NAACL-HLT 2012

BioNLP 2012

Workshop on Biomedical Natural Language Processing

Proceedings of the Workshop

June 8, 2012
Montréal, Canada

Production and Manufacturing by
Omnipress, Inc.
2600 Anderson Street
Madison, WI 53707
USA

©2012 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

ISBN13: 978-1-937284-20-6
ISBN10: 1-937284-20-4

Introduction

BioNLP 2012 received 31 submissions exceeding even the traditionally high quality of the preceding eleven years of BioNLP. Due to uniformly positive reviews, eleven submissions were accepted as full papers and 19 as poster presentations.

The themes in this year's papers and posters continue reflecting researchers' growing interest in clinical text processing, while maintaining a steady mature work in biological language processing. This year presents a wide range of innovative methods applied to interesting problems in both domains.

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 not only noteworthy and stimulating, but also one of the largest, and some years the largest workshop, at ACL/NAACL.

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

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 Microsoft Research Asia
Bonnie Webber,University of Edinburgh, UK

Program Committee:

Sophia Ananiadou
Galia Angelova
Emilia Apostolova
Alan Aronson
Olivier Bodenreider
Wendy Chapman
Kevin Cohen
Nigel Collier
Dina Demner-Fushman
Noemie Elhadad
Marcelo Fiszman
Filip Ginter
Su Jian
Jin-Dong Kim
Zhiyong Lu
Aurelie Neveol
Jon Patrick
John Pestian
Sampo Pyysalo
Bastien Rance
Fabio Rinaldi
Thomas Rindflesch
Brian Roark
Andrey Rzhetsky
Daniel Rubin
Guergana Savova
Hagit Shatkay
Matthew Simpson
Pontus Stenetorp
Yuka Tateisi

Jun’ichi Tsujii
Yoshimasa Tsuruoka
Ozlem Uzuner
Karin Verspoor
Bonnie Webber
Peter White
W. John Wilbur
Limsoon Wong
Antonio Yepes
Guodong Zhou
Pierre Zweigenbaum

Invited Speaker:

Wendy W. Chapman, University of California San Diego
Challenges and Opportunities in Clinical Text Annotation

Table of Contents

Graph-based alignment of narratives for automated neurological assessment
Emily Prud'hommeaux and Brian Roark . . . 1

Bootstrapping Biomedical Ontologies for Scientific Text using NELL
Dana Movshovitz-Attias and William W. Cohen . . . 11

Semantic distance and terminology structuring methods for the detection of semantically close terms
Marie Dupuch, Laëtitia Dupuch, Thierry Hamon and Natalia Grabar . . . 20

Temporal Classification of Medical Events
Preethi Raghavan, Eric Fosler-Lussier and Albert Lai . . . 29

Analyzing Patient Records to Establish If and When a Patient Suffered from a Medical Condition
James Cogley, Nicola Stokes, Joe Carthy and John Dunnion . . . 38

Alignment-HMM-based Extraction of Abbreviations from Biomedical Text
Dana Movshovitz-Attias and William W. Cohen . . . 47

Medical diagnosis lost in translation – Analysis of uncertainty and negation expressions in English and Swedish clinical texts
Danielle L. Mowery, Sumithra Velupillai and Wendy W. Chapman . . . 56

A Hybrid Stepwise Approach for De-identifying Person Names in Clinical Documents
Oscar Ferrandez, Brett South, Shuying Shen and Stephane Meystre . . . 65

Active Learning for Coreference Resolution
Timothy Miller, Dmitriy Dligach and Guergana Savova . . . 73

PubMed-Scale Event Extraction for Post-Translational Modifications, Epigenetics and Protein Structural Relations
Jari Björne, Sofie Van Landeghem, Sampo Pyysalo, Tomoko Ohta, Filip Ginter, Yves Van de Peer, Sophia Ananiadou and Tapio Salakoski . . . 82

An improved corpus of disease mentions in PubMed citations
Rezarta Islamaj Dogan and Zhiyong Lu . . . 91

New Resources and Perspectives for Biomedical Event Extraction
Sampo Pyysalo, Pontus Stenetorp, Tomoko Ohta, Jin-Dong Kim and Sophia Ananiadou . . . 100

Combining Compositionality and Pagerank for the Identification of Semantic Relations between Biomedical Words
Thierry Hamon, Christopher Engström, Mounira Manser, Zina Badji, Natalia Grabar and Sergei Silvestrov . . . 109

