Natural Language Processing Tutorial for Biomedical Text Mining - Abstract

Şenay Kafkas[1,2,*], Sumyyah Toonsi[1,2] and Sakhaa Alsaedi[1,2]

[1]Computer, Electrical and Mathematical Sciences & Engineering (CEMSE) Division, King Abdullah University of Science and Technology (KAUST), Thuwal, 23955, Kingdom of Saudi Arabia

[2]Computational Bioscience Research Center, King Abdullah University of Science and Technology (KAUST), Thuwal, 23955, Kingdom of Saudi Arabia

Abstract

In this tutorial, we introduce Natural Language Processing (NLP) for text mining (TM) in the biomedical domain. The tutorial is structured such that each main concept is backed by hands-on exercises. We start by introducing the difference between NLP and TM. Then continue with motivating the need for text mining in the biomedical domain. Next, we introduce basic concepts of NLP tasks such as Named Entity Recognition (NER), Named Entity Normalization (NEN), and Relationship Extraction (RE. We also cover in detail the widely used methods being used to implement these tasks. These methods include dictionary/ontology-based, rule-based, and advanced machine/deep learning-based approaches. In particular, we cover language models like Word2Vec and transformers (e.g. BERT) and their applications. Furthermore, we discuss the recent advancements in NLP by focusing on the large language models covering GPT, ChatGPT, and others. We conclude our tutorial by discussing limitations and ethics in NLP where we cover the best practices to develop state-of-the-art NLP and TM tools.
The learning objectives of this tutorial are:

- Differences between NLP and TM
- The need for biomedical TM
- Implementation of fundamental NLP tasks: NER, NEN and RE
- Current and future trends in NLP
- Limitations and ethics in NLP and biomedical TM

The learning outcomes of this tutorial are:

- Familiarity with current NLP techniques/tools being used for biomedical TM
- Basic skills to use and develop fundamental NLP tools such as NER and RE
- Familiarity with the current as well as expected future trends in NLP
- Familiarity with the ethics and limitations of biomedical TM

Materials of this tutorial are available from https://github.com/stoonsi/ICBO-NLP-for-Biomedical-Text-Mining-tutorial/tree/main

Keywords

Natural Language Processing, Text mining

Proceedings of the International Conference on Biomedical Ontologies 2023, August 28th-September 1st, 2023, Brasilia, Brazil

*Corresponding author.

senay.kafkas@kaust.edu.sa (Ş. Kafkas); sumyyah.toonsi@kaust.edu.sa (S. Toonsi); sakhaa.alsaedi@kaust.edu.sa (S. Alsaedi)

0000-0001-7509-5786 (Ş. Kafkas); 0000-0003-4746-4649 (S. Toonsi); 0000-0001-7142-8715 (S. Alsaedi)