NAACL HLT 2009

BioNLP 2009

Proceedings of the Workshop

June 4-5, 2009
Boulder, Colorado

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

BioNLP Sponsor:

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 978-1-932432-30-5

BioNLP 2009

K. Bretonnel Cohen, Dina Demner-Fushman, Sophia Ananiadou,
John Pestian, Jun'ichi Tsujii, and Bonnie Webber

1 Introduction

Yearly BioNLP workshops have been held in conjunction with Association for Computational Linguistics and North American Association for Computational Linguistics conferences since 2002. Whereas other venues, such as NLP sessions at biomedical informatics and computational biology meetings, provide excellent opportunities for presenting applications of NLP in the biomedical domain, the ACL BioNLP workshop has become the venue that is most characterized by representation of work in a wide variety of areas of NLP. The BioNLP workshop has consistently been a venue for presenting work that is innovative, novel, and challenging from an NLP perspective. In addition to providing a venue for fundamental BioNLP research, this workshop exposes BioNLP researchers to the latest achievements in other NLP areas and facilitates dissemination of knowledge acquired in the BioNLP domain to the wider NLP community.

Compared to previous years, BioNLP 2009 was novel in two ways. The first is that it is the first workshop since formation of the SIGBIOMED Association for Computational Linguistics Special Interest Group. The second is that for the first time, there was a shared task associated with the workshop. This shared task is documented in a separate proceedings volume.

2 Submissions, acceptance rate, and themes

The workshop received 29 submissions, of which twelve were accepted as full papers and an additional twelve were accepted as posters. A number of themes were evident in this year's papers and posters. Lexical semantics was especially well-represented this year, with papers on ontology selection [10], lexicon construction [12], and synonymy [3]. Information extraction was also well-represented, with papers in this area tackling both the genomic [2, 8], and the clinical [1] domain. This included work that is novel in the biomedical domain in terms of dealing with speech and with the dental domain [1]. This year also saw continued work on contextual issues in biomedical text mining [6, 7]. Finally, the program was rounded out with work on a new formulation of the named entity recognition problem [11], the hot topic of species identification [5], and word sense disambiguation [9] and summarization [4].

Acknowledgments

The greatest debt owed by the organizers of a workshop like this is to the authors who graciously chose BioNLP 2009 as the venue in which to share the fruits of the countless hours of research that went into the work submitted for consideration. The next-biggest debt is, without question, to the many program committee

members (listed elsewhere in this volume); they produced three reviews per paper on a tight review schedule and with an admirable level of insight. Finally, we acknowledge the gracious sponsorship of the Computational Medicine Center and Division of Biomedical Informatics, Cincinnati Children's Hospital Medical Center.

References

[1] Lee Christensen, Henk Harkema, Peter Haug, Jeannie Irwin, and Wendy Chapman. ONYX: A system for the semantic analysis of clinical text. In *BioNLP*, 2009.

[2] Udo Hahn, Katrin Tomanek, Ekaterina Buyko, Jung-jae Kim, and Dietrich Rebholz-Schuhmann. How feasible and robust is the automatic extraction of gene regulation events? A cross-method evaluation under lab and real-life conditions. In *BioNLP*, 2009.

[3] Thierry Hamon and Natalia Grabar. Exploring graph structure for detection of reliability zones within synonym resources: Experiment with the Gene Ontology. In *BioNLP*, 2009.

[4] Feng Jin, Minlie Huang, Zhiyong Lu, and Xiaoyan Zhu. Towards automatic generation of gene summary. In *BioNLP*, 2009.

[5] Thomas Kappeler, Kaarel Kaljurand, and Fabio Rinaldi. TX Task: Automatic detection of focus organisms in biomedical publications. In *BioNLP*, 2009.

[6] Roser Morante and Walter Daelemans. Learning the scope of hedge cues in biomedical texts. In *BioNLP*, 2009.

[7] Danielle Mowery, Henk Harkema, John Dowling, Jonathan Lustgarten, and Wendy Chapman. Distinguishing historical from current problems in clinical reports – Which textual features help? In *BioNLP*, 2009.

[8] Sampo Pyysalo, Tomoko Ohta, Jin-Dong Kim, and Jun'ichi Tsujii. Static relations: a piece in the biomedical information extraction puzzle. In *BioNLP*, 2009.

[9] Mark Stevenson, Yikun Guo, Abdulaziz Alamri, and Robert Gaizauskas. Disambiguation of biomedical abbreviations. In *BioNLP*, 2009.

[10] He Tan and Patrick Lambrix. Selecting an ontology for biomedical text mining. In *BioNLP*, 2009.

[11] Wern Wong, David Martinez, and Lawrence Cavedon. Extraction of named entities from tables in gene mutation literature. In *BioNLP*, 2009.

[12] Rong Xu, Alexander A. Morgan, Amar Das, and Alan Garber. Investigation of unsupervised pattern learning techniques for bootstrap construction of a medical treatment lexicon. In *BioNLP*, 2009.

