ACL 2016

The 4th BioNLP Shared Task

Proceedings of the 4th BioNLP Shared Task Workshop

August 13, 2016
Berlin, Germany

©2016 The Association for Computational Linguistics

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 978-1-945626-21-0

Introduction

The BioNLP Shared Task series pursues establishing a community-wide effort for fine-grained information extraction (IE) in biology domains. As in the third edition (2013), the fourth edition focuses on knowledge base construction through IE. Nowadays some BioNLP tools, e.g., protein name recognizers or protein-protein interaction extractors, are reaching a level of applicability which allows them to be integrated into bioinformatics systems and thus to significantly contribute to the bioinformatics and biology research. The BioNLP community then faces the challenge of the integration of information extracted from the text to external resources like -omics databases, biological ontologies, or systems biology operational models. BioNLP-ST 2016 offers tasks by which new or existing data, and evaluation methods are expected to be adapted to this trend.

BioNLP-ST 2016 comprises three tasks that address different aspects of knowledge acquisition from text and also encompasses a wide range of biological diversity:

- SeeDev, which aims at extracting the regulation of the seed development in plants using a rich model;
- Bacteria Biotopes 3 (BB3) for the construction of a bacteria habitat database using external ontologies;
- Genia 4 (GE4), which aims at delivering new shared task framework to construct knowledge base of NFκB synthesis and regulation through IE.

The previous editions had attracted many participants and it has then supported active research on information extraction and database/knowledge base integration. In this edition, 26 final results were submitted by 15 distinct teams from 10 different countries of four continents. This year, BioNLP-ST is organized as a joint event with the BioASQ challenge that has converging goals on biological question answering and semantic indexing. The BioNLP-ST/BioASQ workshop is collocated with the BioNLP workshop hosted by the ACL/HLT 2016 conference in Berlin, Germany. In addition to the participating systems, an overview of each task is also presented at the workshop.

Thanks to the many excellent manuscripts received from participants and the efforts of the programme committee, it is our pleasure to present these proceedings that describe the BioNLP Shared Task and the participating systems.

Claire Nédellec, Robert Bossy and Jin-Dong Kim

Organizing Committee:

Claire Nédellec (INRA) - **Chair**
Robert Bossy (INRA)
Jin-Dong Kim (DBCLS)

Program Committee:

Robert Bossy (INRA) - **Chair**
Jin-Dong Kim (DBCLS) - **Chair**
Fabio Rinaldi (University of Zürich)
Jari Björne (University of Turku)
Jörg Hakenberg (Illumina)
Tomoko Ohta (Textimi)
Philippe Bessières (INRA)
Filip Ginter (University of Turku)
Roser Morante (Vrije University of Amsterdam)
David McClosky (Google)
Yuka Tateisi (NBDC)
Claire Nédellec (INRA)
Georgios Paliouras (Demokritos)
Özlem Uzuner (University of Albany)
Berry de Bruijn (National Research Council Canada)
Anastasia Krithara (Demokritos)
Sabine Bergler (Concordia University)
Jung-Jae Kim (Institute for Infocomm Research)
Kevin B. Cohen (University of Colorado)
Pierre Zweigenbaum (CNRS)
Louise Deléger (INRA)

Task Organizers:

SeeDev Task
Robert Bossy (INRA) - **Chair**
Estelle Chaix (INRA)
Claire Nédellec (INRA)

BB3 Task
Robert Bossy (INRA) - **Chair**
Claire Nédellec (INRA)

GE4 Task
Jin-Dong Kim (DBCLS) - **Chair**
Yue Wang (DBCLS)

Table of Contents

Overview of the Regulatory Network of Plant Seed Development (SeeDev) Task at the BioNLP Shared Task 2016.
Estelle Chaix, Bertrand Dubreucq, Abdelhak Fatihi, Dialekti Valsamou, Robert Bossy, Mouhamadou Ba, Louise Deléger, Pierre Zweigenbaum, Philippe Bessières, Loïc Lepiniec and Claire Nédellec 1

Overview of the Bacteria Biotope Task at BioNLP Shared Task 2016
Louise Deléger, Robert Bossy, Estelle Chaix, Mouhamadou Ba, Arnaud Ferré, Philippe Bessières and Claire Nédellec 12

Refactoring the Genia Event Extraction Shared Task Toward a General Framework for IE-Driven KB Development
Jin-Dong Kim, Yue Wang, Nicola Colic, Seung Han Beak, Yong Hwan Kim and Min Song 23

LitWay, Discriminative Extraction for Different Bio-Events
Chen Li, Zhiqiang Rao and Xiangrong Zhang 32

VERSE: Event and Relation Extraction in the BioNLP 2016 Shared Task
Jake Lever and Steven JM Jones 42

