EMNLP 2020

The 11th International Workshop on Health Text Mining and Information Analysis LOUHI 2020

Proceedings of the Workshop

November 20, 2020

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 978-1-952148-81-1

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 10 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; in 2017 was held in Sydney, Australia; in 2018 was co-located with EMNLP 2018 in Brussels, Belgium; and in 2019 was co-located with EMNLP 2019 in Hong Kong. This year the workshop is co-located with EMNLP 2020 and takes place online due to the COVID-19 pandemics.

The aim of the LOUHI 2020 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
- Terminology and information model standards (SNOMED CT, FHIR) for health text mining

- Bridging gaps between formal ontology and biomedical NLP

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 (43), 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 16 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 37%.

Our special thanks go to Guergana Savova 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:

Eben Holderness, Brandeis University, USA
Antonio Jimeno Yepes, IBM Research, Australia
Alberto Lavelli, FBK, Trento, Italy
Anne-Lyse Minard, University of Orleans, France
James Pustejovsky, Brandeis University, USA
Fabio Rinaldi, Dalle Molle Institute for Artificial Intelligence Research - IDSIA, University of Zurich, Switzerland & FBK, Trento, Italy

Program Committee:

Mohammad Akbari, National University of Singapore, Singapore
Rafael Berlanga Llavori, Universitat Jaume I, Spain
Georgeta Bordea, Université de Bordeaux, France
Leonardo Campillos Llanos, LIMSI, CNRS, France
Francisco Couto, University of Lisbon, Portugal
Hercules Dalianis, Stockholm University, Sweden
Kerstin Denecke, Bern University of Applied Sciences, Switzerland
Martin Duneld (xmartin@dsv.su.se), Stockholm University, Sweden
Natalia Grabar, CNRS UMR 8163, STL Université de Lille3, France
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
Eben Holderness, Brandeis University, USA
Rezarta Islamaj-Dogan, NIH/NLM/NCBI, USA
Hyeju Jang, University of British Columbia, Canada
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), Spain
Alberto Lavelli, FBK, Trento, Italy
Analia Lourenco, Universidade de Vigo, Spain
David Martinez, University of Melbourne and MedWhat.com, Australia
Sérgio Matos, University of Aveiro, Portugal
Timothy Miller, Harvard Medical School, USA
Anne-Lyse Minard, University of Orleans, France
Hans Moen, University of Turku, Finland
Roser Morante, VU Amsterdam, Netherlands
Danielle L Mowery, University of Utah, USA
Henning Müller, University of Applied Sciences Western Switzerland, Switzerland
Aakanksha Naik, CMU, USA
Aurélie Névéol, LIMSI, CNRS, Université Paris-Saclay, Orsay, France
Mariana Lara Neves, German Federal Institute for Risk Assessment, Germany
Jong C. Park, KAIST Computer Science, Korea
Laura Plaza, Universidad Complutense de Madrid, Spain

James Pustejovsky, Brandeis University, USA
Fabio Rinaldi, Dalle Molle Institute for Artificial Intelligence Research - IDSIA, University of Zurich, Switzerland & FBK, Trento, Italy
Thomas Brox Røst, Norwegian University of Science and Technology, Norway
Tapio Salakoski, University of Turku, Finland
Stefan Schulz, Graz General Hospital and University Clinics, Austria
Maria Skeppstedt, Linneus University, Sweden, and Potsdam University, Germany
Amber Stubbs, Simmons College, USA
Hanna Suominen, Australian National University, Australia
Suzanne Tamang, Stanford University School of Medicine, USA
Sumithra Velupillai, KTH, Royal Institute of Technology, Sweden, and King's College London, UK
Özlem Uzuner, MIT, USA
Yanshan Wang, Mayo Clinic, USA
Pierre Zweigenbaum, LIMSI, CNRS, Université Paris-Saclay, Orsay, France

Additional Reviewers:

Marcia Barros, University of Lisbon, Portugal
Jari Björne, University of Turku, Finland
Juho Heimonen, University of Turku, Finland
Huije Lee, KAIST Computer Science, Korea
Kahyun Lee, MIT, USA
Pedro Ruas, University of Lisbon, Portugal
Hoyun Song, KAIST Computer Science, Korea
Diana Sousa, University of Lisbon, Portugal
Wonsuk Yang, KAIST Computer Science, Korea

