11th Workshop on Biomedical Natural Language Processing 2012

(BioNLP 2012)

Held at the 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT 2012)

Montreal, Quebec, Canada
3-8 June 2012

ISBN: 978-1-62276-552-2

Printed from e-media with permission by:

Curran Associates, Inc.
57 Morehouse Lane
Red Hook, NY 12571

Some format issues inherent in the e-media version may also appear in this print version.

Copyright© (2012) by the Association for Computational Linguistics
All rights reserved.

Printed by Curran Associates, Inc. (2013)

For permission requests, please contact the Association for Computational Linguistics
at the address below.

Association for Computational Linguistics
209 N. Eighth Street
Stroudsburg, Pennsylvania 18360

Phone: 1-570-476-8006
Fax: 1-570-476-0860

acl@aclweb.org

Additional copies of this publication are available from:

Curran Associates, Inc.
57 Morehouse Lane
Red Hook, NY 12571 USA
Phone: 845-758-0400
Fax: 845-758-2634
Email: curran@proceedings.com
Web: www.proceedings.com

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