

BioNLP 2018

**SIGBioMed Workshop on  
Biomedical Natural Language Processing**

**Proceedings of the 17th BioNLP Workshop**

July 19, 2018  
Melbourne, Australia

©2018 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 978-1-948087-33-9

## **Biomedical natural language processing in 2018: Spotlight on Deep Learning**

*Dina Demner-Fushman, Kevin Bretonnel Cohen, Sophia Ananiadou, and Jun-ichi Tsujii*

The number of community challenges, corpora and publicly available tools in the domain continues to grow rapidly. The past year has seen several hackathons, a variety of shared tasks and growing numbers of workshops dedicated to specific biomedical and clinical sublanguages and tasks. The BioNLP meeting has now been ongoing for 17 years. BioNLP continues to stay the flagship and the generalist in biomedical language processing, accepting noteworthy work independently of the tasks and sublanguages studied. The quality of submissions continues to impress the program committee and the organizers. BioNLP 2018 received 28 submissions, of which 13 were accepted for oral presentation and 12 as poster presentations. This year, Deep Learning approaches are explored in the overwhelming majority of the papers, with focus on interesting new models and in-depth exploration of the state-of-the-art publicly available tools. As for the past several years, the themes in this year's papers and posters continue to focus equally on clinical text and biological language processing, as well as reveal growing interest in consumer language processing. The morning session presents clinical text processing for extraction of causes of death, risk factors identification and named entity recognition, among others. The next session presents work on fundamental NLP problems, such as ontology alignment and key-phrase extraction, whereas the afternoon session presents exceptionally strong work on complex text mining tasks, such as event extraction and question answering.

The invited talk and the invited presentation reflect thus growing interest in automated support for systematic reviews of the literature. In the invited talk, professor Paul Glasziou discusses progress and challenges in automating systematic reviews. Paul Glasziou, FRACGP, PhD is Professor of Evidence-Based Medicine at Bond University and a part-time General Practitioner. He was the Director of the Centre for Evidence-Based Medicine in Oxford from 2003-2010. His key interests include identifying and removing the barriers to using high quality research in everyday clinical practice. He is the author of six books related to evidence based practice: *Systematic Reviews in Health Care*, *Decision Making in Health Care and Medicine: integrating evidence and values*, *An Evidence-Based Medicine Workbook*, *Clinical Thinking: Evidence, Communication and Decision-making*, *Evidence-Based Medicine: How to Practice and Teach EBM*, and *Evidence-Based Medical Monitoring: Principles and Practice*. He has authored over 160 peer-reviewed journal articles and his h-index is currently 43. He is the recipient of an NHRMC Australia Fellowship which he commenced at Bond University in July, 2010.

The invited presentation follows suit by bringing to our attention a new corpus of about 5,000 abstracts of randomized control trials annotated with granular information regarding the study populations, interventions, comparators and outcomes.



**Organizers:**

Kevin Bretonnel Cohen, University of Colorado School of Medicine, USA  
Dina Demner-Fushman, US National Library of Medicine  
Sophia Ananiadou, National Centre for Text Mining and University of Manchester, UK  
Jun-ichi Tsujii, National Institute of Advanced Industrial Science and Technology, Japan