Invited Speaker:

Guergana Savova, Boston Children's Hospital and Harvard Medical School, USA

Table of Contents

Conference Program

November 20, 2020

09:00–10:45 Session 1

9:00 ***Introduction***

09:15 *The Impact of De-identification on Downstream Named Entity Recognition in Clinical Text*
Hanna Berg, Aron Henriksson and Hercules Dalianis

09:30 *Simple Hierarchical Multi-Task Neural End-To-End Entity Linking for Biomedical Text*
Maciej Wiatrak and Juha Iso-Sipila

09:40 *Medical Concept Normalization in User-Generated Texts by Learning Target Concept Embeddings*
Katikapalli Subramanyam Kalyan and Sivanesan Sangeetha

09:50 *Not a cute stroke: Analysis of Rule- and Neural Network-based Information Extraction Systems for Brain Radiology Reports*
Andreas Grivas, Beatrice Alex, Claire Grover, Richard Tobin and William Whiteley

10:05 *GGPONC: A Corpus of German Medical Text with Rich Metadata Based on Clinical Practice Guidelines*
Florian Borchert, Christina Lohr, Luise Modersohn, Thomas Langer, Markus Follmann, Jan Philipp Sachs, Udo Hahn and Matthieu-P. Schapranow

10:20 ***Session 1 QA***

November 20, 2020 (continued)

10:45–11:00 Break

11:00–12:20 Session 2

11:00 *Normalization of Long-tail Adverse Drug Reactions in Social Media*
Emmanouil Manousogiannis, Sepideh Mesbah, Alessandro Bozzon, Robert-Jan Sips, Zoltan Szlanik and Selene Baez

11:15 *Evaluation of Machine Translation Methods applied to Medical Terminologies*
Konstantinos Skianis, Yann Briand and Florent Desgrippes

11:30 *Information retrieval for animal disease surveillance: a pattern-based approach.*
Sarah Valentin, Mathieu Roche and Renaud Lancelot

11:45 *Multitask Learning of Negation and Speculation using Transformers*
Aditya Khandelwal and Benita Kathleen Britto

12:00 *Session 2 QA*

12:20–14:00 Break

14:00–14:45 Invited Talk

14:00 *TBA*
Guergana Savova

14:30 *Invited Talk QA*

November 20, 2020 (continued)

14:45–15:00 Break

15:00–16:15 Session 3

15:00 *Biomedical Event Extraction as Multi-turn Question Answering*
Xing David Wang, Leon Weber and Ulf Leser

15:15 *An efficient representation of chronological events in medical texts*
Andrey Kormilitzin, Nemanja Vaci, Qiang Liu, Hao Ni, Goran Nenadic and Alejo Nevado-Holgado

15:25 *Defining and Learning Refined Temporal Relations in the Clinical Narrative*
Kristin Wright-Bettner, Chen Lin, Timothy Miller, Steven Bethard, Dmitriy Dligach, Martha Palmer, James H. Martin and Guergana Savova

15:40 *Context-Aware Automatic Text Simplification of Health Materials in Low-Resource Domains*
Tarek Sakakini, Jong Yoon Lee, Aditya Duri, Renato F.L. Azevedo, Victor Sadauskas, Kuangxiao Gu, Suma Bhat, Dan Morrow, James Graumlich, Saqib Walayat, Mark Hasegawa-Johnson, Thomas Huang, Ann Willemsen-Dunlap and Donald Halpin

15:55 *Session 3 QA*

16:15–16:30 Break

16:30–17:30 Session 4

16:30 *Identifying Personal Experience Tweets of Medication Effects Using Pre-trained RoBERTa Language Model and Its Updating*
Minghao Zhu, Youzhe Song, Ge Jin and Keyuan Jiang

16:45 *Detecting Foodborne Illness Complaints in Multiple Languages Using English Annotations Only*
Ziyi Liu, Giannis Karamanolakis, Daniel Hsu and Luis Gravano

17:00 *Detection of Mental Health from Reddit via Deep Contextualized Representations*
Zhengping Jiang, Sarah Ita Levitan, Jonathan Zomick and Julia Hirschberg

17:15 *Session 4 QA*