

ACL HLT 2011

**The 49th Annual Meeting of the  
Association for Computational Linguistics:  
Human Language Technologies**

**Proceedings of BioNLP Shared Task 2011 Workshop**

24 June, 2011  
Portland, Oregon, USA

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

©2011 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](mailto:acl@aclweb.org)

ISBN-13 9781937284091

## Introduction

The requirements of improved access to the massive amount of scientific literature in biomedical domain - through applications such as semantic search, assisted pathway annotation, and the automatic identification of specific biomolecular reactions for database curation support - place continuing demands on the development of methods and resources for advanced biomedical information extraction and text mining. The BioNLP Shared Task series seeks to advance this development through an increased focus on detailed structured representations of extracted information, novel corpus resources with fully text-bound annotation, and precise task definitions, support and evaluation.

The BioNLP Shared Task 2011 is the second in the series, following up on the first event organized in 2009. Seeking to build on the success of the previous event, the task was organized as a collaboration between several groups in Asia, Europe and the US who defined in total eight specific tasks involving diverse challenges, including in addition to structured event extraction also relation extraction and supporting tasks such as coreference resolution. The main theme of the 2011 event was generalization, and the main tasks further broadened on the 2009 setup in three aspects: text types, subject domains, and novel event extraction targets.

The task attracted broad interest from the community, and a total of 46 final submissions were received from 24 groups, maintaining the 2009 task participation numbers while nearly doubling its number of submissions. In addition to the continued interest from the biomedical text mining community, we were glad to welcome the participation of many new groups from academia and industry. The submissions demonstrated substantial progress at the established event extraction task and showed that event extraction methods generalize well, among other aspects, to full papers, new subject domains such as infectious diseases and bacterial interactions, and new sets of events such as protein post-translational modifications.

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.

BioNLP Shared Task 2011 chairs

Jun'ichi Tsujii  
Jin-Dong Kim  
Sampo Pyysalo



**Scientific Advisory Committee:**

Jun'ichi Tsujii (Chair), Microsoft Research Asia  
Sophia Ananiadou, National Centre for Text Mining (NaCTeM)  
Kevin Cohen, University of Colorado, and MITRE  
Claire Nédellec, French National Institute for Agricultural Research (INRA)  
Andrey Rzhetsky, University of Chicago  
Bruno Sobral, Virginia Bioinformatics Institute  
Tapio Salakoski, University of Turku  
Toshihisa Takagi, Database Center for Life Science (DBCLS)

**Organizing Committee:**

Jin-Dong Kim (Chair), Database Center for Life Science (DBCLS)  
Sampo Pyysalo (Chair), University of Tokyo  
Tomoko Ohta, University of Tokyo  
Robert Bossy, French National Institute for Agricultural Research (INRA)  
Chunhong Mao, Virginia Bioinformatics Institute  
Dan Sullivan, Virginia Bioinformatics Institute  
Rafal Rak, National Centre for Text Mining (NaCTeM)  
Ngan Nguyen, University of Tokyo

**Program Committee:**

Timothy Baldwin, University of Melbourne  
Sabine Bergler, Concordia University  
Olivier Bodenreider, National Library of Medicine (NLM)  
Wendy Chapman, University of California, San Diego (UCSD)  
Kevin Cohen, University of Colorado, and MITRE  
Nigel Collier, National Institute of Informatics (NII)  
Filip Ginter, University of Turku  
Jörg Hakenberg, Arizona State University  
Minlie Huang, Tsinghua University  
Su Jian, Institute for Infocomm Research  
Min-Yen Kan, National University of Singapore  
Jung-Jae Kim, Nanyang Technological University  
Martin Krallinger, National Biotechnology Center (CNB)  
Zhiyong Lu, National Library of Medicine (NLM)  
David McClosky, Stanford University  
Roser Morante, University of Antwerp  
Claire Nédellec, French National Institute for Agricultural Research (INRA)  
Serguei Pakhomov, University of Minnesota  
Thierry Poibeau, French Institute for Fundamental Research (CNRS)  
Hoifung Poon, University of Washington

Sebastian Riedel, University of Massachusetts  
Fabio Rinaldi, University of Zurich  
Thomas Rindflesch, National Library of Medicine (NLM)  
Yvan Saeys, Ghent University  
Tapio Salakoski, University of Turku  
Hagit Shatkay, University of Delaware  
Rune Sætre, Norwegian University of Science and Technology (NTNU)  
Yuka Tateisi, Kogakuin University  
Yoshimasa Tsuruoka, Japan Advanced Institute of Science and Technology (JAIST)  
Karin Verspoor, University of Colorado  
Xinglong Wang, National Centre for Text Mining  
Hong Yu, University of Wisconsin-Milwaukee  
Pierre Zweigenbaum, French National Center for Scientific Research (CNRS)

