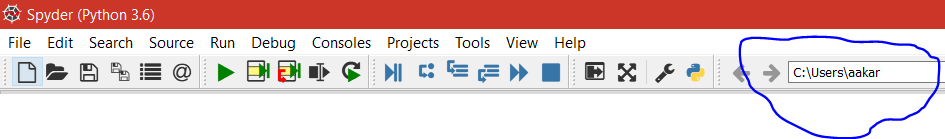
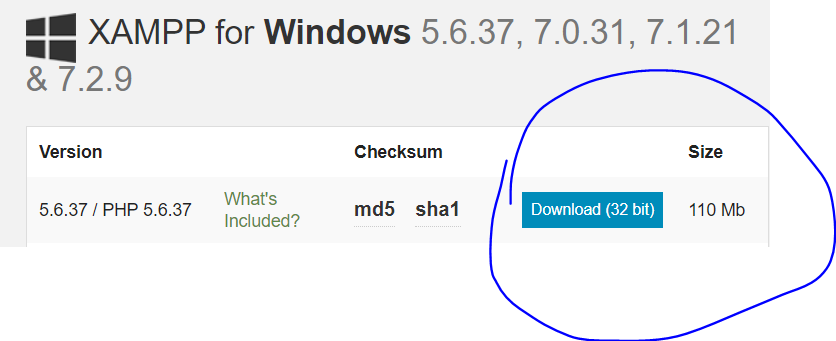
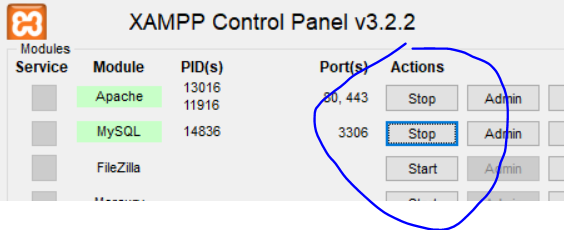
1. Install latest version (3.6) of Python from <https://www.anaconda.com/download/> according to your system requirements.
2. From the search bar of the start menu type ‘*Spyder*’, to run python in Spyder IDE.
3. Search for ‘*Anaconda Navigator*’ from the start menu to install required packages. After opening Anaconda Navigator, click on Environment at the left and search for the packages below changing the drop down to ‘Not Installed’. The packages to be downloaded are:
   1. Nltk
   2. Pymysql
   3. Biopython
4. Download MedPost POS tagger from <ftp://ftp.ncbi.nlm.nih.gov/pub/lsmith/MedPost/medpost.tar.gz> to python root directory (You can see this in Spyder window as shown below).



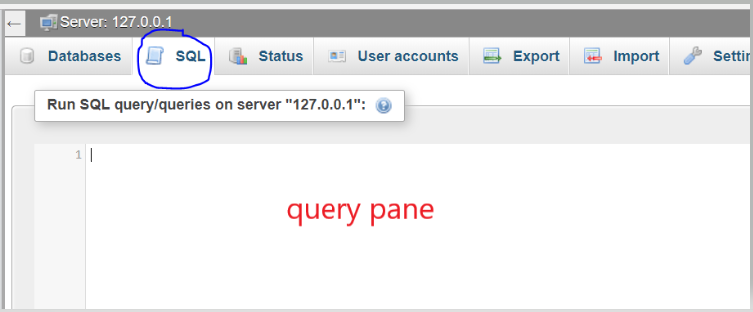
1. Make sure you have WinRar installed, to extract or unzip the above file in the same location.
2. Save ‘*words.db*’ file as ‘words.txt’ in side medpost folder, so that it would be easy to read text file in python (make sure you provide .txt extension while saving).
3. Download MRCONSO.RRF, MRSTY.RRF and *queries.txt* files provided which is UMLS data we need to access in a single folder at your desired location. *The files are downloaded from UMLS website. Look for the optional steps for more details.*
4. Open *queries.txt file* and provide the path of both files (MRCONSO .RRF and MRSTY.RRF) in the format as of *'C:/xampp/htdocs/umls/META/MRSTY.RRF'* (make sure to convert the backslash to forward slash when providing your path location)
5. To host UMLS database on SQL Server install XAMPP on your machine. XAMPP can be installed from <https://www.apachefriends.org/download.html>



1. To host data in XAMPP SQL server, open XAMPP application, then start Apache and MySQL actions as shown below. Now in chrome browser, type ‘*localhost*’ in address bar and click ‘*phpMyAdmin*’ to get SQL server environment.



1. Select SQL option to run all the queries provided in ‘queries.txt’ file in the query pane and click on *Go.* (Be patient as it takes time to execute all the queries).



1. Open the *‘geneprotein.py’* file in Spyder IDE and run as per the steps mentioned in the file. See the intermediate files generated to get to know about the process.
2. ‘*finaldataset.csv’* will generate the sentences containing more than one gene/protein and the action between them.
3. Use the above generated “*finaldataset.csv*” for generating visualization dataset in the next step of code(Step : 6) Generated file will be “*Visualization.csv*”
4. Download Cytoscape from the below mentioned link to create a network visualization.

<http://www.cytoscape.org/download.php>

1. Import “Visualization.csv” into Cytoscape as follows.

At the left corner Go to File>Import>Network>File

New window is opened and select “*Visualization.csv*” and click open

Again New window is opened and click “OK”

You can see the Interaction network.

**OPTIONAL**

* UMLS data can be downloaded from [https://www.nlm.nih.gov/research/umls/licensedcontent/umlsknowledgesources.html](https://www.google.com/url?q=https://www.nlm.nih.gov/research/umls/licensedcontent/umlsknowledgesources.html&sa=D&source=hangouts&ust=1531687550615000&usg=AFQjCNEIWf13-67nQNnNvXNKFRwHuod7Eg) . Make sure you create an account before downloading the data.

**Note: Please refer to NLPCodeSteps.doc for more information on steps involved for preparation of data that is suitable for Deep Learning**