just what was said, but why; from identifying conclusions naturally drawn from what has been said and what hasn't; and from representing domains in a sufficiently rich way to reduce reliance on only what a text makes explicit. As such, we believe there can be a broad and positive impact of reviving early aspirations in the current context of large data sets and "deep" and probabilistic methods.

The workshop program is split into four panel sessions and a poster session. Each panel leads a discussion on a different area of natural language processing: document understanding, natural language generation, dialogue and speech, and language grounding. Each panel session consists of four short (10 minute) presentations, two by established researchers who carried out early work in the area, and two by more junior researchers who are known for their work on specific problems in the area. Following the presentations, workshop participants are invited to discuss challenges and potential approaches for challenges in that field. In addition, the program includes twelve research abstracts that were selected out of 16 submissions. These abstracts are presented as poster boosters at the workshop, as well as in a poster session.

Our program committee consisted of 25 researchers who provided constructive and thoughtful reviews. This workshop would not have been possible without their hard work. Many thanks to you all. We also thank the U.S. National Science Foundation for financial support. Finally, a huge thank you to all the authors who submitted abstracts to this workshop and made it a big success.

Annie, Michael, Bonnie, Mike and Luke

Organizers:

Annie Louis, University of Essex
Michael Roth, University of Illinois Urbana-Champaign / Saarland University
Bonnie Webber, University of Edinburgh
Michael White, The Ohio State University
Luke Zettlemoyer, University of Washington

Program Committee:

Omri Abend, The Hebrew University of Jerusalem
Timothy Baldwin, University of Melbourne
Nate Chambers, United States Naval Academy
Ann Copestake, University of Cambridge
Vera Demberg, Saarland University
Anette Frank, Heidelberg University
Aurelie Herbelot, University of Trento
Graeme Hirst, University of Toronto
Eduard Hovy, Carnegie Mellon University
Beata Beigman Klebanov, Educational Testing Service
Anna Korhonen, University of Cambridge
Daniel Marcu, Information Sciences Institute, USC
Katja Markert, Heidelberg University
Rada Mihalcea, University of Michigan
Hwee Tou Ng, National University of Singapore
Alexis Palmer, Heidelberg University
Manfred Pinkal, Saarland University
Sameer Pradhan, Boulder Learning
Sabine Schulte im Walde, University of Stuttgart
Jan Snajder, University of Zagreb
Swapna Somasundaran, Educational Testing Service
Manfred Stede, University of Potsdam
Joel Tetreault, Yahoo! Labs
Simone Teufel, University of Cambridge
Lucy Vanderwende, Microsoft Research

Invited Speakers:

James Allen, University of Rochester / IHMC
Joyce Chai, Michigan State University
Yejin Choi, University of Washington
Hal Daumé III, University of Maryland, College Park
Marie-Catherine de Marneffe, Ohio State University
David DeVault, University of Southern California

Andrew Kehler, University of California, San Diego
Ioannis Konstas, University of Washington
Mark Liberman, University of Pennsylvania
Diane Litman, University of Pittsburgh
Chris Manning, Stanford University
Kathleen McKeown, Columbia University
Margaret Mitchell, Microsoft Research
Donia Scott, University of Sussex
Mark Steedman, University of Edinburgh
Amanda Stent, Bloomberg