A dictionary- and rule-based system for identification of bacteria and habitats in text
Helen V Cook, Evangelos Pafilis and Lars Juhl Jensen 50

Ontology-Based Categorization of Bacteria and Habitat Entities using Information Retrieval Techniques
Mert Tiftikci, Hakan Şahin, Berfu Büyüköz, Alper Yayıkçı and Arzucan Özgür 56

Identification of Mentions and Relations between Bacteria and Biotope from PubMed Abstracts
Cyril Grouin 64

Deep Learning with Minimal Training Data: TurkuNLP Entry in the BioNLP Shared Task 2016
Farrokh Mehryary, Jari Björne, Sampo Pyysalo, Tapio Salakoski and Filip Ginter 73

SeeDev Binary Event Extraction using SVMs and a Rich Feature Set
Nagesh C. Panyam, Gitansh Khirbat, Karin Verspoor, Trevor Cohn and Kotagiri Ramamohanarao 82

Extraction of Regulatory Events using Kernel-based Classifiers and Distant Supervision
Andre Lamurias, Miguel J. Rodrigues, Luka A. Clarke and Francisco M. Couto 88

DUTIR in BioNLP-ST 2016: Utilizing Convolutional Network and Distributed Representation to Extract Complicate Relations
Honglei Li, Jianhai Zhang, Jian Wang, Hongfei Lin and Zhihao Yang 93

Extracting Biomedical Event Using Feature Selection and Word Representation
Xinyu He, Lishuang Li, Jieqiong Zheng and Meiyue Qin 101

Conference Program

Saturday, August 13, 2016

14:00–14:15 Overview of BioNLP-ST

Overview of the Regulatory Network of Plant Seed Development (SeeDev) Task at the BioNLP Shared Task 2016.
Estelle Chaix, Bertrand Dubreucq, Abdelhak Fatihi, Dialekti Valsamou, Robert Bossy, Mouhamadou Ba, Louise Deléger, Pierre Zweigenbaum, Philippe Bessières, Loïc Lepiniec and Claire Nédellec

Overview of the Bacteria Biotope Task at BioNLP Shared Task 2016
Louise Deléger, Robert Bossy, Estelle Chaix, Mouhamadou Ba, Arnaud Ferré, Philippe Bessières and Claire Nédellec

Refactoring the Genia Event Extraction Shared Task Toward a General Framework for IE-Driven KB Development
Jin-Dong Kim, Yue Wang, Nicola Colic, Seung Han Beak, Yong Hwan Kim and Min Song

14:15–15:30 BioNLP-ST participant session 1

14:15–14:30 *LitWay, Discriminative Extraction for Different Bio-Events*
Chen Li, Zhiqiang Rao and Xiangrong Zhang

14:30–14:45 *VERSE: Event and Relation Extraction in the BioNLP 2016 Shared Task*
Jake Lever and Steven JM Jones

14:45–15:00 *A dictionary- and rule-based system for identification of bacteria and habitats in text*
Helen V Cook, Evangelos Pafilis and Lars Juhl Jensen

15:00–15:15 *Ontology-Based Categorization of Bacteria and Habitat Entities using Information Retrieval Techniques*
Mert Tiftikci, Hakan Şahin, Berfu Büyüköz, Alper Yayıkçı and Arzucan Özgür

15:15–15:30 *Identification of Mentions and Relations between Bacteria and Biotope from PubMed Abstracts*
Cyril Grouin

Saturday, August 13, 2016 (continued)

15:30–16:00 Coffee break

16:00–17:00 BioNLP-ST participant session 2

16:00–16:15 *Deep Learning with Minimal Training Data: TurkuNLP Entry in the BioNLP Shared Task 2016*
Farrokh Mehryary, Jari Björne, Sampo Pyysalo, Tapio Salakoski and Filip Ginter

16:15–16:30 *SeeDev Binary Event Extraction using SVMs and a Rich Feature Set*
Nagesh C. Panyam, Gitansh Khirbat, Karin Verspoor, Trevor Cohn and Kotagiri Ramamohanarao

16:30–16:45 *Extraction of Regulatory Events using Kernel-based Classifiers and Distant Supervision*
Andre Lamurias, Miguel J. Rodrigues, Luka A. Clarke and Francisco M. Couto

16:45–17:00 *DUTIR in BioNLP-ST 2016: Utilizing Convolutional Network and Distributed Representation to Extract Complicate Relations*
Honglei Li, Jianhai Zhang, Jian Wang, Hongfei Lin and Zhihao Yang

17:00–17:30 Closing session

Poster

Extracting Biomedical Event Using Feature Selection and Word Representation
Xinyu He, Lishuang Li, Jieqiong Zheng and Meiyue Qin