## Table of Contents

<i>Overview of BioNLP Shared Task 2011</i>	
Jin-Dong Kim, Sampo Pyysalo, Tomoko Ohta, Robert Bossy, Ngan Nguyen and Jun'ichi Tsujii	1
<i>Overview of Genia Event Task in BioNLP Shared Task 2011</i>	
Jin-Dong Kim, Yue Wang, Toshihisa Takagi and Akinori Yonezawa	7
<i>Overview of the Epigenetics and Post-translational Modifications (EPI) task of BioNLP Shared Task 2011</i>	
Tomoko Ohta, Sampo Pyysalo and Jun'ichi Tsujii	16
<i>Overview of the Infectious Diseases (ID) task of BioNLP Shared Task 2011</i>	
Sampo Pyysalo, Tomoko Ohta, Rafal Rak, Dan Sullivan, Chunhong Mao, Chunxia Wang, Bruno Sobral, Jun'ichi Tsujii and Sophia Ananiadou	26
<i>Biomedical Event Extraction from Abstracts and Full Papers using Search-based Structured Prediction</i>	
Andreas Vlachos and Mark Craven	36
<i>Event Extraction as Dependency Parsing for BioNLP 2011</i>	
David McClosky, Mihai Surdeanu and Christopher Manning	41
<i>Robust Biomedical Event Extraction with Dual Decomposition and Minimal Domain Adaptation</i>	
Sebastian Riedel and Andrew McCallum	46
<i>Model Combination for Event Extraction in BioNLP 2011</i>	
Sebastian Riedel, David McClosky, Mihai Surdeanu, Andrew McCallum and Christopher D. Manning	51
<i>BioNLP Shared Task 2011 - Bacteria Biotope</i>	
Robert Bossy, Julien Jourde, Philippe Bessières, Maarten van de Guchte and Claire Nédellec	56
<i>BioNLP Shared Task 2011 – Bacteria Gene Interactions and Renaming</i>	
Julien Jourde, Alain-Pierre Manine, Philippe Veber, Karën Fort, Robert Bossy, Erick Alphonse and Philippe Bessières	65
<i>Overview of BioNLP 2011 Protein Coreference Shared Task</i>	
Ngan Nguyen, Jin-Dong Kim and Jun'ichi Tsujii	74
<i>Overview of the Entity Relations (REL) supporting task of BioNLP Shared Task 2011</i>	
Sampo Pyysalo, Tomoko Ohta and Jun'ichi Tsujii	83
<i>The Taming of Reconcile as a Biomedical Coreference Resolver</i>	
Youngjun Kim, Ellen Riloff and Nathan Gilbert	89
<i>Extracting Bacteria Biotopes with Semi-supervised Named Entity Recognition and Coreference Resolution</i>	
Nhung T. H. Nguyen and Yoshimasa Tsuruoka	94

<i>BioNLP 2011 Task Bacteria Biotope – The Alvis system</i>	
Zorana Ratkovic, Wiktorija Golik, Pierre Warnier, Philippe Veber and Claire Nédellec . . . . .	102
<i>BioNLP Shared Task 2011: Supporting Resources</i>	
Pontus Stenetorp, Goran Topić, Sampo Pyysalo, Tomoko Ohta, Jin-Dong Kim and Jun’ichi Tsujii	112
<i>Sentence Filtering for BioNLP: Searching for Renaming Acts</i>	
Pierre Warnier and Claire Nédellec . . . . .	121
<i>Complex Biological Event Extraction from Full Text using Signatures of Linguistic and Semantic Features</i>	
Liam R. McGrath, Kelly Domico, Courtney D. Corley and Bobbie-Jo Webb-Robertson . . . . .	130
<i>Using Kybots for Extracting Events in Biomedical Texts</i>	
Arantza Casillas, Arantza Díaz de Ilarraza, Koldo Gojenola, Maite Oronoz and German Rigau	138
<i>Extracting Biological Events from Text Using Simple Syntactic Patterns</i>	
Quoc-Chinh Bui and Peter. M.A. Sloot . . . . .	143
<i>Detecting Entity Relations as a Supporting Task for Bio-Molecular Event Extraction</i>	
Sofie Van Landeghem, Thomas Abeel, Bernard De Baets and Yves Van de Peer . . . . .	147
<i>A Pattern Approach for Biomedical Event Annotation</i>	
Quang Le Minh, Son Nguyen Truong and Quoc Ho Bao . . . . .	149
<i>An Incremental Model for the Coreference Resolution Task of BioNLP 2011</i>	
Don Tugener, Manfred Klenner, Gerold Schneider, Simon Clematide and Fabio Rinaldi . . . .	151
<i>Double Layered Learning for Biological Event Extraction from Text</i>	
Ehsan Emadzadeh, Azadeh Nikfarjam and Graciela Gonzalez . . . . .	153
<i>MSR-NLP Entry in BioNLP Shared Task 2011</i>	
Chris Quirk, Pallavi Choudhury, Michael Gamon and Lucy Vanderwende . . . . .	155
<i>From Graphs to Events: A Subgraph Matching Approach for Information Extraction from Biomedical Text</i>	
Haibin Liu, Ravikumar Komandur and Karin Verspoor . . . . .	164
<i>Adapting a General Semantic Interpretation Approach to Biological Event Extraction</i>	
Halil Kilicoglu and Sabine Bergler . . . . .	173
<i>Generalizing Biomedical Event Extraction</i>	
Jari Björne and Tapio Salakoski . . . . .	183