**Program Committee:**

Sophia Ananiadou, National Centre for Text Mining and University of Manchester, UK  
Emilia Apostolova, Language.ai, USA  
Eiji Aramaki, University of Tokyo, Japan  
Asma Ben Abacha, US National Library of Medicine  
Olivier Bodenreider, US National Library of Medicine  
Leonardo Campillos Llanos, LIMSI - CNRS, France  
Brian Connolly, Kroger Digital, USA  
Dina Demner-Fushman, US National Library of Medicine  
Filip Ginter, University of Turku, Finland  
Graciela Gonzalez-Hernandez, University of Pennsylvania, USA  
Travis Goodwin, The University of Texas at Dallas, USA  
Cyril Grouin, LIMSI - CNRS, France  
Tudor Groza, The Garvan Institute of Medical Research, Australia  
Antonio Jimeno Yepes, IBM, Melbourne Area, Australia  
Halil Kilicoglu, US National Library of Medicine  
Robert Leaman, US National Library of Medicine  
Ulf Leser, Humboldt-Universität zu Berlin, Germany  
Zhiyong Lu, US National Library of Medicine  
Timothy Miller, Children's Hospital Boston, USA  
Makoto Miwa, Toyota Technological Institute, Japan  
Danielle L Mowery, VA Salt Lake City Health Care System, USA  
Yassine M'Rabet, US National Library of Medicine  
Aurelie Neveol, LIMSI - CNRS, France  
Claire Nedellec, INRA, France  
Mariana Neves, Hasso Plattner Institute and University of Potsdam, Germany  
Nhung Nguyen, The University of Manchester, UK  
Naoaki Okazaki, Tohoku University, Japan  
Sampo Pyysalo, University of Cambridge, UK  
Francisco J. Ribadas-Pena, University of Vigo, Spain  
Fabio Rinaldi, University of Zurich, Switzerland  
Kirk Roberts, The University of Texas Health Science Center at Houston, USA  
Angus Roberts, The University of Sheffield, UK  
Hagit Shatkay, University of Delaware, USA  
Pontus Stenetorp, University College London, UK  
Karin Verspoor, The University of Melbourne, Australia  
Byron C. Wallace, University of Texas at Austin, USA  
W John Wilbur, US National Library of Medicine  
Pierre Zweigenbaum, LIMSI - CNRS, France

**Additional Reviewers:**

Pramod Chandrashekar, University of Pennsylvania, USA  
Nicolas Fiorini, US National Library of Medicine  
Arjun Magge, University of Pennsylvania, USA  
Yijia Zhang, US National Library of Medicine

**Invited Speaker:**

Paul Glasziou, Bond University, Australia

## Table of Contents

<i>Embedding Transfer for Low-Resource Medical Named Entity Recognition: A Case Study on Patient Mobility</i>	
Denis Newman-Griffis and Ayah Zirikly .....	1
<i>Multi-task learning for interpretable cause of death classification using key phrase prediction</i>	
Serena Jeblee, Mireille Gomes and Graeme Hirst .....	12
<i>Identifying Risk Factors For Heart Disease in Electronic Medical Records: A Deep Learning Approach</i>	
Thanat Chokwijitkul, Anthony Nguyen, Hamed Hassanzadeh and Siegfried Perez .....	18
<i>Keyphrases Extraction from User-Generated Contents in Healthcare Domain Using Long Short-Term Memory Networks</i>	
Ilham Fathy Saputra, Rahmad Mahendra and Alfian Farizki Wicaksono .....	28
<i>Identifying Key Sentences for Precision Oncology Using Semi-Supervised Learning</i>	
Jurica Ševa, Martin Wackerbauer and Ulf Leser .....	35
<i>Ontology alignment in the biomedical domain using entity definitions and context</i>	
Lucy Wang, Chandra Bhagavatula, Mark Neumann, Kyle Lo, Chris Wilhelm and Waleed Ammar	47
<i>Sub-word information in pre-trained biomedical word representations: evaluation and hyper-parameter optimization</i>	
Dieter Galea, Ivan Laponogov and Kirill Veselkov .....	56
<i>PICO Element Detection in Medical Text via Long Short-Term Memory Neural Networks</i>	
Di Jin and Peter Szolovits .....	67
<i>Coding Structures and Actions with the COSTA Scheme in Medical Conversations</i>	
Nan Wang, Yan Song and Fei Xia .....	76
<i>A Neural Autoencoder Approach for Document Ranking and Query Refinement in Pharmacogenomic Information Retrieval</i>	
Jonas Pfeiffer, Samuel Broscheit, Rainer Gemulla and Mathias Göschl .....	87
<i>Biomedical Event Extraction Using Convolutional Neural Networks and Dependency Parsing</i>	
Jari Björne and Tapio Salakoski .....	98
<i>BioAMA: Towards an End to End BioMedical Question Answering System</i>	
Vasu Sharma, Nitish Kulkarni, Srividya Pranavi, Gabriel Bayomi, Eric Nyberg and Teruko Mitamura .....	109
<i>Phrase2VecGLM: Neural generalized language model-based semantic tagging for complex query reformulation in medical IR</i>	
Manirupa Das, Eric Fosler-Lussier, Simon Lin, Soheil Moosavinasab, David Chen, Steve Rust, Yungui Huang and Rajiv Ramnath .....	118
<i>Convolutional neural networks for chemical-disease relation extraction are improved with character-based word embeddings</i>	
Dat Quoc Nguyen and Karin Verspoor .....	129