Domain Adaptation of Coreference Resolution for Radiology Reports
Emilia Apostolova, Noriko Tomuro, Pattanasak Mongkolwat and Dina Demner-Fushman . . . 118

What can NLP tell us about BioNLP?
Attapol Thamrongrattanarit, Michael Shafir, Michael Crivaro, Bensiin Borukhov and Marie Meteer 122

A Prototype Tool Set to Support Machine-Assisted Annotation
Brett South, Shuying Shen, Jianwei Leng, Tyler Forbush, Scott DuVall and Wendy Chapman 130

MedLingMap: A growing resource mapping the Bio-Medical NLP field
Marie Meteer, Borukhov Bensiin, Mike Crivaro, Michael Shafir and Attapol Thamrongrattanarit 140

Exploring Label Dependency in Active Learning for Phenotype Mapping
Shefali Sharma, Leslie Lange, Jose Luis Ambite, Yigal Arens and Chun-Nan Hsu 146

Evaluating Joint Modeling of Yeast Biology Literature and Protein-Protein Interaction Networks
Ramnath Balasubramanyan, Kathryn Rivard, William W. Cohen, Jelena Jakovljevic and John L. Woolford 155

RankPref: Ranking Sentences Describing Relations between Biomedical Entities with an Application
Catalina Oana Tudor and K Vijay-Shanker 163

Finding small molecule and protein pairs in scientific literature using a bootstrapping method
Ying Yan, Jee-Hyub Kim, Samuel Croset and Dietrich Rebholz-Schuhmann 172

Grading the Quality of Medical Evidence
Binod Gyawali, Thamar Solorio and Yassine Benajiba 176

Classifying Gene Sentences in Biomedical Literature by Combining High-Precision Gene Identifiers
Sun Kim, Won Kim, Don Comeau and W. John Wilbur 185

Effect of small sample size on text categorization with support vector machines
Pawel Matykiewicz and John Pestian 193

PubAnnotation - a persistent and sharable corpus and annotation repository
Jin-Dong Kim and Yue Wang 202

Using Natural Language Processing to Extract Drug-Drug Interaction Information from Package Inserts
Richard Boyce, Gregory Gardner and Henk Harkema 206

Automatic Approaches for Gene-Drug Interaction Extraction from Biomedical Text: Corpus and Comparative Evaluation
Nate Sutton, Laura Wojtulewicz, Neel Mehta and Graciela Gonzalez 214

A Preliminary Work on Symptom Name Recognition from Free-Text Clinical Records of Traditional Chinese Medicine using Conditional Random Fields and Reasonable Features
Yaqiang Wang, Yiguang Liu, Zhonghua Yu, Li Chen and Yongguang Jiang 223

Scaling up WSD with Automatically Generated Examples
Weiwei Cheng, Judita Preiss and Mark Stevenson 231

Boosting the protein name recognition performance by bootstrapping on selected text
Yue Wang and Jin-Dong Kim 240

Conference Program

Friday, June 8, 2012

8:40–8:50 Opening Remarks

Session 1: Alignment, similarity, classification

8:50–9:10 *Graph-based alignment of narratives for automated neurological assessment*
Emily Prud'hommeaux and Brian Roark

9:10–9:30 *Bootstrapping Biomedical Ontologies for Scientific Text using NELL*
Dana Movshovitz-Attias and William W. Cohen

9:30–9:50 *Semantic distance and terminology structuring methods for the detection of semantically close terms*
Marie Dupuch, Laëtitia Dupuch, Thierry Hamon and Natalia Grabar

9:50–10:10 *Temporal Classification of Medical Events*
Preethi Raghavan, Eric Fosler-Lussier and Albert Lai

10:10–10:30 *Analyzing Patient Records to Establish If and When a Patient Suffered from a Medical Condition*
James Cogley, Nicola Stokes, Joe Carthy and John Dunnion

10:30–11:00 Morning coffee break

11:00–12:10 Invited Talk by Wendy Chapman

12:10–12:30 *Alignment-HMM-based Extraction of Abbreviations from Biomedical Text*
Dana Movshovitz-Attias and William W. Cohen

