EMNLP 2018

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

Proceedings of the Workshop

31 October, 2018
Brussels, Belgium

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

ISBN 978-1-948087-74-2

Introduction

The International Workshop on Health Text Mining and Information Analysis (LOUHI) 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. The eight previous editions of the workshop 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, with EMNLP 2015 in Lisbon, Portugal, with EMNLP 2016 in Austin, Texas; and in 2017 was held in Sydney, Australia. This year the workshop is co-located with EMNLP 2018 in Brussels, Belgium.

The aim of the LOUHI 2018 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.

This year we received a high number of submissions (49), therefore the selection process was very competitive. Due to time and space limitations, we could only choose a small number of the submitted papers to appear in the program.

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 23 papers. Although the selection was entirely based on the scores provided by the reviewers, we regretfully had to set a relatively high threshold for acceptance. The overall acceptance rate is 46%. During the workshop, 13 papers will be presented orally, and 10 papers will be presented as posters.

Our special thanks go to Goran Nenadic for accepting to give an invited talk.

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

Organizers:

Alberto Lavelli, FBK, Trento, Italy
Anne-Lyse Minard, IRISA, CNRS, Rennes, France
Fabio Rinaldi, University of Zurich, Switzerland & FBK, Trento, Italy

Program Committee:

Sophia Ananiadou, University of Manchester, UK
Georgeta Bordea, Université de Bordeaux, France
Leonardo Campillos Llanos, LIMSI, CNRS, France
Wendy Chapman, University of Utah, USA
Vincent Claveau, IRISA, CNRS, France
Kevin B Cohen, University of Colorado/School of Medicine, USA
Francisco Couto, University of Lisbon, Portugal
Hercules Dalianis, Stockholm University, Sweden
Martin Duneld, Stockholm University, Sweden
Filip Ginter, University of Turku, Finland
Natalia Grabar, CNRS UMR 8163, STL Université de Lille3, France
Gintaré Grigonyté, Stockholm University, Sweden
Cyril Grouin, LIMSI, CNRS, Université Paris-Saclay, Orsay, France
Thierry Hamon, LIMSI, CNRS, Université Paris-Saclay, Orsay, France & Université Paris 13, Villetaneuse, France
Aron Henriksson, Stockholm University, Sweden
Rezarta Islamaj-Dogan, NIH/NLM/NCBI, USA
Antonio Jimeno Yepes, IBM Research, Australia
Yoshinobu Kano, Shizuoka University, Japan
Jin-Dong Kim, Research Organization of Information and Systems, Japan
Dimitrios Kokkinakis, University of Gothenburg, Sweden
Martin Krallinger, Spanish National Cancer Research Centre (CNIO)
Michael Krauthammer, Yale University, USA
Ivano Lauriola, University of Padova and FBK, Trento, Italy
Analia Lourenco, Universidade de Vigo, Spain
David Martinez, University of Melbourne and MedWhat.com, Australia
Sérgio Matos, University of Aveiro, Portugal
Marie-Jean Meurs, UQAM & Concordia University, QC, Canada
Timothy Miller, Harvard Medical School, USA
Hans Moen, University of Turku
Diego Molla, Maquaire University, Australia
Roser Morante, VU Amsterdam, Netherlands
Danielle L Mowery, University of Utah, USA
Henning Müller, University of Applied Sciences Western Switzerland, Switzerland
Goran Nenadic, University of Manchester, UK
Aurélie Névéol, LIMSI, CNRS, Université Paris-Saclay, Orsay, France
Mariana Lara Neves, German Federal Institute for Risk Assessment, Germany
Richard Nock, CSIRO, Australia
Øystein Nytrø, NTNU, Norway

Naoaki Okazaki, Tokyo Institute of Technology, Japan
Jong C. Park, KAIST Computer Science, Korea
Thomas Brox Røst, Norwegian University of Science and Technology, Norway
Patrick Ruch, SIB Swiss Institute of Bioinformatics, Switzerland
Tapio Salakoski, University of Turku, Finland
Sanna Salanterä, University of Turku, Finland
Stefan Schulz, Graz General Hospital and University Clinics, Austria
Isabel Segura-Bedmar, Universidad Carlos III de Madrid, Spain
Maria Skeppstedt, Linneus University, Sweden, and Potsdam University, Germany
Manfred Stede, University of Potsdam, Germany
Hanna Suominen, CSIRO, Australia
Sumithra Velupillai, KTH, Royal Institute of Technology, Sweden, and King's College London, UK
Özlem Uzuner, MIT, USA
Pierre Zweigenbaum, LIMSI, CNRS, Université Paris-Saclay, Orsay, France

Invited Speaker:

Goran Nenadic, University of Manchester, UK

Table of Contents

Workshop Program

October 31, 2018

9:00–10:30 Session 1

9:00 ***Introduction***

9:05 *Detecting Diabetes Risk from Social Media Activity*
Dane Bell, Egoitz Laparra, Aditya Kousik, Terron Ishihara, Mihai Surdeanu and Stephen Kobourov

