EMNLP 2016

Seventh International Workshop on Health Text Mining and Information Analysis (LOUHI)

Proceedings of the Workshop

November 5, 2016
Austin, Texas, USA

Order copies of this and other ACL proceedings from:

Curran Associates
57 Morehouse Lane
Red Hook, New York 12571
USA
Tel: +1-845-758-0400
Fax: +1-845-758-2633
curran@proceedings.com

ISBN 978-1-945626-33-3

Preface

The Seventh International Workshop on Health Text Mining and Information Analysis provides an interdisciplinary forum for researchers interested in automated processing of health documents. Health documents encompass electronic health records, clinical guidelines, spontaneous reports for pharmacovigilance, biomedical literature, health forums/blogs or any other type of health-related documents. The LOUHI workshop series fosters interactions between the Computational Linguistics, Medical Informatics and Artificial Intelligence communities. Following the six previous edition of the workshop which were co-located with SMBM 2008 in Turku, Finland, with NAACL 2010 in Los Angeles, California, with Artificial Intelligence in Medicine (AIME 2011) in Bled, Slovenia, during NICTA Techfest 2013 in Sydney, Australia, co-located with EACL 2014 in Gothenburg, Sweden, and with EMNLP 2015 in Lisbon, Portugal, this workshop is co-located this year with EMNLP 2016 in Austin, Texas.

The aim of the LOUHI 2016 workshop is to bring together research work on topics related to health documents, particularly emphasizing multidisciplinary aspects of health documentation and the interplay between nursing and medical sciences, information systems, computational linguistics and computer science. The topics include, but are not limited to, the following Natural Language Processing techniques and related areas:

- Techniques supporting information extraction, e.g. named entity recognition, negation and uncertainty detection
- Classification and text mining applications (e.g. diagnostic classifications such as ICD-10 and nursing intensity scores) and problems (e.g. handling of unbalanced data sets)
- Text representation, including dealing with data sparsity and dimensionality issues
- Domain adaptation, e.g. adaptation of standard NLP tools (incl. tokenizers, PoS-taggers, etc) to the medical domain
- Information fusion, i.e. integrating data from various sources, e.g. structured and narrative documentation
- Unsupervised methods, including distributional semantics
- Evaluation, gold/reference standard construction and annotation
- Syntactic, semantic and pragmatic analysis of health documents
- Anonymization/de-identification of health records and ethics
- Supporting the development of medical terminologies and ontologies
- Individualization of content, consumer health vocabularies, summarization and simplification of text
- NLP for supporting documentation and decision making practices

- Predictive modeling of adverse events, e.g. adverse drug events and hospital acquired infections

The call for papers encouraged authors to submit papers describing substantial and completed work but also focus on a contribution, a negative result, a software package or work in progress. We also encouraged to report work on low-resourced languages, addressing the challenges of data sparsity and language characteristic diversity.

We received 21 submissions. Each submission went through a double-blind review process which involved three program committee members. Based on comments and rankings supplied by the reviewers, we accepted 13 papers. The overall acceptance rate is 62%. During the workshop, 6 papers will be presented orally, and 7 papers will be presented as posters.

Our special thanks go to Nigel Collier for accepting to give an invited talk.

Finally, we would like to thank the members of the program committee for the quality of theirs reviews in a very short period, and the authors for their submissions and the quality of their work.

Cyril Grouin, Thierry Hamon, Aurélie Névéol, Pierre Zweigenbaum.

Organizers:

Cyril Grouin, CNRS, LIMSI (France)
Thierry Hamon, Université Paris 13, CNRS, LIMSI (France)
Aurélie Névéol, CNRS, LIMSI (France)
Pierre Zweigenbaum, CNRS, LIMSI (France)

Program Committee:

Sophia Ananiadou, University of Manchester (UK)
Sabine Bergler, Concordia University (Canada)
Thomas Brox Røst, Norwegian University of Science and Technology (Norway)
Kevin B Cohen, University of Colorado/School of Medicine (USA)
Hercules Dalianis, Stockholm University (Sweden)
Louise Deléger, INRA (France)
Filip Ginter, University of Turku (Finland)
Natalia Grabar, CNRS UMR 8163, STL Université de Lille 3 (France)
Gintaré Grigonyté, Stockholm University (Sweden)
Aron Henriksson, Stockholm University (Sweden)
Antonio Jimeno Yepes, IBM Research (Australia)
Jussi Karlgren, KTH, Royal Institute of Technology (Sweden)
Dimitrios Kokkinakis, University of Gothenburg (Sweden)
Maria Kvist, Stockholm University (Sweden)
Alberto Lavelli, Fondazione Bruno Kessler (Italy)
David Martinez, University of Melbourne and MedWhat.com (Australia)
Beáta Megyesi, Uppsala University (Sweden)
Marie-Jean Meurs, UQAM & Concordia University (Canada)
Fleur Mougin, Université de Bordeaux, ERIAS, Centre INSERM U897, ISPED (France)
Danielle L Mowery, University of Utah (USA)
Henning Müller, University of Applied Sciences Western Switzerland (Switzerland)
Mariana Lara Neves, Hasso-Plattner-Institute at the University of Potsdam (Germany)
Jong C. Park, KAIST Computer Science (Korea)
Rezarta Islamaj-Dogan, NIH/NLM/NCBI (USA)
Tapio Salakoski, University of Turku (Finland)
Stefan Schulz, Graz General Hospital and University Clinics (Austria)
Isabel Segura-Bedmar, Universidad Carlos III de Madrid (Spain)
Maria Skeppstedt, Stockholm University (Sweden)
Hanna Suominen, NICTA (Australia)
Suzanne Tamang, Stanford University School of Medicine (USA)
Özlem Uzuner, MIT (USA)
Sumithra Velupillai, Stockholm University (Sweden)
Karin Verspoor, University of Melbourne (Australia)
Mats Wirén, Stockholm University, Stockholm (Sweden)

