

# **Fifth Workshop on Building and Evaluating Resources for Biomedical Text Mining (BioTxtM 2016)**

Osaka, Japan  
11 - 16 December 2016

ISBN: 978-1-5108-3331-9

**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>Cancer Hallmark Text Classification Using Convolutional Neural Networks</i> Simon Baker, Anna Korhonen and Sampo Pyysalo .....	1
<i>Learning Orthographic Features in Bi-directional LSTM for Biomedical Named Entity Recognition</i> Nut Limsopatham and Nigel Collier .....	10
<i>Building Content-driven Entity Networks for Scarce Scientific Literature using Content Information</i> Reinald Kim Amplayo and Min Song .....	20
<i>Named Entity Recognition in Swedish Health Records with Character-Based Deep Bidirectional LSTMs</i> Simon Almgren, Sean Pavlov and Olof Mogren .....	30
<i>Entity-Supported Summarization of Biomedical Abstracts</i> Frederik Schulze and Mariana Neves .....	40
<i>Fully unsupervised low-dimensional representation of adverse drug reaction events through distributional semantics</i> Alicia Pérez, Arantza Casillas and Koldo Gojenola .....	50
<i>A Dataset for ICD-10 Coding of Death Certificates: Creation and Usage</i> Thomas Lavergne, Aurelie Neveol, Aude Robert, Cyril Grouin, Grégoire Rey and Pierre Zweigenbaum .....	60
<i>A Corpus of Tables in Full-Text Biomedical Research Publications</i> Tatyana Shmanina, Ingrid Zukerman, Ai Lee Cheam, Thomas Bochynek and Lawrence Cavedon .....	70
<i>Supervised classification of end-of-lines in clinical text with no manual annotation</i> Pierre Zweigenbaum, Cyril Grouin and Thomas Lavergne .....	80
<i>BioDCA Identifier: A System for Automatic Identification of Discourse Connective and Arguments from Biomedical Text</i> Sindhuja Gopalan and Sobha Lalitha Devi .....	89
<i>Data, tools and resources for mining social media drug chatter</i> Abeed Sarker and Graciela Gonzalez .....	99
<i>Detection of Text Reuse in French Medical Corpora</i> Eva D'hondt, Cyril Grouin, Aurelie Neveol, Efstathios Stamatatos and Pierre Zweigenbaum ..	108
<i>Negation Detection in Clinical Reports Written in German</i> Viviana Cotik, Roland Roller, Feiyu Xu, Hans Uszkoreit, Klemens Budde and Danilo Schmidt	115
<i>Scoring Disease-Medication Associations using Advanced NLP, Machine Learning, and Multiple Content Sources</i> Bharath Dandala, Murthy Devarakonda, Mihaela Bornea and Christopher Nielson .....	125
<i>Author Name Disambiguation in MEDLINE Based on Journal Descriptors and Semantic Types</i> Dina Vishnyakova, Raul Rodriguez-Esteban, Khan Ozol and Fabio Rinaldi .....	134