<i>Domain Adaptation for Disease Phrase Matching with Adversarial Networks</i>	
Miaofeng Liu, Jialong Han, Haisong Zhang and Yan Song .....	137
<i>Predicting Discharge Disposition Using Patient Complaint Notes in Electronic Medical Records</i>	
Mohamad Salimi and Alla Rozovskaya .....	142
<i>Bacteria and Biotope Entity Recognition Using A Dictionary-Enhanced Neural Network Model</i>	
Qiuyue Wang and Xiaofeng Meng .....	147
<i>SingleCite: Towards an improved Single Citation Search in PubMed</i>	
Lana Yeganova, Donald C Comeau, Won Kim, W John Wilbur and Zhiyong Lu .....	151
<i>A Framework for Developing and Evaluating Word Embeddings of Drug-named Entity</i>	
Mengnan Zhao, Aaron J. Masino and Christopher C. Yang .....	156
<i>MeSH-based dataset for measuring the relevance of text retrieval</i>	
Won Gyu KIM, Lana Yeganova, Donald comeau, W John Wilbur and Zhiyong Lu .....	161
<i>CRF-LSTM Text Mining Method Unveiling the Pharmacological Mechanism of Off-target Side Effect of Anti-Multiple Myeloma Drugs</i>	
Kaiyin Zhou, Sheng Zhang, Xiangyu Meng, Qi Luo, Yuxing Wang, Ke Ding, Yukun Feng, Mo Chen, Kevin Cohen and Jingbo Xia .....	166
<i>Prediction Models for Risk of Type-2 Diabetes Using Health Claims</i>	
Masatoshi Nagata, Kohichi Takai, Keiji Yasuda, Panikos Heracleous and Akio Yoneyama .....	172
<i>On Learning Better Embeddings from Chinese Clinical Records: Study on Combining In-Domain and Out-Domain Data</i>	
Yaqiang Wang, Yunhui Chen, Hongping Shu and Yongguang Jiang .....	177
<i>Investigating Domain-Specific Information for Neural Coreference Resolution on Biomedical Texts</i>	
Long Trieu, Nhung Nguyen, Makoto Miwa and Sophia Ananiadou .....	183
<i>Toward Cross-Domain Engagement Analysis in Medical Notes</i>	
Sara Rosenthal and Adam Faulkner .....	189

# Conference Program

**Thursday July 19, 2018**

**9:00–9:15**      **Opening remarks**

**9:15–10:30**    **Session 1: Clinical NLP**

9:15–9:30      *Embedding Transfer for Low-Resource Medical Named Entity Recognition: A Case Study on Patient Mobility*

Denis Newman-Griffis and Ayah Zirikly

9:30–9:45      *Multi-task learning for interpretable cause of death classification using key phrase prediction*

Serena Jeblee, Mireille Gomes and Graeme Hirst

9:45–10:00    *Identifying Risk Factors For Heart Disease in Electronic Medical Records: A Deep Learning Approach*

Thanat Chokwijitkul, Anthony Nguyen, Hamed Hassanzadeh and Siegfried Perez

10:00–10:15    *Keyphrases Extraction from User-Generated Contents in Healthcare Domain Using Long Short-Term Memory Networks*

Ilham Fathy Saputra, Rahmad Mahendra and Alfan Farizki Wicaksono

10:15–10:30    *Identifying Key Sentences for Precision Oncology Using Semi-Supervised Learning*

Jurica Ševa, Martin Wackerbauer and Ulf Leser

**10:30–11:00**    *Coffee Break*

**Thursday July 19, 2018 (continued)**

**11:00–12:30 Session 2: Foundations**

11:00–11:15 *Ontology alignment in the biomedical domain using entity definitions and context*  
Lucy Wang, Chandra Bhagavatula, Mark Neumann, Kyle Lo, Chris Wilhelm and Waleed Ammar

