BiomedNLP 2011

Proceedings of the Workshop on Biomedical Natural Language Processing

held in conjunction with
the 8th International Conference on
Recent Advances in Natural Language Processing
(RANLP 2011)

15 September, 2011
Hissar, Bulgaria

INTERNATIONAL WORKSHOP
BIOMEDICAL NATURAL LANGUAGE PROCESSING

PROCEEDINGS

Hissar, Bulgaria
15 September 2011

ISBN 978-954-452-020-5

Designed and Printed by INCOMA Ltd.
Shoumen, BULGARIA

Preface

Biomedical natural language processing deals with the application of text mining techniques to clinical documents and to scientific publications in the areas of biology and medicine. Recent years have seen dramatic changes in the types and amount of data available to researchers in this field. Where most research on publications in the past has dealt with the abstracts of journal articles, we now have access to the full texts of journal articles via PubMedCentral. Where research on clinical documents has been hampered by a lack of availability of data, we now have access to large bodies of data through the auspices of the Cincinnati Children's Hospital NLP Challenge, the i2b2 shared tasks (www.i2b2.org), and the new TREC Electronic Medical Records track, as well as gold standard data being generated under US-funded Strategic Health Advanced Research Projects Area 4 (www.sharpn.org). Meanwhile, the number of abstracts in PubMed continues to grow exponentially. These are exciting times for biomedical NLP.

The Biomedical Information Extraction Workshop at RANLP 2011 provides a venue for presentation of current work in this field. Submissions in the areas of medicine and biology were solicited, as was research on the genres of publications, clinical documents, and web-based materials.

One theme present in the workshop was medical coding, represented by the work in Henriksson and Hassel (2011) and Boytcheva (2011). Other papers dealt with parsing (Kokkinakis 2011), evaluation of detection of personal health information (Sokolova 2011), and named entity recognition (Georgiev and Nakov 2011). There was a clear trend in the workshop towards working with clinically oriented problems or data rather than publication data, exemplified by most of the preceding papers, as well as work by Proux et al. on hospital-acquired infections and by Angelova and Boytcheva (2011) on the difficult problem of temporality in discharge notes. This is encouraging for the field in terms of its implications for the growing availability of clinically oriented data. A final encouraging note is that work was presented on data other than English, making this workshop nearly unique in its provision of a venue for work on this important topic.

We would like to thank all the authors for their efforts in making the event a highly productive workshop and a lively venue for exchange of scientific ideas. We also thank the Programme Committee members and the reviewers for providing high quality reviews. The workshop is organised with the partial support of project D0-02-292 EVTIMA "Effective search of conceptual information with applications in medical informatics" funded by the Bulgarian National Science Fund in 2009-2012.

September 2011

Guergana Savova
Kevin Bretonnel Cohen
Galia Angelova

Organizers:

Guergana Savova (Children's Hospital Boston and Harvard Medical School)
Kevin Bretonnel Cohen (University of Colorado School of Medicine)
Galia Angelova (IICT Bulgarian Academy of Sciences)

Programme Committee:

Vangelis Karkaletsis (Institute of Informatics and Telecommunications, Athens)
Dimitris Kokkinakis (Gothenburg University, Norway)
Frédérique Segond (Xerox Research Centre Europe, Grenoble)
Preslav Nakov (National University of Singapore, Singapore)
Pinar Wennerberg (Bayer AG, Germany)

Additional Reviewers:

Svetla Boytcheva (State University of Library Studies and Information Technologies, Bulgaria)
Georgi Georgiev (OntoText AD, Bulgaria)
Jiaping Zheng (Children's Hospital Boston and Harvard Medical School)
Tim Miller (Children's Hospital Boston and Harvard Medical School)

Table of Contents

Workshop Programme

Thursday, 15 September, 2011

9:30–10:30 *Invited talk: Assignment of Ontology-based Broad Semantic Classes to Biomedical Text*
Kevin Bretonnel Cohen

10:30–11:00 Coffee break

11:00–11:30 *Exploiting Structured Data, Negation Detection and SNOMED CT Terms in a Random Indexing Approach to Clinical Coding*
Aron Henriksson and Martin Hassel

11:30–12:00 *Automatic Matching of ICD-10 codes to Diagnoses in Discharge Letters*
Svetla Boytcheva

12:00–12:30 *Evaluation Measures for Detection of Personal Health Information*
Marina Sokolova

12:30–14:30 Lunch break

14:30-15:00 *Building a Named Entity Recognizer in Three Days: Application to Disease Name Recognition in Bulgarian Epicrises*
Georgi Georgiev, Valentin Zhivkov, Borislav Popov and Preslav Nakov

15:00–15:30 *Reducing Complexity in Parsing Scientific Medical Data, a Diabetes Case Study*
Dimitrios Kokkinakis

15:30–16:00 Coffee break

16:00–16:40 Poster Presentations

Architecture and Systems for Monitoring Hospital Acquired Infections inside Hospital Information Workflows
Denys Proux, Caroline Hagège, Quentin Gicquel, Suzanne Pereira, Stefan Darmoni, Frédérique Segond and Marie-Hélène Metzger

Towards Temporal Segmentation of Patient History in Discharge Letters
Galia Angelova and Svetla Boytcheva