9:30 *Treatment Side Effect Prediction from Online User-generated Content*
Hoang Nguyen, Kazunari Sugiyama, Min-Yen Kan and Kishaloy Halder

9:55 ***Poster booster***

10:15 ***Poster session***

Revisiting neural relation classification in clinical notes with external information
Simon Suster, Madhumita Sushil and Walter Daelemans

Supervised Machine Learning for Extractive Query Based Summarisation of Biomedical Data
Mandeep Kaur and Diego Molla

Comparing CNN and LSTM character-level embeddings in BiLSTM-CRF models for chemical and disease named entity recognition
Zenan Zhai, Dat Quoc Nguyen and Karin Verspoor

Deep learning for language understanding of mental health concepts derived from Cognitive Behavioural Therapy
Lina M. Rojas Barahona, Bo-Hsiang Tseng, Yinpei Dai, Clare Mansfield, Osman Ramadan, Stefan Ultes, Michael Crawford and Milica Gasic

Investigating the Challenges of Temporal Relation Extraction from Clinical Text
Diana Galvan, Naoaki Okazaki, Koji Matsuda and Kentaro Inui

De-identifying Free Text of Japanese Dummy Electronic Health Records
Kohei Kajiyama, Hiromasa Horiguchi, Takashi Okumura, Mizuki Morita and Yoshinobu Kano

October 31, 2018 (continued)

Unsupervised Identification of Study Descriptors in Toxicology Research: An Experimental Study
Drahomira Herrmannova, Steven Young, Robert Patton, Christopher Stahl, Nicole Kleinstreuer and Mary Wolfe

Identification of Parallel Sentences in Comparable Monolingual Corpora from Different Registers
Rémi Cardon and Natalia Grabar

Evaluation of a Prototype System that Automatically Assigns Subject Headings to Nursing Narratives Using Recurrent Neural Network
Hans Moen, Kai Hakala, Laura-Maria Peltonen, Henry Suhonen, Petri Loukasmäki, Tapio Salakoski, Filip Ginter and Sanna Salanterä

Automatically Detecting the Position and Type of Psychiatric Evaluation Report Sections
Deya Banisakher, Naphtali Rishe and Mark A. Finlayson

10:30–11:00 Break

11:00–12:30 Session 2

11:00 *Iterative development of family history annotation guidelines using a synthetic corpus of clinical text*
Taraka Rama, Pål Brekke, Øystein Nytrø and Lilja Øvrelid

11:25 *CAS: French Corpus with Clinical Cases*
Natalia Grabar, Vincent Claveau and Clément Dalloux

11:40 *Analysis of Risk Factor Domains in Psychosis Patient Health Records*
Eben Holderness, Nicholas Miller, Kirsten Bolton, Philip Cawkwell, Marie Meteer, James Pustejovsky and Mei Hua-Hall

12:05 *Patient Risk Assessment and Warning Symptom Detection Using Deep Attention-Based Neural Networks*
Ivan Girardi, Pengfei Ji, An-phi Nguyen, Nora Hollenstein, Adam Ivankay, Lorenz Kuhn, Chiara Marchiori and Ce Zhang

October 31, 2018 (continued)

12:30–14:00 Lunch

14:00–15:30 Session 3

14:00 *Invited Talk - Distributed text mining in healthcare: linking data, methods and people*
Goran Nenadic

14:50 *Syntax-based Transfer Learning for the Task of Biomedical Relation Extraction*
Joël Legrand, Yannick Toussaint, Chedy Raïssi and Adrien Coulet

15:15 *In-domain Context-aware Token Embeddings Improve Biomedical Named Entity Recognition*
Golnar Sheikhshabbafghi, Inanc Birol and Anoop Sarkar

15:30–16:00 Break

16:00–17:30 Session 4

16:00 *Self-training improves Recurrent Neural Networks performance for Temporal Relation Extraction*
Chen Lin, Timothy Miller, Dmitriy Dligach, Hadi Amiri, Steven Bethard and Guergana Savova

16:25 *Listwise temporal ordering of events in clinical notes*
Serena Jeblee and Graeme Hirst

16:40 *Time Expressions in Mental Health Records for Symptom Onset Extraction*
Natalia Viani, Lucia Yin, Joyce Kam, Ayunni Alawi, André Bittar, Rina Dutta, Rashmi Patel, Robert Stewart and Sumithra Velupillai

16:55 *Evaluation of a Sequence Tagging Tool for Biomedical Texts*
Julien Tourille, Matthieu Doutreligne, Olivier Ferret, Aurélie Névéol, Nicolas Paris and Xavier Tannier

17:10 *Learning to Summarize Radiology Findings*
Yuhao Zhang, Daisy Yi Ding, Tianpei Qian, Christopher D. Manning and Curtis P. Langlotz