EACL 2021

The 12th International Workshop on Health Text Mining and Information Analysis LOUHI 2021

Proceedings of the Workshop

April 19, 2021

©2021 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

ISBN 978-1-954085-16-9

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 11 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 colocated with EMNLP 2018 in Brussels, Belgium; in 2019 was co-located with EMNLP 2019 in Hong Kong; and in 2020 was co-located with EMNLP 2020 and took place online due to the COVID-19 pandemics. This year the workshop is co-located with EACL 2021 and takes place online due to the persistence of the COVID-19 pandemics.

The aim of the LOUHI 2021 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 20 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 11 papers. The selection was entirely based on the scores provided by the reviewers. The overall acceptance rate is 55%.

Our special thanks go to Karin Verspoor 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) Eben Holderness (Brandeis University, USA) Antonio Jimeno Yepes (IBM Research Australia) Anne-Lyse Minard (LLL, CNRS, University of Orléans, France) James Pustejovsky (Brandeis University, USA) Fabio Rinaldi (Dalle Molle Institute for Artificial Intelligence Research - IDSIA, 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 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 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 (LLL, CNRS, University of Orléans, France) Hans Moen, University of Turku, Finland Diego Molla, Maguaire University, Australia Danielle L Mowery, University of Utah, USA 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, 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 Özlem Uzuner, MIT, USA Yanshan Wang, Mayo Clinic, USA Pierre Zweigenbaum, LIMSI, CNRS, Université Paris-Saclay, Orsay, France

Additional Reviewers:

Fatemah Husain, Kuwait University (Kuwait)

Invited Speaker:

Karin Verspoor, RMIT University (Australia)

Table of Contents

ArCorona: Analyzing Arabic Tweets in the Early Days of Coronavirus (COVID-19) Pandemic Hamdy Mubarak and Sabit Hassan
Multilingual Negation Scope Resolution for Clinical Text Mareike Hartmann and Anders Søgaard 7
Understanding Social Support Expressed in a COVID-19 Online Forum Anietie Andy, Brian Chu, Ramie Fathy, Barrington Bennett, Daniel Stokes and Sharath Chandra Guntuku
Fast and Effective Biomedical Entity Linking Using a Dual EncoderRajarshi Bhowmik, Karl Stratos and Gerard de Melo28
Leveraging knowledge sources for detecting self-reports of particular health issues on social media Parsa Bagherzadeh and Sabine Bergler
Integrating Higher-Level Semantics into Robust Biomedical Name Representations Pieter Fivez, Simon Suster and Walter Daelemans
Classification of mental illnesses on social media using RoBERTa Ankit Murarka, Balaji Radhakrishnan and Sushma Ravichandran
<i>Topic Modeling for Maternal Health Using Reddit</i> Shuang Gao, Shivani Pandya, Smisha Agarwal and João Sedoc
FuzzyBIO: A Proposal for Fuzzy Representation of Discontinuous Entities Anne Dirkson, Suzan Verberne and Wessel Kraaij 77
Cluster Analysis of Online Mental Health Discourse using Topic-Infused Deep Contextualized Represen- tations
Atharva Kulkarni, Amey Hengle, Pradnya Kulkarni and Manisha Marathe
Scientific Claim Verification with VerT5erini
Ronak Pradeep, Xueguang Ma, Rodrigo Nogueira and Jimmy Lin