ACL 2013

BioNLP 2013

2013 Workshop on Biomedical Natural Language Processing

Proceedings of the Workshop

August 8, 2013
Sofia, Bulgaria

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

Sponsored by the Computational Medicine Center and Division of Biomedical Informatics, Cincinnati Children's Hospital Medical Center

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 978-1-937284-54-1

Introduction

BioNLP 2013 has accepted 11 outstanding full papers and five posters. The themes in this year's papers and posters are divided equally between clinical and biomedical text processing. In addition to the customary research in practical and theoretical issues, such as domain adaptation, question answering, temporal relations extraction, and evaluation of text mining methods, this year, we see a growing body of research in languages other than English. The issues with clinical text processing in resource-poor languages are also discussed in the keynote presentation.

Keynote: Processing clinical narratives in less-resourced languages: the challenge to start from scratch

Galia Angelova, Ph.D. Linguistic Modeling Department, Institute of Information and Communication Technologies, Bulgarian Academy of Sciences

Dr. Angelova presents automatic analysis of free texts in Bulgarian hospital discharge letters of patients with endocrine and metabolic diseases. Processing Bulgarian clinical texts is challenging due to some specific reasons: the notes contain about 37% Latin terms that might occur in Latin alphabet characters as well as transliterated to Cyrillic alphabet (34% of all tokens); the lack of important medical nomenclatures in Bulgarian: for example, the ATC classification is supported in Latin only and requires manual augmentation with Bulgarian drug names in Cyrillic alphabet; no electronic resource with medical terminology is available so the collection of terms and important phrases involves analysis of documents, such as manuals for coding to ICD-10 terms, or collection of collocations directly from the corpus of discharge letters, among others. Currently available resources and methods include automatic recognition of ICD-10 diagnoses; drugs, especially those taken during hospitalization; patient status; values of laboratory tests; and the temporal structure of diabetic case histories. Dr. Angelova discusses scenarios for application of the extraction components in practical settings when cleaning and validation of patient data is required.

Dr. Claire Nedellec presents an overview of the BioNLP Shared Task 2013.

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 among the increasing numbers of available venues. We are equally indebted to the program committee members (listed elsewhere in this volume) who produced at least three thorough 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.

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, Microsoft Research Asia
and National Centre for Text Mining, UK

Program Committee:

Emilia Apostolova, DePaul University, USA
Eiji Aramaki, University of Tokyo, Japan
Alan Aronson, US National Library of Medicine
Sabine Bergler, Concordia University, Canada
Olivier Bodenreider, US National Library of Medicine
Kevin Cohen, University of Colorado, USA
Nigel Collier, National Institute of Informatics, Japan
Dina Demner-Fushman, US National Library of Medicine
Noemie Elhadad, Columbia University, USA
Marcelo Fiszman, US National Library of Medicine
Filip Ginter, University of Turku, Finland
Graciela Gonzalez, Arizona State University, USA
Antonio Jimeno Yepes, NICTA, Australia
Halil Kilicoglu, US National Library of Medicine
Jin-Dong Kim, University of Tokyo, Japan
Robert Leaman, US National Library of Medicine
Ulf Leser, Humboldt University of Berlin, Germany
Zhiyong Lu, US National Library of Medicine
Makoto Miwa, National Centre for Text Mining, UK
Naoaki Okazaki, Tohoku University, Japan
Jong Park, KAIST, South Korea
Rashmi Prasad, University of Wisconsin-Milwaukee, USA
Sampo Pyysalo, National Centre for Text Mining, UK
Bastien Rance, Georges Pompidou European Hospital, France
Andrey Rzhetsky, University of Chicago, USA
Matthew Simpson, US National Library of Medicine
Pontus Stenetorp, University of Tokyo, Japan
Yoshimasa Tsuruoka, University of Tokyo, Japan
Karin Verspoor, NICTA, Australia
W. John Wilbur, US National Library of Medicine
Pierre Zweigenbaum, LIMSI, France

