

# **15th Workshop on Biomedical Natural Language Processing (BioNLP 2016)**

Held at ACL 2016

Berlin, Germany  
12 August 2016

ISBN: 978-1-5108-2766-0

**Printed from e-media with permission by:**

Curran Associates, Inc.  
57 Morehouse Lane  
Red Hook, NY 12571



**Some format issues inherent in the e-media version may also appear in this print version.**

Copyright© (2016) by the Association for Computational Linguistics  
All rights reserved.

Printed by Curran Associates, Inc. (2016)

For permission requests, please contact the Association for Computational Linguistics  
at the address below.

Association for Computational Linguistics  
209 N. Eighth Street  
Stroudsburg, Pennsylvania 18360

Phone: 1-570-476-8006  
Fax: 1-570-476-0860

[acl@aclweb.org](mailto:acl@aclweb.org)

**Additional copies of this publication are available from:**

Curran Associates, Inc.  
57 Morehouse Lane  
Red Hook, NY 12571 USA  
Phone: 845-758-0400  
Fax: 845-758-2633  
Email: [curran@proceedings.com](mailto:curran@proceedings.com)  
Web: [www.proceedings.com](http://www.proceedings.com)

## Table of Contents

<i>A Machine Learning Approach to Clinical Terms Normalization</i> Jose Castano, María Laura Gambarte, Hee Joon Park, Maria del Pilar Avila Williams, David Perez, Fernando Campos, Daniel Luna, Sonia Benitez, Hernan Berinsky and Sofía Zanetti . . . . .	1
<i>Improved Semantic Representation for Domain-Specific Entities</i> Mohammad Taher Pilehvar and Nigel Collier . . . . .	12
<i>Identification, characterization, and grounding of gradable terms in clinical text</i> Chaitanya Shivade, Marie-Catherine de Marneffe, Eric Fosler-Lussier and Albert M. Lai . . . . .	17
<i>Graph-based Semi-supervised Gene Mention Tagging</i> Golnar Sheikhshab, Elizabeth Starks, Aly Karsan, Anoop Sarkar and Inanc Birol . . . . .	27
<i>Feature Derivation for Exploitation of Distant Annotation via Pattern Induction against Dependency Parses</i> Dayne Freitag and John Niekrasz . . . . .	36
<i>Inferring Implicit Causal Relationships in Biomedical Literature</i> Halil Kilicoglu . . . . .	46
<i>SnapToGrid: From Statistical to Interpretable Models for Biomedical Information Extraction</i> Marco A. Valenzuela-Escárcega, Gus Hahn-Powell, Dane Bell and Mihai Surdeanu . . . . .	56
<i>Character based String Kernels for Bio-Entity Relation Detection</i> Ritambhara Singh and Yanjun Qi . . . . .	66
<i>Disambiguation of entities in MEDLINE abstracts by combining MeSH terms with knowledge</i> Amy Siu, Patrick Ernst and Gerhard Weikum . . . . .	72
<i>Using Distributed Representations to Disambiguate Biomedical and Clinical Concepts</i> Stephan Tulkens, Simon Suster and Walter Daelemans . . . . .	77
<i>Unsupervised Document Classification with Informed Topic Models</i> Timothy Miller, Dmitriy Dligach and Guergana Savova . . . . .	83
<i>Vocabulary Development To Support Information Extraction of Substance Abuse from Psychiatry Notes</i> Sumithra Velupillai, Danielle L Mowery, Mike Conway, John Hurdle and Brent Kious . . . . .	92
<i>Syntactic analyses and named entity recognition for PubMed and PubMed Central — up-to-the-minute</i> Kai Hakala, Suwisa Kaewphan, Tapio Salakoski and Filip Ginter . . . . .	102
<i>Improving Temporal Relation Extraction with Training Instance Augmentation</i> Chen Lin, Timothy Miller, Dmitriy Dligach, Steven Bethard and Guergana Savova . . . . .	108
<i>Using Centroids of Word Embeddings and Word Mover’s Distance for Biomedical Document Retrieval in Question Answering</i> Georgios-Ioannis Brokos, Prodromos Malakasiotis and Ion Androutsopoulos . . . . .	114
<i>Measuring the State of the Art of Automated Pathway Curation Using Graph Algorithms - A Case Study of the mTOR Pathway</i> Michael Spranger, Sucheendra Palaniappan and Samik Gosh . . . . .	119

<i>Construction of a Personal Experience Tweet Corpus for Health Surveillance</i>	
Keyuan Jiang, Ricardo Calix and Matrika Gupta .....	128
<i>Modelling the Combination of Generic and Target Domain Embeddings in a Convolutional Neural Network for Sentence Classification</i>	
Nut Limsopatham and Nigel Collier .....	136
<i>PubTermVariants: biomedical term variants and their use for PubMed search</i>	
Lana Yeganova, Won Kim, Sun Kim, Rezarta Islamaj Doğan, Wanli Liu, Donald C Comeau, Zhiyong Lu and W John Wilbur .....	141
<i>This before That: Causal Precedence in the Biomedical Domain</i>	
Gus Hahn-Powell, Dane Bell, Marco A. Valenzuela-Escárcega and Mihai Surdeanu .....	146
<i>Syntactic methods for negation detection in radiology reports in Spanish</i>	
Viviana Cotik, Vanesa Stricker, Jorge Vivaldi and Horacio Rodriguez .....	156
<i>How to Train good Word Embeddings for Biomedical NLP</i>	
Billy Chiu, Gamal Crichton, Anna Korhonen and Sampo Pyysalo .....	166
<i>An Information Foraging Approach to Determining the Number of Relevant Features</i>	
Brian Connolly, Benjamin Glass and John Pestian .....	175
<i>Assessing the Feasibility of an Automated Suggestion System for Communicating Critical Findings from Chest Radiology Reports to Referring Physicians</i>	
Brian E. Chapman, Danielle L Mowery, Evan Narasimhan, Neel Patel, Wendy Chapman and Marta Heilbrun .....	181
<i>Building a dictionary of lexical variants for phenotype descriptors</i>	
Simon Kocbek and Tudor Groza .....	186
<i>Applying deep learning on electronic health records in Swedish to predict healthcare-associated infections</i>	
Olof Jacobson and Hercules Dalianis .....	191
<i>Identifying First Episodes of Psychosis in Psychiatric Patient Records using Machine Learning</i>	
Genevieve Gorrell, Sherifat Oduola, Angus Roberts, Tom Craig, Craig Morgan and Rob Stewart	196
<i>Relation extraction from clinical texts using domain invariant convolutional neural network</i>	
Sunil Sahu, Ashish Anand, Krishnadev Oruganty and Mahanandeeshwar Gattu .....	206