BioMedical NLP - final project topics 2023

1. Shared task participation - potential publication! (watch out for the registration/submission deadline):
    - **eRisk** 2023: https://erisk.irlab.org/ (Tasks on depression/gambling addiction/eatcing disorders)
    - **MentalRiskES** @ IberLEF: https://sites.google.com/view/mentalriskes (detecting mental health problems in Spanish)
    - **MEDIQA-Sum** / **MEDIQA-Chat** 2023: https://www.imageclef.org/2023/medical/mediqa / https://sites.google.com/view/mediqa2023/clinicalnlp-mediqa-chat-2023 (summarize conversations between doctors and patients, generate clinical notes)
    - **BioLaySumm 2023**: https://biolaysumm.org/ (summarize biomedical articles such that they are readable by laymen).
    - **CLINAIS:** https://ixa2.si.ehu.eus/clinais/ (Automatic Identification of Sections in Clinical Documents for Spanish)
    - **BioNLP Workshop 2023 Shared Task 1A: Problem List Summarization**: https://physionet.org/content/bionlp-workshop-2023-task-1a/1.0.0/
    - **BioNLP Workshop 2023 Shared Task 1B: Radiology Report Summarization:** https://vilmedic.app/misc/bionlp23/sharedtask
    - **TESTLINK** (Relation extraction from clinical data for Spanish and Basque): https://e3c.fbk.eu/testlinkiberlef
    - **SV-Ident 2023** (Survey Variable Identification in Social Science): https://sdproc.org/2023/sharedtasks.html#svident , https://vadis-project.github.io/sv-ident-sdp2022/

    - Other relevant shared tasks from the **Scholarly Document Processing Workshop**: https://sdproc.org/2023/sharedtasks.html#mup

2. Survey of research on a specific sub-topic in bio-medical NLP + prototype implementation of one solution
3. Health related fake news: detect topics in existing datasets / collect new dataset
4. Stance detection for COVID-19 vaccination
5. Semantic change & domain sensitivity in the medical domain: compare embeddings and pretrained models across time and across domains (standard vs medical)
6. Collect and annotate (semi-automatic) mental health dataset for Romanian / other low-resource language / from novel platform (Instagram?), or for less studied disorders (anxiety, …)
7. Implement medical NER on existing dataset
8. Topics in health related scientific papers over time, scientific trend detection
9. Detecting symptoms in social media data
10. Reproduce methods described in a paper discussed during the seminar
11. … (contact me for approval)

Requirements

Deliverables include:

- implementation of a solution to a problem in bio-medical NLP
- document describing the methods used (4-8 pages)
- presentation

The document/paper will follow the classical structure of a research article:

- short summary (abstract)
- analysis of main idea
- related work: state of the art (SOTA) where it exists, short history, recent and/or related results
- conclusions and future work, directions for further improvement
- references

In case you're presenting a survey: explain main methodologies and selection process (i.e. you are surveying either chronologically, or in order of SOTA achievements), discuss advantages and disadvantages to the methods used.
In the case of presenting specific applications: describe the method, compare it with other results in the field

Teams are 1 to 4 students. max 4 teams can approach the same topic, except for shared tasks where there is no limit (as long as you participate as separate teams and solutions are different!)
Add your choice next to your name on the google sheets document with your name.

More instructions on projects and tools for developing NLP solutions & reading/writing scientific research: https://github.com/ananana/nlp-projects/blob/main/README.md