Invited Speakers:

Galia Angelova, Bulgarian Academy of Sciences
Processing clinical narratives in less-resourced languages: the challenge to start from scratch
Claire Nedellec, INRA
Overview of the BioNLP Shared Task 2013

Table of Contents

Conference Program

Thursday, August 8, 2013

8:40–8:50 Opening Remarks

Session 1: Clinical text processing

8:50–9:10 *Earlier Identification of Epilepsy Surgery Candidates Using Natural Language Processing*
Pawel Matykiewicz, Kevin Cohen, Katherine D. Holland, Tracy A. Glauser, Shannon M. Standridge, Karen M. Verspoor and John Pestian

9:10–9:30 *Identification of Patients with Acute Lung Injury from Free-Text Chest X-Ray Reports*
Meliha Yetisgen-Yildiz, Cosmin Bejan and Mark Wurfel

9:30–9:50 *Discovering Temporal Narrative Containers in Clinical Text*
Timothy Miller, Steven Bethard, Dmitriy Dligach, Sameer Pradhan, Chen Lin and Guergana Savova

9:50–10:10 *Identifying Pathological Findings in German Radiology Reports Using a Syntacto-semantic Parsing Approach*
Claudia Bretschneider, Sonja Zillner and Matthias Hammon

10:10–10:30 *Corpus-Driven Terminology Development: Populating Swedish SNOMED CT with Synonyms Extracted from Electronic Health Records*
Aron Henriksson, Maria Skeppstedt, Maria Kvist, Martin Duneld and Mike Conway

10:30–11:00 Morning coffee break

11:00–12:00 Invited Talk by Galia Angelova

12:00–12:30 BioNLP Shared Task overview by Claire Nedellec

12:30–14:00 Lunch break

Thursday, August 8, 2013 (continued)

Session 2: Biomedical language processing

14:00–14:20 *Unsupervised Linguistically-Driven Reliable Dependency Parses Detection and Self-Training for Adaptation to the Biomedical Domain*
Felice Dell'Orletta, Giulia Venturi and Simonetta Montemagni

14:20–14:40 *Interpreting Consumer Health Questions: The Role of Anaphora and Ellipsis*
Halil Kilicoglu, Marcelo Fiszman and Dina Demner-Fushman

14:40–15:00 *Evaluating Large-scale Text Mining Applications Beyond the Traditional Numeric Performance Measures*
Sofie Van Landeghem, Suwisa Kaewphan, Filip Ginter and Yves Van de Peer

15:00–15:20 *Recognizing Sublanguages in Scientific Journal Articles through Closure Properties*
Irina Temnikova and Kevin Cohen

15:30–16:00 Afternoon coffee break

16:00–16:20 *BEL Networks Derived from Qualitative Translations of BioNLP Shared Task Annotations*
Juliane Fluck, Alexander Klenner, Sumit Madan, Sam Ansari, Tamara Bobic, Julia Hoeng, Martin Hofmann-Apitius and Manuel Peitsch

16:20–16:40 *Exploring Word Class N-grams to Measure Language Development in Children*
Gabriela Ramirez-de-la-Rosa, Thamar Solorio, Manuel Montes, Yang Liu, Lisa Bedore, Elizabeth Pena and Aquiles Iglesias

Poster Session (16:40–17:30)

Adapting a Parser to Clinical Text by Simple Pre-processing Rules
Maria Skeppstedt

Using the Argumentative Structure of Scientific Literature to Improve Information Access
Antonio Jimeno Yepes, James Mork and Alan Aronson

Using Latent Dirichlet Allocation for Child Narrative Analysis
Khairun-nisa Hassanali, Yang Liu and Thamar Solorio

Effect of Out Of Vocabulary Terms on Inferring Eligibility Criteria for a Retrospective Study in Hebrew EHR
Raphael Cohen and Michael Elhadad

Parallels between Linguistics and Biology
Sutanu Chakraborti and Ashish Tendulkar