

# **Clinical Natural Language Processing Workshop (ClinicalNLP 2016)**

Osaka, Japan  
11 December 2016

ISBN: 978-1-5108-3322-7

**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>The impact of simple feature engineering in multilingual medical NER</i> Rebecka Weegar, Arantza Casillas, Arantza Diaz de Ilarraza, Maite Oronoz, Alicia Pérez and Koldo Gojenola.....	1
<i>Bidirectional LSTM-CRF for Clinical Concept Extraction</i> Raghavendra Chalapathy, Ehsan Zare Borzeshi and Massimo Piccardi.....	7
<i>MedNLPDoc: Japanese Shared Task for Clinical NLP</i> Eiji Aramaki, Yoshinobu Kano, Tomoko Ohkuma and Mizuki Morita.....	13
<i>Feature-Augmented Neural Networks for Patient Note De-identification</i> Ji Young Lee, Franck Dernoncourt, Ozlem Uzuner and Peter Szolovits.....	17
<i>Semi-supervised Clustering of Medical Text</i> Pracheta Sahoo, Asif Ekbal, Sriparna Saha, Diego Molla and Kaushik Nandan.....	23
<i>Deep Learning Architecture for Patient Data De-identification in Clinical Records</i> Shweta Yadav, Asif Ekbal, Sriparna Saha and Pushpak Bhattacharyya.....	32
<i>Neural Clinical Paraphrase Generation with Attention</i> Sadid A. Hasan, Bo Liu, Joey Liu, Ashequl Qadir, Kathy Lee, Vivek Datla, Aaditya Prakash and Oladimeji Farri.....	42
<i>Assessing the Corpus Size vs. Similarity Trade-off for Word Embeddings in Clinical NLP</i> Kirk Roberts.....	54
<i>Inference of ICD Codes from Japanese Medical Records by Searching Disease Names</i> Masahito Sakishita and Yoshinobu Kano.....	64
<i>A fine-grained corpus annotation schema of German nephrology records</i> Roland Roller, Hans Uszkoreit, Feiyu Xu, Laura Seiffe, Michael Mikhailov, Oliver Staeck, Klemens Budde, Fabian Halleck and Danilo Schmidt.....	69
<i>Detecting Japanese Patients with Alzheimer’s Disease based on Word Category Frequencies</i> Daisaku Shibata, Shoko Wakamiya, Ayae Kinoshita and Eiji Aramaki.....	78
<i>Prediction of Key Patient Outcome from Sentence and Word of Medical Text Records</i> Takanori Yamashita, Yoshifumi Wakata, Hidehisa Soejima, Naoki Nakashima and Sachio Hirokawa	86
<i>Unsupervised Abbreviation Detection in Clinical Narratives</i> Markus Kreuzthaler, Michel Oleynik, Alexander Avian and Stefan Schulz.....	91
<i>Automated Anonymization as Spelling Variant Detection</i> Steven Kester Yuwono, Hwee Tou Ng and Kee Yuan Ngiam.....	99