11:15–11:30 *Sub-word information in pre-trained biomedical word representations: evaluation and hyper-parameter optimization*  
Dieter Galea, Ivan Laponogov and Kirill Veselkov

11:30–11:45 *PICO Element Detection in Medical Text via Long Short-Term Memory Neural Networks*  
Di Jin and Peter Szolovits

11:45–12:00 *Coding Structures and Actions with the COSTA Scheme in Medical Conversations*  
Nan Wang, Yan Song and Fei Xia

**12:00–13:30 Lunch break**

**13:30–14:30 Invited Talk: "Automating systematic reviews: progress and challenges" – Paul Glasziou**

**14:30–15:30 Session 3 Literature mining and retrieval; Question Answering**

14:30–14:45 *A Neural Autoencoder Approach for Document Ranking and Query Refinement in Pharmacogenomic Information Retrieval*  
Jonas Pfeiffer, Samuel Broscheit, Rainer Gemulla and Mathias Göschl

14:45–15:00 *Biomedical Event Extraction Using Convolutional Neural Networks and Dependency Parsing*  
Jari Björne and Tapio Salakoski

15:00–15:15 *BioAMA: Towards an End to End BioMedical Question Answering System*  
Vasu Sharma, Nitish Kulkarni, Srividya Pranavi, Gabriel Bayomi, Eric Nyberg and Teruko Mitamura

15:15–15:30 *Phrase2VecGLM: Neural generalized language model-based semantic tagging for complex query reformulation in medical IR*  
Manirupa Das, Eric Fosler-Lussier, Simon Lin, Soheil Moosavinasab, David Chen, Steve Rust, Yungui Huang and Rajiv Ramnath

Thursday July 19, 2018 (continued)

15:30–16:00 *Coffee Break*

16:00–16:15 **Invited Presentation: "A Corpus with Multi-Level Annotations of Patients, Interventions and Outcomes to Support Language Processing for Medical Literature" – Ben Nye**

16:15–18:00 **Poster Session**

*Convolutional neural networks for chemical-disease relation extraction are improved with character-based word embeddings*

Dat Quoc Nguyen and Karin Verspoor

*Domain Adaptation for Disease Phrase Matching with Adversarial Networks*

Miaofeng Liu, Jialong Han, Haisong Zhang and Yan Song

*Predicting Discharge Disposition Using Patient Complaint Notes in Electronic Medical Records*

Mohamad Salimi and Alla Rozovskaya

*Bacteria and Biotope Entity Recognition Using A Dictionary-Enhanced Neural Network Model*

Qiuyue Wang and Xiaofeng Meng

*SingleCite: Towards an improved Single Citation Search in PubMed*

Lana Yeganova, Donald C Comeau, Won Kim, W John Wilbur and Zhiyong Lu

*A Framework for Developing and Evaluating Word Embeddings of Drug-named Entity*

Mengnan Zhao, Aaron J. Masino and Christopher C. Yang

*MeSH-based dataset for measuring the relevance of text retrieval*

Won Gyu KIM, Lana Yeganova, Donald comeau, W John Wilbur and Zhiyong Lu

*CRF-LSTM Text Mining Method Unveiling the Pharmacological Mechanism of Off-target Side Effect of Anti-Multiple Myeloma Drugs*

Kaiyin Zhou, Sheng Zhang, Xiangyu Meng, Qi Luo, Yuxing Wang, Ke Ding, Yukun Feng, Mo Chen, Kevin Cohen and Jingbo Xia

*Prediction Models for Risk of Type-2 Diabetes Using Health Claims*

Masatoshi Nagata, Kohichi Takai, Keiji Yasuda, Panikos Heracleous and Akio Yoneyama

**Thursday July 19, 2018 (continued)**

*On Learning Better Embeddings from Chinese Clinical Records: Study on Combining In-Domain and Out-Domain Data*

Yaqiang Wang, Yunhui Chen, Hongping Shu and Yongguang Jiang

*Investigating Domain-Specific Information for Neural Coreference Resolution on Biomedical Texts*

Long Trieu, Nhung Nguyen, Makoto Miwa and Sophia Ananiadou

*Toward Cross-Domain Engagement Analysis in Medical Notes*

Sara Rosenthal and Adam Faulkner