Invited Speaker:

Nigel Collier, University of Cambridge, UK

Table of Contents

An Investigation of Recurrent Neural Architectures for Drug Name Recognition
Raghavendra Chalapathy, Ehsan Zare Borzeshi and Massimo Piccardi 1

Clinical Text Prediction with Numerically Grounded Conditional Language Models
Georgios Spithourakis, Steffen Petersen and Sebastian Riedel 6

Modelling Radiological Language with Bidirectional Long Short-Term Memory Networks
Savelie Cornegruta, Robert Bakewell, Samuel Withey and Giovanni Montana 17

Data Resource Acquisition from People at Various Stages of Cognitive Decline – Design and Exploration Considerations
Dimitrios Kokkinakis, Kristina Lundholm Fors and Arto Nordlund 28

Analysis of Anxious Word Usage on Online Health Forums
Nicolas Rey-Villamizar, Prasha Shrestha, Farig Sadeque, Steven Bethard, Ted Pedersen, Arjun Mukherjee and Thamar Solorio 37

Retrofitting Word Vectors of MeSH Terms to Improve Semantic Similarity Measures
Zhiguo Yu, Trevor Cohen, Byron Wallace, Elmer Bernstam and Todd Johnson 43

Unsupervised Resolution of Acronyms and Abbreviations in Nursing Notes Using Document-Level Context Models
Katrin Kirchhoff and Anne M. Turner 52

Low-resource OCR error detection and correction in French Clinical Texts
Eva D'hondt, Cyril Grouin and Brigitte Grau 61

Citation Analysis with Neural Attention Models
Tsendsuren Munkhdalai, John Lalor and Hong Yu 69

Replicability of Research in Biomedical Natural Language Processing: a pilot evaluation for a coding task
Aurelie Neveol, Kevin Cohen, Cyril Grouin and Aude Robert 78

NLP and Online Health Reports: What do we say and what do we mean?
Nigel Collier 85

Leveraging coreference to identify arms in medical abstracts: An experimental study
Elisa Ferracane, Iain Marshall, Byron C. Wallace and Katrin Erk 86

Hybrid methods for ICD-10 coding of death certificates
Pierre Zweigenbaum and Thomas Lavergne 96

Exploring Query Expansion for Entity Searches in PubMed
Chung-Chi Huang and Zhiyong Lu 106

Workshop Program

Saturday, November 5, 2016

09:00–10:15 Session I - Machine-Learning

09:00–09:25 *An Investigation of Recurrent Neural Architectures for Drug Name Recognition*
Raghavendra Chalapathy, Ehsan Zare Borzeshi and Massimo Piccardi

09:25–09:50 *Clinical Text Prediction with Numerically Grounded Conditional Language Models*
Georgios Spithourakis, Steffen Petersen and Sebastian Riedel

09:50–10:15 *Modelling Radiological Language with Bidirectional Long Short-Term Memory Networks*
Savelie Cornegruta, Robert Bakewell, Samuel Withey and Giovanni Montana

10:15–10:30 Session II - Boosters

10:30–11:00 ***Coffee Break***

11:00–12:30 Session III - Posters

Data Resource Acquisition from People at Various Stages of Cognitive Decline – Design and Exploration Considerations
Dimitrios Kokkinakis, Kristina Lundholm Fors and Arto Nordlund

Analysis of Anxious Word Usage on Online Health Forums
Nicolas Rey-Villamizar, Prasha Shrestha, Farig Sadeque, Steven Bethard, Ted Pedersen, Arjun Mukherjee and Thamar Solorio

Retrofitting Word Vectors of MeSH Terms to Improve Semantic Similarity Measures
Zhiguo Yu, Trevor Cohen, Byron Wallace, Elmer Bernstam and Todd Johnson

Unsupervised Resolution of Acronyms and Abbreviations in Nursing Notes Using Document-Level Context Models
Katrin Kirchhoff and Anne M. Turner

Low-resource OCR error detection and correction in French Clinical Texts
Eva D’hondt, Cyril Grouin and Brigitte Grau

Saturday, November 5, 2016 (continued)

	Citation Analysis with Neural Attention Models Tsendsuren Munkhdalai, John Lalor and Hong Yu
	Replicability of Research in Biomedical Natural Language Processing: a pilot evaluation for a coding task Aurelie Neveol, Kevin Cohen, Cyril Grouin and Aude Robert
12:30–14:00	***Lunch break***
14:00–15:30	**Session IV - Invited talk**
14:00–15:30	*NLP and Online Health Reports: What do we say and what do we mean?* Nigel Collier
15:30–16:00	***Coffee Break***
16:00–17:15	**Session V - NLP for literature and clinical documents**
16:00–16:25	*Leveraging coreference to identify arms in medical abstracts: An experimental study* Elisa Ferracane, Iain Marshall, Byron C. Wallace and Katrin Erk
16:25–16:50	*Hybrid methods for ICD-10 coding of death certificates* Pierre Zweigenbaum and Thomas Lavergne
16:50–17:15	*Exploring Query Expansion for Entity Searches in PubMed* Chung-Chi Huang and Zhiyong Lu