# Conference Program

**Friday, June 24, 2011**

## **Session 1: Oral presentations and discussion**

- 09:00–09:25 Overview of BioNLP Shared Task 2011 (I) - *Overall Organization*, and *GE*, *IDE* and *ID* Tasks  
Shared Task Organizers
- 09:25–09:40 *Biomedical Event Extraction from Abstracts and Full Papers using Search-based Structured Prediction*  
Andreas Vlachos and Mark Craven
- 09:40–09:55 *Event Extraction as Dependency Parsing for BioNLP 2011*  
David McClosky, Mihai Surdeanu and Christopher Manning
- 09:55–10:10 *Robust Biomedical Event Extraction with Dual Decomposition and Minimal Domain Adaptation*  
Sebastian Riedel and Andrew McCallum
- 10:10–10:25 *Model Combination for Event Extraction in BioNLP 2011*  
Sebastian Riedel, David McClosky, Mihai Surdeanu, Andrew McCallum and Christopher D. Manning
- 10:25–10:30 Discussion
- 10:30–11:00 Morning break

**Friday, June 24, 2011 (continued)**

**Session 2: Oral presentations and discussion**

- 11:00–11:20 Overview of BioNLP Shared Task 2011 (II) - Bacteria Track (*BB*, *BI*) and Supporting Tasks (*CO*, *REL* and *REN*)  
Shared Task Organizers
- 11:20–11:35 *The Taming of Reconcile as a Biomedical Coreference Resolver*  
Youngjun Kim, Ellen Riloff and Nathan Gilbert
- 11:35–11:55 *Extracting Bacteria Biotopes with Semi-supervised Named Entity Recognition and Coreference Resolution*  
Nhung T. H. Nguyen and Yoshimasa Tsuruoka
- 11:55–12:15 *BioNLP 2011 Task Bacteria Biotope – The Alvis system*  
Zorana Ratkovic, Wiktorina Golik, Pierre Warnier, Philippe Veber and Claire Nédellec
- 12:15–12:30 Discussion
- 12:30–14:00 Lunch break

**Friday, June 24, 2011 (continued)**

**Session 3: Poster presentations**

14:00–14:10 Spotlight presentation

14:10–15:30 Poster presentations

*BioNLP Shared Task 2011: Supporting Resources*

Pontus Stenetorp, Goran Topić, Sampo Pyysalo, Tomoko Ohta, Jin-Dong Kim and Jun'ichi Tsujii

*Sentence Filtering for BioNLP: Searching for Renaming Acts*

Pierre Warnier and Claire Nédellec

*Complex Biological Event Extraction from Full Text using Signatures of Linguistic and Semantic Features*

Liam R. McGrath, Kelly Domico, Courtney D. Corley and Bobbie-Jo Webb-Robertson

*Using Kybots for Extracting Events in Biomedical Texts*

Arantza Casillas, Arantza Díaz de Ilarraza, Koldo Gojenola, Maite Oronoz and German Rigau

*Extracting Biological Events from Text Using Simple Syntactic Patterns*

Quoc-Chinh Bui and Peter. M.A. Sloom

*Detecting Entity Relations as a Supporting Task for Bio-Molecular Event Extraction*

Sofie Van Landeghem, Thomas Abeel, Bernard De Baets and Yves Van de Peer

*A Pattern Approach for Biomedical Event Annotation*

Quang Le Minh, Son Nguyen Truong and Quoc Ho Bao

*An Incremental Model for the Coreference Resolution Task of BioNLP 2011*

Don Tuggener, Manfred Klenner, Gerold Schneider, Simon Clematide and Fabio Rinaldi

*Double Layered Learning for Biological Event Extraction from Text*

Ehsan Emadzadeh, Azadeh Nikfarjam and Graciela Gonzalez

15:30–16:00 Afternoon break

**Friday, June 24, 2011 (continued)**

**Session 4: Oral presentations and discussion**

- 16:00–16:20 *MSR-NLP Entry in BioNLP Shared Task 2011*  
Chris Quirk, Pallavi Choudhury, Michael Gamon and Lucy Vanderwende
- 16:20–16:40 *From Graphs to Events: A Subgraph Matching Approach for Information Extraction from Biomedical Text*  
Haibin Liu, Ravikumar Komandur and Karin Verspoor
- 16:40–16:55 *Adapting a General Semantic Interpretation Approach to Biological Event Extraction*  
Halil Kilicoglu and Sabine Bergler
- 16:55–17:15 *Generalizing Biomedical Event Extraction*  
Jari Björne and Tapio Salakoski
- 17:15–17:30 Discussion