ACL 2013

BioNLP Shared Task 2013

Proceedings of the Workshop

August 9, 2013
Sofia, Bulgaria

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

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: 978-1-937284-55-8

Introduction

The BioNLP Shared Task (BioNLP-ST) series represents a community-wide trend in text-mining for biology toward fine-grained information extraction (IE). The two previous events, BioNLP-ST 2009 and 2011, attracted wide attention, with over 30 teams submitting final results. The tasks and their data have since served as the basis of numerous studies, released event extraction systems, and published datasets. As in previous events, the results of BioNLP-ST 2013 are presented at the ACL/HLT BioNLP-ST workshop colocated with the BioNLP workshop in Sofia, Bulgaria (9 August 2013).

BioNLP-ST 2013 follows the general outline and goals of the previous tasks. It identifies biologically relevant extraction targets and proposes a linguistically motivated approach to event representation. The tasks in BioNLP-ST 2013 cover many new hot topics in biology that are close to biologist needs. BioNLP-ST 2013 broadens the scope of the text-mining application domains in biology by introducing new issues on cancer genetics and pathway curation. It also builds on the well-known previous datasets GENIA, LLL/BI and BB to propose more realistic tasks that considered previously, closer to the actual needs of biological data integration.

The first event in 2009 triggered active research in the community on a specific fine-grained IE task. Expanding on this, the second BioNLP-ST was organized under the theme "Generalization", which was well received by participants, who introduced numerous systems that could be straightforwardly applied to multiple tasks. This time, the BioNLP-ST takes a step further and pursues the grand theme of "Knowledge base construction", which is addressed in various ways: semantic web (GE, GRO), pathways (PC), molecular mechanisms of cancer (CG), regulation networks (GRN) and ontology population (GRO, BB). A general overview paper in this volume summarizes the organization and participation in the shared tasks, with 22 teams submitted 38 final results this year. Each specific task is additionally covered by an overview paper.

As in previous events, manually annotated data were provided for training, development and evaluation of information extraction methods. According to their relevance for biological studies, the annotations are either bound to specific expressions in the text or represented as structured knowledge. Tools for the evaluation of system outputs are publicly available. Support in performing linguistic processing was provided to the participants in the form of analyses created by various state-of-the art tools on the dataset texts. A last overview paper is dedicated to the preparation of these supporting resources.

Thanks to the many excellent manuscripts received from participants and the efforts of the programme committee, it is our pleasure to present these proceedings describing the task and the participating systems.

Claire Nédellec — Organizing Chair
Robert Bossy — BB and GRN Task Chair
Jin-Dong Kim — GE Task Chair
Jung-jae Kim — GRO Task Chair
Tomoko Ohta — PC Task Chair
Sampo Pyysalo — CG Task Chair
Pierre Zweigenbaum — PC Chair

Committees

Scientific Advisory Board

Jun'ichi Tsujii (Microsoft) .... *Chair*
Philippe Bessières (INRA)
Sung-Pil Choi (KISTI)
Kevin Cohen (Univ. Colorado)
Yuji Kohara (DBCLS)
Tapio Salakoski (Univ. Turku)
Pierre Zweigenbaum (CNRS)

Organizing Committee

Claire Nédellec (INRA) .... *Organizing Chair*
Sophia Ananiadou (NaCTeM, Univ. Manchester)
Robert Bossy (INRA) .... *Task BB and GRN Chair*
Julien Jourde (INRA)
Jin-Dong Kim (DBCLS) .... *Task GE Chair*
Jung-jae Kim (NTU, Singapore) .... *Task GRO Chair*
Tomoko Ohta (NaCTeM, Univ. Manchester) .. *Task PC Chair*
Sampo Pyysalo (NaCTeM, Univ. Manchester) *Task CG Chair*
Pontus Stenetorp (Univ. Tokyo)
Yue Wang (DBCLS)

