

Answering consumer health questions with non-expert language

Courses: MEI; MCD

Supervisor: Sérgio Matos (aleixomatos@ua.pt)

Collaborator: Tiago Almeida (tiagomeloalmeida@ua.pt)

Context and objectives

Online search for health information by patients, family or informal caregivers is a common and important activity that has received deserved attention over the years. With the appearance of chatbot applications, searches and conversations regarding health issues can be expected to increase, which raises the challenges on how to reply with verified and reliable information. This work will consider the task of patient information search, which requires simpler but knowledge grounded answers. Towards that end, the work will focus particularly on retrieval and lay language summarization of scientific documents.

The objective of this work is to develop and evaluate a method for retrieval and summarization of scientific articles that contain answers to consumer health questions. The work will make use of the BioLaySumm challenge datasets, and participation on the next challenge will be considered.

Technologies and frameworks: The work will be based on deep learning frameworks, namely PyTorch and HuggingFace.

Work plan

- Literature review on LLMs, neural information retrieval, and summarization
- Literature review on question answering, focusing on consumer health questions
- Design and development of the retrieval and summarization model
- Validation of the resulting models on available benchmark datasets
- Application and validation on the BioLaySumm challenge (subject to results and timeline of work)
- Dissertation writing

More information

<https://lhncbc.nlm.nih.gov/LHC-research/LHC-projects/NLP/chqa.html>

<https://biolaysumm.org>