ACL-HLT 2011

BioNLP 2011

Proceedings of the Workshop

23-24 June, 2011
Portland, Oregon, USA

Production and Manufacturing by
Omnipress, Inc.
2600 Anderson Street
Madison, WI 53704 USA

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 978-1-932432-91-6

Introduction

BioNLP 2011 received 31 submissions that with very few exceptions maintain the tradition of excellence established by the BioNLP authors over the past 10 years. Eleven submissions were accepted as full papers and 14 as poster presentations.

The themes in this year's papers and posters cover complex NLP problems in biological and clinical language processing.

Acknowledgments

We are profoundly grateful to the authors who chose BioNLP as venue for presenting their innovative research.

The authors' willingness to share their work through BioNLP consistently makes the workshop noteworthy and stimulating.

We are equally indebted to the program committee members (listed elsewhere in this volume) who produced at least two thorough reviews per paper on a tight review schedule and with an admirable level of insight.

We are particularly grateful to reviewers who reviewed late submissions from Japan in even shorter period of time. And we admire our Japanese colleagues who continued focusing on their research in the middle of an unfathomable natural disaster.

Organizers:

Kevin Bretonnel Cohen, University of Colorado School of Medicine
Dina Demner-Fushman, US National Library of Medicine
Sophia Ananiadou, University of Manchester and National Centre for Text Mining, UK
John Pestian, Computational Medical Center, University of Cincinnati,
Cincinnati Children's Hospital Medical Center
Jun'ichi Tsujii, University of Tokyo
and University of Manchester and National Centre for Text Mining, UK
Bonnie Webber,University of Edinburgh, UK

Program Committee:

Alan Aronson
Emilia Apostolova
Olivier Bodenreider
Wendy Chapman
Aaron Cohen
Nigel Collier
Noemie Elhadad
Marcelo Fiszman
Filip Ginter
Su Jian
Halil Kilicoglu
Jin-Dong Kim
Marc Light
Zhiyong Lu
Aurelie Neveol
Sampo Pyysalo
Thomas Rindflesch
Andrey Rzhetsky
Daniel Rubin
Hagit Shatkay
Matthew Simpson
Larry Smith
Yuka Tateisi
Yoshimasa Tsuruoka
Karin Verspoor
W. John Wilbur
Limsoon Wong
Pierre Zweigenbaum

Table of Contents

Conference Program

Thursday June 23, 2011

9:00–9:10 Opening Remarks

Session 1: Text Mining

9:10–9:30 *Not all links are equal: Exploiting Dependency Types for the Extraction of Protein-Protein Interactions from Text*
Philippe Thomas, Stefan Pietschmann, Illés Solt, Domonkos Tikk and Ulf Leser

9:30–9:50 *Unsupervised Entailment Detection between Dependency Graph Fragments*
Marek Rei and Ted Briscoe

9:50–10:10 *Learning Phenotype Mapping for Integrating Large Genetic Data*
Chun-Nan Hsu, Cheng-Ju Kuo, Congxing Cai, Sarah Pendergrass, Marylyn Ritchie and Jose Luis Ambite

10:10–10:30 *EVEX: A PubMed-Scale Resource for Homology-Based Generalization of Text Mining Predictions*
Sofie Van Landeghem, Filip Ginter, Yves Van de Peer and Tapio Salakoski

10:30–11:00 Morning coffee break

11:00–11:20 *Fast and simple semantic class assignment for biomedical text*
K. Bretonnel Cohen, Thomas Christiansen, William Baumgartner Jr., Karin Verspoor and Lawrence Hunter

11:20–12:30 Invited Talk

12:30–14:00 Lunch break

Thursday June 23, 2011 (continued)

Session 2: Information extraction and corpora

14:00–14:20 *The Role of Information Extraction in the Design of a Document Triage Application for Biocuration*
Sandeep Pokkunuri, Cartic Ramakrishnan, Ellen Riloff, Eduard Hovy and Gully Burns

14:20–14:40 *Medical Entity Recognition: A Comparaison of Semantic and Statistical Methods*
Asma Ben Abacha and Pierre Zweigenbaum

14:40–15:00 *Automatic Acquisition of Huge Training Data for Bio-Medical Named Entity Recognition*
Yu Usami, Han-Cheol Cho, Naoaki Okazaki and Jun'ichi Tsujii

15:00–15:20 *Building frame-based corpus on the basis of ontological domain knowledge*
He Tan, Rajaram Kaliyaperumal and Nirupama Benis

15:30–16:00 Afternoon coffee break

16:00–16:20 *Building a Coreference-Annotated Corpus from the Domain of Biochemistry*
Riza Theresa Batista-Navarro and Sophia Ananiadou

16:20–16:40 *Towards Morphologically Annotated Corpus of Hospital Discharge Reports in Polish*
Malgorzata Marciniak and Agnieszka Mykowiecka

16:40–17:00 Poster boaster and wrap-up

Poster Session

17:00–17:30 *In Search of Protein Locations*
Catherine Blake and Wu Zheng

17:00–17:30 *Automatic extraction of data deposition statements: where do the research results go?*
Aurelie Neveol, W. John Wilbur and Zhiyong Lu

17:00–17:30 *From Pathways to Biomolecular Events: Opportunities and Challenges*
Tomoko Ohta, Sampo Pyysalo and Jun'ichi Tsujii

17:00–17:30 *Towards Exhaustive Event Extraction for Protein Modifications*
Sampo Pyysalo, Tomoko Ohta, Makoto Miwa and Jun'ichi Tsujii

Thursday June 23, 2011 (continued)

17:00–17:30 *A Study on Dependency Tree Kernels for Automatic Extraction of Protein-Protein Interaction*
Faisal Md. Chowdhury, Alberto Lavelli and Alessandro Moschitti

17:00–17:30 *Hypothesis and Evidence Extraction from Full-Text Scientific Journal Articles*
Elizabeth White, K. Bretonnel Cohen and Larry Hunter

17:00–17:30 *SimSem: Fast Approximate String Matching in Relation to Semantic Category Disambiguation*
Pontus Stenetorp, Sampo Pyysalo and Jun'ichi Tsujii

17:00–17:30 *Building Timelines from Narrative Clinical Records: Initial Results Based-on Deep Natural Language Understanding*
Hyuckchul Jung, James Allen, Nate Blaylock, William de Beaumont, Lucian Galescu and Mary Swift

17:00–17:30 *Text Mining Techniques for Leveraging Positively Labeled Data*
Lana Yeganova, Donald C. Comeau, Won Kim and W. John Wilbur

17:00–17:30 *Parsing Natural Language Queries for Life Science Knowledge*
Tadayoshi Hara, Yuka Tateisi, Jin-Dong Kim and Yusuke Miyao

17:00–17:30 *Unlocking Medical Ontologies for Non-Ontology Experts*
Shao Fen Liang, Donia Scott, Robert Stevens and Alan Rector

17:00–17:30 *Self-training and co-training in biomedical word sense disambiguation*
Antonio Jimeno Yepes and Alan Aronson

17:00–17:30 *Medstract - The Next Generation*
Marc Verhagen and James Pustejovsky

17:00–17:30 *ThaiHerbMiner: A Thai Herbal Medicine Mining and Visualizing Tool*
Choochart Haruechaiyasak, Jaruwat Pailai, Wasna Viratyosin and Rachada Kongkachandra