Organizers:

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

Program Committee:

Alan Aronson, LHNCBC, US National Library of Medicine
Catherine Blake, University of North Carolina
Olivier Bodenreider, LHNCBC, US National Library of Medicine
Bob Carpenter, Alias-i
Wendy Chapman, University of Pittsburgh
Aaron Cohen, Oregon Health and Science University
Nigel Collier, National Institute of Informatics, Tokyo
Noemie Elhadad, Columbia University
Marcelo Fiszman, US National Library of Medicine
Carol Friedman, Columbia College of Physicians and Surgeons
Jin-Dong Kim, University of Tokyo
Marc Light, Thomson
Zhiyong Lu, NCBI, US National Library of Medicine
Aurelie Neveol, LHNCBC, US National Library of Medicine
Serguei Pakhomov, University of Minnesota
Thomas Rindflesch, LHNCBC, US National Library of Medicine
Daniel Rubin, Stanford University
Hagit Shatkay, Queen's University, Canada
Larry Smith, NCBI, US National Library of Medicine
Yuka Tateisi, University of Tokyo
Yoshimasa Tsuruoka, University of Manchester
Alfonso Valencia, Centro Nacional de Biotecnologia
Karin Verspoor, Center for Computational Pharmacology, University of Colorado School of Medicine
Peter White, Children's Hospital of Philadelphia
W. John Wilbur, NCBI, US National Library of Medicine
Limsoon Wong, National University of Singapore
Hong Yu, University of Wisconsin
Pierre Zweigenbaum, LIMSI

Table of Contents

Conference Program

Thursday, June 4, 2009

9:00–9:10 Opening Remarks

Session 1: Paper presentations

9:10–9:35 *Static Relations: a Piece in the Biomedical Information Extraction Puzzle*
Sampo Pyysalo, Tomoko Ohta, Jin-Dong Kim and Jun'ichi Tsujii

9:35–10:00 *Distinguishing Historical from Current Problems in Clinical Reports – Which Textual Features Help?*
Danielle Mowery, Henk Harkema, John Dowling, Jonathan Lustgarten and Wendy Chapman

10:00–10:25 *ONYX: A System for the Semantic Analysis of Clinical Text*
Lee Christensen, Henk Harkema, Peter Haug, Jeannie Irwin and Wendy Chapman

10:30–11:00 morning break

11:00–11:25 *Learning the Scope of Hedge Cues in Biomedical Texts*
Roser Morante and Walter Daelemans

11:25–11:50 *How Feasible and Robust is the Automatic Extraction of Gene Regulation Events? A Cross-Method Evaluation under Lab and Real-Life Conditions*
Udo Hahn, Katrin Tomanek, Ekaterina Buyko, Jung-jae Kim and Dietrich Rebholz-Schuhmann

11:50–12:15 *Extraction of Named Entities from Tables in Gene Mutation Literature*
Wern Wong, David Martinez and Lawrence Cavedon

12:15–10:40 *Selecting an Ontology for Biomedical Text Mining*
He Tan and Patrick Lambrix

12:40–2:00 lunch break

2:00–2:30 Invited Talk

2:35–3:00 *Investigation of Unsupervised Pattern Learning Techniques for Bootstrap Construction of a Medical Treatment Lexicon*
Rong Xu, Alexander A. Morgan, Amar Das and Alan Garber

Thursday, June 4, 2009 (continued)

3:00–3:25 *Disambiguation of Biomedical Abbreviations*
Mark Stevenson, Yikun Guo, Abdulaziz Alamri and Robert Gaizauskas

3:30–4:00 afternoon break

4:00–4:25 *TX Task: Automatic Detection of Focus Organisms in Biomedical Publications*
Thomas Kappeler, Kaarel Kaljurand and Fabio Rinaldi

4:25–4:50 *Exploring Graph Structure for Detection of Reliability Zones within Synonym Resources: Experiment with the Gene Ontology*
Thierry Hamon and Natalia Grabar

4:50–5:15 *Towards Automatic Generation of Gene Summary*
Feng Jin, Minlie Huang, Zhiyong Lu and Xiaoyan Zhu

Session 2: Poster presentations

5:00–6:00 *Incorporating GENETAG-style annotation to GENIA corpus*
Tomoko Ohta, Jin-Dong Kim, Sampo Pyysalo, Yue Wang and Jun'ichi Tsujii

5:00–6:00 *User-Driven Development of Text Mining Resources for Cancer Risk Assessment*
Lin Sun, Anna Korhonen, Ilona Silins and Ulla Stenius

5:00–6:00 *Transforming Controlled Natural Language Biomedical Queries into Answer Set Programs*
Esra Erdem and Reyyan Yeniterzi

5:00–6:00 *Incorporating Syntactic Dependency Information towards Improved Coding of Lengthy Medical Concepts in Clinical Reports*
Vijayaraghavan Bashyam and Ricky K Taira

5:00–6:00 *Identifying Interaction Sentences from Biological Literature Using Automatically Extracted Patterns*
Haibin Liu, Christian Blouin and Vlado Keselj

5:00–6:00 *Using Hedges to Enhance a Disease Outbreak Report Text Mining System*
Mike Conway, Son Doan and Nigel Collier

5:00–6:00 *Exploring Two Biomedical Text Genres for Disease Recognition*
Aurelie Neveol, Won Kim, W. John Wilbur and Zhiyong Lu

Thursday, June 4, 2009 (continued)

5:00–6:00 *Towards Retrieving Relevant Information for Answering Clinical Comparison Questions*
Annette Leonhard

5:00–6:00 *Bridging the Gap between Domain-Oriented and Linguistically-Oriented Semantics*
Sumire Uematsu, Jin-Dong Kim and Jun'ichi Tsujii

5:00–6:00 *Evaluation of the Clinical Question Answering Presentation*
Yong-Gang Cao, John Ely, Lamont Antieau and Hong Yu

5:00–6:00 *Clustering Semantic Spaces of Suicide Notes and Newsgroups Articles.*
Pawel Matykiewicz, Wlodzislaw Duch and John Pestian

5:00–6:00 *TEXT2TABLE: Medical Text Summarization System Based on Named Entity Recognition and Modality Identification*
Eiji Aramaki, Yasuhide Miura, Masatsugu Tonoike, Tomoko Ohkuma, Hiroshi Mashuichi and Kazuhiko Ohe

5:00–6:00 *Semantic Annotation of Papers: Interface & Enrichment Tool (SAPIENT)*
Maria Liakata, Claire Q and Larisa N. Soldatova