Table of Contents

<i>An Analysis of Prerequisite Skills for Reading Comprehension</i> Saku Sugawara and Akiko Aizawa	1
<i>Bridging the gap between computable and expressive event representations in Social Media</i> Darina Benikova and Torsten Zesch	6
<i>Statistical Script Learning with Recurrent Neural Networks</i> Karl Pichotta and Raymond Mooney	11
<i>Moving away from semantic overfitting in disambiguation datasets</i> Marten Postma, Filip Ilievski, Piek Vossen and Marieke van Erp	17
<i>Unsupervised Event Coreference for Abstract Words</i> Dheeraj Rajagopal, Eduard Hovy and Teruko Mitamura	22
<i>Towards Broad-coverage Meaning Representation: The Case of Comparison Structures</i> Omid Bakhshandeh and James Allen	27
<i>DialPort: A General Framework for Aggregating Dialog Systems</i> Tiancheng Zhao, Kyusong Lee and Maxine Eskenazi	32
<i>C2D2E2: Using Call Centers to Motivate the Use of Dialog and Diarization in Entity Extraction</i> Ken Church, Weizhong Zhu and Jason Pelecanos	35
<i>Visualizing the Content of a Children’s Story in a Virtual World: Lessons Learned</i> Quynh Ngoc Thi Do, Steven Bethard and Marie-Francine Moens	39
<i>Stylistic Transfer in Natural Language Generation Systems Using Recurrent Neural Networks</i> Jad Kabbara and Jackie Chi Kit Cheung	43
<i>Using Language Groundings for Context-Sensitive Text Prediction</i> Timothy Lewis, Cynthia Matuszek, Amy Hurst and Matthew Taylor	48
<i>Towards a continuous modeling of natural language domains</i> Sebastian Ruder, Parsa Ghaffari and John G. Breslin	53

Workshop Program

Saturday, November 5, 2016

09:00–10:20 Session S1: Text Understanding

09:00–10:20 *Invited talks, followed by discussion*
Hal Daume III, Andrew Kehler, Chris Manning, Marie-Catherine de Marneffe

10:20–10:30 Session S2: Poster Boasters

An Analysis of Prerequisite Skills for Reading Comprehension
Saku Sugawara and Akiko Aizawa

Bridging the gap between computable and expressive event representations in Social Media
Darina Benikova and Torsten Zesch

Statistical Script Learning with Recurrent Neural Networks
Karl Pichotta and Raymond Mooney

Moving away from semantic overfitting in disambiguation datasets
Marten Postma, Filip Ilievski, Piek Vossen and Marieke van Erp

Unsupervised Event Coreference for Abstract Words
Dheeraj Rajagopal, Eduard Hovy and Teruko Mitamura

Towards Broad-coverage Meaning Representation: The Case of Comparison Structures
Omid Bakhshandeh and James Allen

10:30–11:00 Coffee break

Saturday, November 5, 2016 (continued)

11:00–12:20 Session S3: Natural Language Generation

11:00–12:20 *Invited talks, followed by discussion*
Ioannis Konstas, Kathleen McKeown, Margaret Mitchell, Donia Scott

12:20–12:30 Session S4: Poster Boosters

DialPort: A General Framework for Aggregating Dialog Systems
Tiancheng Zhao, Kyusong Lee and Maxine Eskenazi

C2D2E2: Using Call Centers to Motivate the Use of Dialog and Diarization in Entity Extraction
Ken Church, Weizhong Zhu and Jason Pelecanos

Visualizing the Content of a Children’s Story in a Virtual World: Lessons Learned
Quynh Nhoc Thi Do, Steven Bethard and Marie-Francine Moens

Stylistic Transfer in Natural Language Generation Systems Using Recurrent Neural Networks
Jad Kabbara and Jackie Chi Kit Cheung

Using Language Groundings for Context-Sensitive Text Prediction
Timothy Lewis, Cynthia Matuszek, Amy Hurst and Matthew Taylor

Towards a continuous modeling of natural language domains
Sebastian Ruder, Parsa Ghaffari and John G. Breslin

12:30–14:00 Lunch break

Saturday, November 5, 2016 (continued)

14:00–15:20 Session S5: Dialogue and Speech

14:00–15:20 *Invited talks, followed by discussion*
David DeVault, Mark Liberman, Diane Litman, Amanda Stent

15:20–16:00 *Coffee break + poster session*

16:00–16:30 Session S6: Poster session (continued)

16:30–17:50 Session S7: Grounded Language

16:30–17:50 *Invited talks, followed by discussion*
James Allen, Joyce Chai, Yejin Choi, Mark Steedman