12:30–14:00 Lunch break

14:00–14:20 *Medical diagnosis lost in translation – Analysis of uncertainty and negation expressions in English and Swedish clinical texts*
Danielle L. Mowery, Sumithra Velupillai and Wendy W. Chapman

14:20–14:40 *A Hybrid Stepwise Approach for De-identifying Person Names in Clinical Documents*
Oscar Ferrandez, Brett South, Shuying Shen and Stephane Meystre

Friday, June 8, 2012 (continued)

14:40–15:00 *Active Learning for Coreference Resolution*
Timothy Miller, Dmitriy Dligach and Guergana Savova

15:00–15:20 *PubMed-Scale Event Extraction for Post-Translational Modifications, Epigenetics and Protein Structural Relations*
Jari Björne, Sofie Van Landeghem, Sampo Pyysalo, Tomoko Ohta, Filip Ginter, Yves Van de Peer, Sophia Ananiadou and Tapio Salakoski

15:20–15:40 *An improved corpus of disease mentions in PubMed citations*
Rezarta Islamaj Dogan and Zhiyong Lu

15:30–16:00 Afternoon coffee break

Poster Session (16:00–18:00)

New Resources and Perspectives for Biomedical Event Extraction
Sampo Pyysalo, Pontus Stenetorp, Tomoko Ohta, Jin-Dong Kim and Sophia Ananiadou

Combining Compositionality and Pagerank for the Identification of Semantic Relations between Biomedical Words
Thierry Hamon, Christopher Engström, Mounira Manser, Zina Badji, Natalia Grabar and Sergei Silvestrov

Domain Adaptation of Coreference Resolution for Radiology Reports
Emilia Apostolova, Noriko Tomuro, Pattanasak Mongkolwat and Dina Demner-Fushman

What can NLP tell us about BioNLP?
Attapol Thamrongrattanarit, Michael Shafir, Michael Crivaro, Bensiin Borukhov and Marie Meteer

A Prototype Tool Set to Support Machine-Assisted Annotation
Brett South, Shuying Shen, Jianwei Leng, Tyler Forbush, Scott DuVall and Wendy Chapman

MedLingMap: A growing resource mapping the Bio-Medical NLP field
Marie Meteer, Borukhov Bensiin, Mike Crivaro, Michael Shafir and Attapol Thamrongrattanarit

Exploring Label Dependency in Active Learning for Phenotype Mapping
Shefali Sharma, Leslie Lange, Jose Luis Ambite, Yigal Arens and Chun-Nan Hsu

Evaluating Joint Modeling of Yeast Biology Literature and Protein-Protein Interaction Networks
Ramnath Balasubramanyan, Kathryn Rivard, William W. Cohen, Jelena Jakovljevic and John L. Woolford

Friday, June 8, 2012 (continued)

RankPref: Ranking Sentences Describing Relations between Biomedical Entities with an Application
Catalina Oana Tudor and K Vijay-Shanker

Finding small molecule and protein pairs in scientific literature using a bootstrapping method
Ying Yan, Jee-Hyub Kim, Samuel Croset and Dietrich Rebholz-Schuhmann

Grading the Quality of Medical Evidence
Binod Gyawali, Thamar Solorio and Yassine Benajiba

Classifying Gene Sentences in Biomedical Literature by Combining High-Precision Gene Identifiers
Sun Kim, Won Kim, Don Comeau and W. John Wilbur

Effect of small sample size on text categorization with support vector machines
Pawel Matykiewicz and John Pestian

PubAnnotation - a persistent and sharable corpus and annotation repository
Jin-Dong Kim and Yue Wang

Using Natural Language Processing to Extract Drug-Drug Interaction Information from Package Inserts
Richard Boyce, Gregory Gardner and Henk Harkema

Automatic Approaches for Gene-Drug Interaction Extraction from Biomedical Text: Corpus and Comparative Evaluation
Nate Sutton, Laura Wojtulewicz, Neel Mehta and Graciela Gonzalez

A Preliminary Work on Symptom Name Recognition from Free-Text Clinical Records of Traditional Chinese Medicine using Conditional Random Fields and Reasonable Features
Yaqiang Wang, Yiguang Liu, Zhonghua Yu, Li Chen and Yongguang Jiang

Scaling up WSD with Automatically Generated Examples
Weiwei Cheng, Judita Preiss and Mark Stevenson

Boosting the protein name recognition performance by bootstrapping on selected text
Yue Wang and Jin-Dong Kim