Makes pycharm work pose sequoia:

/usr/libexec/ApplicationFirewall/socketfilterfw --add /Applications/"PyCharm CE.app"

Useful

For the data scraping of clinical trial data from the .gov.

<https://www.learndatasci.com/tutorials/ultimate-guide-web-scraping-w-python-requests-and-beautifulsoup/>

Biobert and spacy are necessary for the nlp ner processing of the exclusion and processing of patient and clinical trial data. I think that scispacy- something I haven’t actually yet heard of may be able to do that? But either way apparently spacy has a entity linking

To elaborate in order to use this spacy function I need to have a model in the spacy to make it work but I need a model that supports entity linking… does biobert do this?

Inclusion and exclusion selector:

#eligibility-criteria-description

Inclusion and exclusion xpath:

//\*[@id="eligibility-criteria-description"]

Inclusion and exclusion full xpath:

/html/body/app-root/main/ctg-study-details/div[2]/ctg-study-info/div/div/div/div[3]/ctg-participation-criteria/div[2]/div/div[2]/div[1]/ctg-long-text

Python 3.8 is the only pyhton that works with all the packages: scispacy mainly also had to use :

* CFLAGS="-mavx -DWARN(a)=(a)" pip install nmslib

To make nmslib work

A screenshot of a computer

Description automatically generatedA screenshot of a computer

Description automatically generated

Here I am on 7 october still struggling with the scraping which I will make todays priority:

Discoveres the v2 api can be used to extract the information however the diffiuculty that im having now is how do I cycle through all the studies that are recruiting???? As well as put this into a callable function later on.?

Questions now are whether to search through CTG with Patient DATA or sequential modular smt? I am really unsure but Also need to automate searching through all recruiting CTs. And how dshould I process the patient data?

Maybe per patient post aggregating all of their content I will put geo and inclsuipn/ condition data as part odf the search quesry? Would narrow them

THIS IS DUE ON FRIDAY…

Read tasks make action but also chill today maybe

Resources:

<https://stackoverflow.com/questions/78415818/how-to-get-full-results-with-clinicaltrials-gov-api-in-python>

<https://github.com/allenai/scispacy>

<https://spacy.io/usage/linguistic-features#entity-linking>

try it!

<https://www.youtube.com/watch?v=dIUTsFT2MeQ>

?what is this?

<https://github.com/daveramseymusic/biobert-bern/blob/main/biobert_bern/biobertApi.py>

10/OCTOBER/24

Additional and useful:

<https://clinicaltrials.gov/data-api/about-api/search-areas#TitleSearch>

<https://clinicaltrials.gov/data-api/about-api/study-data-structure>

What I can see is that I can search for things that match the query data quite easily. I am not sure however about the UNKNOWN titled status clinical trials I could ignore the unknown ones. Also the invitation ones are a bit odd as perhaps with prompting they could incorporate the patient.

Nonetheless what I will do next is use patient data to query the API for relevant studies <Maybe irrelevant of location….>

This does mean that I have to sort the patient data or find a measure of the most informative information to use for the matching process.

Have to manually split eligibility criteria