Programme Committee

Pierre Zweigenbaum, National Center for Scientific Research (CNRS) .... *PC Chair*
Sophia Ananiadou, University of Manchester (NaCTeM)
Nathalie Aussenac-Gilles, National Center for Scientific Research (CNRS)
Sabine Bergler, Concordia University
Philippe Bessières, National Institute for Agricultural Research (INRA)
Robert Bossy, National Institute for Agricultural Research (INRA)
Kevin Bretonnel Cohen, University of Colorado
Berry de Bruijn, National Research Council (NRC)
Dina Demner-Fushman, National Library of Medicine (NLM)
Jörg Hakenberg, Arizona State University
Jin-Dong Kim, Database Center for Life Science (DBCLS)
Jung-Jae Kim, Nanyang Technological University
Martin Krallinger, National Biotechnology Center (CNB)
David McClosky, Stanford University
Roser Morante, University of Antwerp
Claire Nédellec, National Institute for Agricultural Research (INRA)
Tomoko Ohta, University of Manchester (NaCTeM)
Thierry Poibeau, National Center for Scientific Research (CNRS)
Sampo Pyysalo, University of Manchester (NaCTeM)
Rafal Rak, University of Manchester (NaCTeM)
Sebastian Riedel, University of Massachusetts
Fabio Rinaldi, University of Zurich
Yvan Saeys, Ghent University
Tapio Salakoski, University of Turku
Rune Sætre, Norwegian University of Science and Technology (NTNU)
Özlem Uzuner, State University of New York
Andreas Vlachos, University of Cambridge

Task organizers

GE task

Jin-Dong Kim (DBCLS)
Yue Wang (DBCLS)
Yasunori Yamamoto (DBCLS)
Sabine Bergler (Concordia Univ.)
Roser Morante (Univ. Antwerp)
Kevin Cohen (Univ. Colorado)

CG task

Sampo Pyysalo (NaCTeM and Univ. Manchester)
Tomoko Ohta (NaCTeM and Univ. Manchester)
Rafal Rak (NaCTeM and Univ. Manchester)
Andrew Rowley (NaCTeM and Univ. Manchester)
Jacob Carter (NaCTeM and Univ. Manchester)
Sophia Ananiadou (NaCTeM and Univ. Manchester)

PC task

Tomoko Ohta (NaCTeM and Univ. Manchester)
Sampo Pyysalo (NaCTeM and Univ. Manchester)
Rafal Rak (NaCTeM and Univ. Manchester)
Andrew Rowley (NaCTeM and Univ. Manchester)
Jacob Carter (NaCTeM and Univ. Manchester)
Sophia Ananiadou (NaCTeM and Univ. Manchester)
Sung-Pil Choi (KISTI)
Hong-woo Chun (KISTI)
Sung-jae Jung (KISTI)
Hyun Uk Kim (KAIST)
Jinki Kim (KAIST)
Kyusang Hwang (KAIST)
Yonghwa Jo
Hyeyeon Choi

GRO task

Jung-jae Kim (NTU)
Han Xu (NTU)
Dietrich Rebholz-Schuhmann (Univ. Zurich)
Vivian Lee (EBI)

GRN task

Robert Bossy (INRA)
Philippe Bessières (INRA)
Frédéric Papazian (INRA)
Claire Nédellec (INRA)

BB task

Robert Bossy (INRA)
Philippe Bessières (INRA)
Wiktoria Golik (INRA)
Frédéric Papazian (INRA)
Zorana Ratkovic (INRA)
Claire Nédellec (INRA)

Table of Contents

Workshop Program

Friday, August 9, 2013

(8:30-9:00) Welcome and Introduction

08:45	*Overview of BioNLP Shared Task 2013* Claire Nédellec, Robert Bossy, Jin-Dong Kim, Jung-Jae Kim, Tomoko Ohta, Sampo Pyysalo and Pierre Zweigenbaum

Session 1: (9:00-10:30) Oral Presentations: Genia Event Extraction and Gene Regulation Ontology

9:00	*The Genia Event Extraction Shared Task, 2013 Edition - Overview* Jin-Dong Kim, Yue Wang and Yamamoto Yasunori
9:10–9:30	*TEES 2.1: Automated Annotation Scheme Learning in the BioNLP 2013 Shared Task* Jari Björne and Tapio Salakoski
9:30–9:50	*EVEX in ST'13: Application of a large-scale text mining resource to event extraction and network construction* Kai Hakala, Sofie Van Landeghem, Tapio Salakoski, Yves Van de Peer and Filip Ginter
9:50–10:10	*Extracting Biomedical Events and Modifications Using Subgraph Matching with Noisy Training Data* Andrew MacKinlay, David Martinez, Antonio Jimeno Yepes, Haibin Liu, W John Wilbur and Karin Verspoor
10:10–10:30	*Biomedical Event Extraction by Multi-class Classification of Pairs of Text Entities* Xiao Liu, Antoine Bordes and Yves Grandvalet

