**Steps to run the files and expected outcome:**

1) Install the required packages available in Requirements.txt

- can use command like pip install -r Requirements.txt which loads all the needed packages if any dependency is missing

2) Git pull

- latest code is available on [https://github.com/sanjshroff/Biomedical](https://github.com/sanjshroff/Biomedical%20) main branch and operation of git pull can be performed to load it to local PC

3) For checking performance of different models tested in the project run python clustering.py

-clear terminal window before running all files

-after running successfully, displays time and performance for different models

- Prepared dataset expected output:

Text

Description automatically generated

* Indicates performance for Agglomerative and kmeans and showcases true labels
* Ground truth TREC dataset expected output:

Text

Description automatically generated

3)Result File generated under Output folder

Graphical user interface, text, application

Description automatically generated

-agglo\_with\_cluster\_num.csv has the cluster numbers mapped to combined input data for Agglomerative clustering

-kmeans\_title\_with\_cluster\_num.csv has only titles and cluster labels for kmeans clustering

-sbert\_topk.csv contains the cosine similarity score within clusters along with document details

4) Result File generated under logs folder

- Contains a subfolder wordcloud which contains the word clouds for prepared dataset

Graphical user interface, diagram

Description automatically generated with medium confidence

-Expected output is shown above which has word clouds for all 10 clusters

5)Running the file through User interface run python tkinter\_UI.py

-Follow these steps to view expected results

-Please not this section is not completed against all different situations

Graphical user interface, application, Word

Description automatically generated

-Step 1 is to enter a PMID recommended to try 34736317 as it has certain topics in the predefined dataset, and it comes from the topic of cheminformatics

-Step 2 Click on the Enter PMID button, input is not saved if this button is not clicked

-Step 3 Enter the number of documents to retrieve

-Step 4 Click on the button enter number of records to fetch (k)

-When both the buttons are pressed above, the dialog is destroyed automatically

-If the dialogue box does not disappear, click on Exit

-Once the UI is disappeared, the following values are displayed in the terminal

Text

Description automatically generated

Additional Information

-UI and NLP tasks are Supported for across all Windows OS

-UI has different appearance on MAC OS

-SBERT not supported on MAC chip M1 and M2

-Tried creating Docker image but unable to host, please run through git