LOUHI 2022

13th International Workshop on Health Text Mining and Information Analysis

Proceedings of the Workshop

December 7, 2022

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 978-1-959429-13-5

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 12 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; in 2019 was co-located with EMNLP 2019 in Hong Kong; in 2020 was co-located with EMNLP 2020 and took place online due to the COVID-19 pandemics; and in 2021 was co-located with EACL 2021 and took place online due to the persistence of the COVID-19 pandemics. This year the workshop is co-located with EMNLP 2022 and takes place with a hybrid modality.

The aim of the LOUHI 2022 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 56 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 25 papers. The selection was entirely based on the scores provided by the reviewers. The overall acceptance rate is 45%.

Our special thanks go to Tim Baldwin 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.

Organizing Committee

Organizers

Alberto Lavelli, FBK, Trento, Italy
Eben Holderness, Brandeis University, USA
Antonio Jimeno Yepes, RMIT University, Australia
Anne-Lyse Minard, LLL, CNRS, University of Orléans, France
James Pustejovsky, Brandeis University, USA
Fabio Rinaldi, IDSIA, University of Zurich, Switzerland, and FBK, Trento, Italy

Program Committee

Reviewers

Rafael Berlanga Llavori

Leonardo Campillos Llanos, Francisco M. Couto

Hercules Dalianis

Natalia Grabar, Cyril Grouin

Thierry Hamon, Eben Holderness

Antonio Jimeno Yepes

Yoshinobu Kano

Alberto Lavelli, Analia Lourenco

David Martinez, Sérgio Matos, Timothy Miller, Anne-Lyse Minard, Hans Moen, Diego Molla, Roser Morante, Danielle L Mowery

Aakanksha Naik, Mariana Lara Neves, Aurélie Névéol

Jong C. Park, Laura Plaza, James Pustejovsky

Fabio Rinaldi, Thomas Brox Røst

Tapio Salakoski, Maria Skeppstedt, Amber Stubbs, Hanna Suominen

Suzanne Tamang

Pierre Zweigenbaum

Keynote Talk: Deep Phonology: Analysing Antimicrobial Stewardship in Veterinary Clinics through NLP

Tim Baldwin
Mohamed bin Zayed University of Artificial Intelligence, UAE

Abstract: Antimicrobial stewardship refers to guidelines on the appropriate use of antimicrobials to optimise patient health and minimise microbial resistance. In this talk, I will present work on the large-scale analysis of veterinary clinical records to perform fine-grained analysis to aid in the implementation and monitoring of antimicrobial stewardship programmes in Australia.

Bio: Tim Baldwin is Associate Provost (Academic and Student Affairs) and Head of the Department of Natural Language Processing, Mohamed bin Zayed University of Artificial Intelligence in addition to being a Melbourne Laureate Professor in the School of Computing and Information Systems, The University of Melbourne. His primary research focus is on natural language processing (NLP), including social media analytics, deep learning, and computational social science.
Tim completed a BSc(CS/Maths) and BA(Linguistics/Japanese) at The University of Melbourne in 1995, and an MEng(CS) and PhD(CS) at the Tokyo Institute of Technology in 1998 and 2001, respectively. Prior to joining The University of Melbourne in 2004, he was a Senior Research Engineer at the Center for the Study of Language and Information, Stanford University (2001-2004). His research has been funded by organisations including the Australia Research Council, Google, Microsoft, Xerox, ByteDance, SEEK, NTT, and Fujitsu, and has been featured in MIT Tech Review, IEEE Spectrum, The Times, ABC News, The Age/Sydney Morning Herald, Australian Financial Review, and The Australian. He is the author of well over 400 peer-reviewed publications across diverse topics in natural language processing and AI, with around 20,000 citations and an h-index of 66 (Google Scholar), in addition to being an ARC Future Fellow, and the recipient of a number of awards at top conferences.

Table of Contents

Program

Wednesday, December 7, 2022

09:00 - 09:10 *Opening Remarks*

09:10 - 10:00 *Invited Talk*

10:00 - 10:30 *TBD*

Can Current Explainability Help Provide References in Clinical Notes to Support Humans Annotate Medical Codes?
Byung-Hak Kim, Zhongfen Deng, Philip Yu and Varun Ganapathi

10:30 - 11:00 *Coffee Break*

11:00 - 12:30 *Session 2*

Assessing the Limits of Straightforward Models for Nested Named Entity Recognition in Spanish Clinical Narratives
Matias Rojas, Casimiro Pio Carrino, Aitor Gonzalez-Agirre, Jocelyn Dunstan and Marta Villegas

A Quantitative and Qualitative Analysis of Schizophrenia Language
Amal Alqahtani, Efsun Sarioglu Kayi, Sardar Hamidian, Michael Compton and Mona Diab

Enriching Deep Learning with Frame Semantics for Empathy Classification in Medical Narrative Essays
Priyanka Dey and Roxana Girju

12:30 - 14:00 *Lunch Break*

14:00 - 15:30 *Session 3*

Exploring the Influence of Dialog Input Format for Unsupervised Clinical Questionnaire Filling
Farnaz Ghassemi Toudeshki, Anna Liednikova, Philippe Jolivet and Claire Gardent

DDI-MuG: Multi-aspect Graphs for Drug-Drug Interaction Extraction
Jie Yang, Yihao Ding, Siqu Long, Josiah Poon and Soyeon Caren Han

Wednesday, December 7, 2022 (continued)

Divide and Conquer: An Extreme Multi-Label Classification Approach for Coding Diseases and Procedures in Spanish
Jose Barros, Matias Rojas, Jocelyn Dunstan and Andres Abeliuk

15:30 - 16:00 *Coffee Break*

16:00 - 17:15 *Session 4 (Poster Session)*

17:15 - 17:30 *Mini Break*

17:30 - 19:00 *Session 5*

Integration of Heterogeneous Knowledge Sources for Biomedical Text Processing
Parsa Bagherzadeh and Sabine Bergler

How Long Is Enough? Exploring the Optimal Intervals of Long-Range Clinical Note Language Modeling
Samuel Cahyawijaya, Bryan Wilie, Holy Lovenia, Huan Zhong, MingQian Zhong, Yuk-Yu Nancy Ip and Pascale Fung

Proxy-based Zero-Shot Entity Linking by Effective Candidate Retrieval
Maciej Wiatrak, Eirini Arvaniti, Angus Brayne, Jonas Vetterle and Aaron Sim