(10:30-11:00) Break

Friday, August 9, 2013 (continued)

Session 2: (11:00-12:30) Oral Presentations: Cancer Genetics and Pathway Curation

11:00 *GRO Task: Populating the Gene Regulation Ontology with events and relations*
Jung-Jae Kim, Xu Han, Vivian Lee and Dietrich Rebholz-Schuhmann

11:10 *Overview of the Cancer Genetics (CG) task of BioNLP Shared Task 2013*
Sampo Pyysalo, Tomoko Ohta and Sophia Ananiadou

11:20 *Overview of the Pathway Curation (PC) task of BioNLP Shared Task 2013*
Tomoko Ohta, Sampo Pyysalo, Rafal Rak, Andrew Rowley, Hong-Woo Chun, Sung-Jae Jung, Sung-Pil Choi, Sophia Ananiadou and Jun'ichi Tsujii

11:30–11:50 *Generalizing an Approximate Subgraph Matching-based System to Extract Events in Molecular Biology and Cancer Genetics*
Haibin Liu, Karin Verspoor, Donald C. Comeau, Andrew MacKinlay and W John Wilbur

11:50–12:10 *Performance and limitations of the linguistically motivated Cocoa/Peaberry system in a broad biological domain.*
SV Ramanan and P. Senthil Nathan

12:10–12:30 *NaCTeM EventMine for BioNLP 2013 CG and PC tasks*
Makoto Miwa and Sophia Ananiadou

(12:30-14:00) Lunch Break

Session 3: (14:00-15:30) Posters

BioNLP Shared Task 2013: Supporting Resources
Pontus Stenetorp, Wiktoria Golik, Thierry Hamon, Donald C. Comeau, Rezarta Islamaj Dogan, Haibin Liu and W John Wilbur

A fast rule-based approach for biomedical event extraction
Quoc-Chinh Bui, David Campos, Erik van Mulligen and Jan Kors

Improving Feature-Based Biomedical Event Extraction System by Integrating Argument Information
Lishuang Li, Yiwen Wang and Degen Huang

UZH in BioNLP 2013
Gerold Schneider, Simon Clematide, Tilia Ellendorff, Don Tuggener, Fabio Rinaldi and Gintarė Grigonytė

Friday, August 9, 2013 (continued)

A Hybrid approach for biomedical event extraction
Xuan Quang Pham, Minh Quang Le and Bao Quoc Ho

Identification of Genia Events using Multiple Classifiers
Roland Roller and Mark Stevenson

Exploring a Probabilistic Earley Parser for Event Composition in Biomedical Texts
Mai-Vu Tran, Nigel Collier, Hoang-Quynh Le, Van-Thuy Phi and Thanh-Binh Pham

Detecting Relations in the Gene Regulation Network
Thomas Provoost and Marie-Francine Moens

Ontology-based semantic annotation: an automatic hybrid rule-based method
Sondes Bannour, Laurent Audibert and Henry Soldano

Building A Contrasting Taxa Extractor for Relation Identification from Assertions: BIOlogical Taxonomy & Ontology Phrase Extraction System
Cyril Grouin

(15:30-16:00) Break

Session 4: (16:00-17:30) Oral Presentations: Bacteria: Gene Regulation Network and Biotope

16:00 *BioNLP Shared Task 2013 – An overview of the Genic Regulation Network Task*
Robert Bossy, Philippe Bessières and Claire Nédellec

16:10 *BioNLP shared Task 2013 – An Overview of the Bacteria Biotope Task*
Robert Bossy, Wiktoria Golik, Zorana Ratkovic, Philippe Bessières and Claire Nédellec

16:20–16:40 *Bacteria Biotope Detection, Ontology-based Normalization, and Relation Extraction using Syntactic Rules*
İlknur Karadeniz and Arzucan Özgür

16:40–17:00 *Extracting Gene Regulation Networks Using Linear-Chain Conditional Random Fields and Rules*
Slavko Zitnik, Marinka Žitnik, Blaž Zupan and Marko Bajec

17:00–17:20 *IRISA participation to BioNLP-ST13: lazy-learning and information retrieval for information extraction tasks*
Vincent Claveau

Session 5: (17:30-18:00) General Discussion