Recommender System Manual (recsys_sbert.ipnyb)

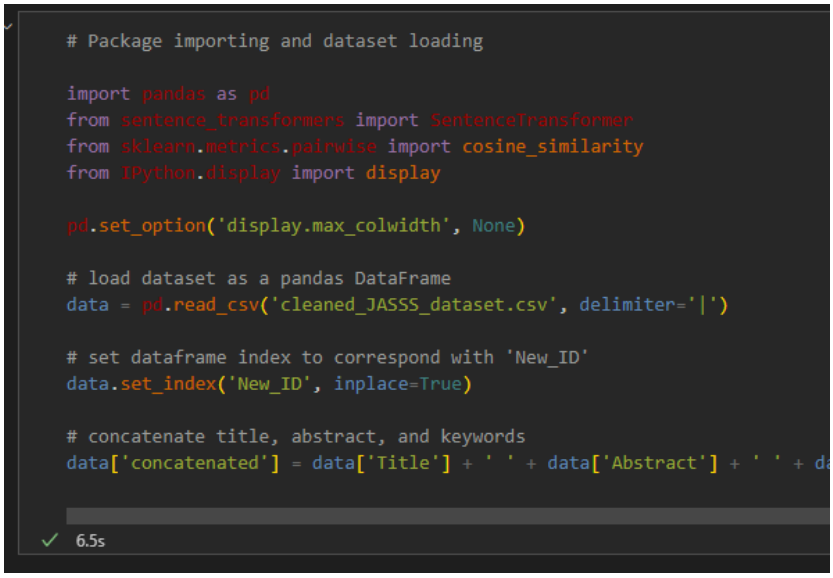
1. Install required packages using ‘pip install <package name>’

You will need to install:

- Scikitlearn (pip install scikit-learn)
- Sentence Transformers (pip install sentence_transformers)

It is assumed you have already installed pandas on a Jupyter notebook or is running the script on an online platform such as Google Collab.

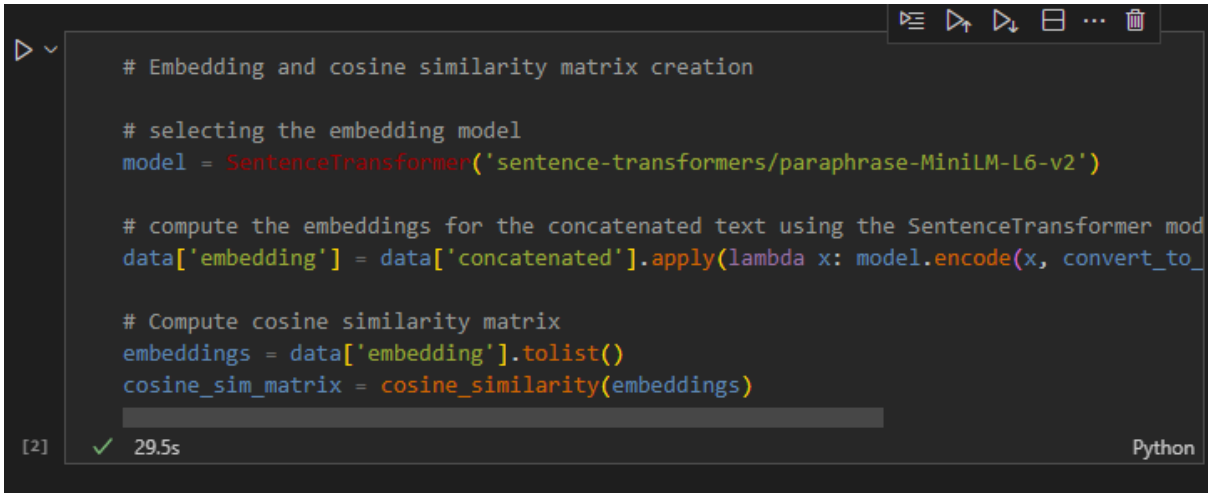
2. Run the first cell. Make sure the name of the dataset you are using is correct and in the correct directory



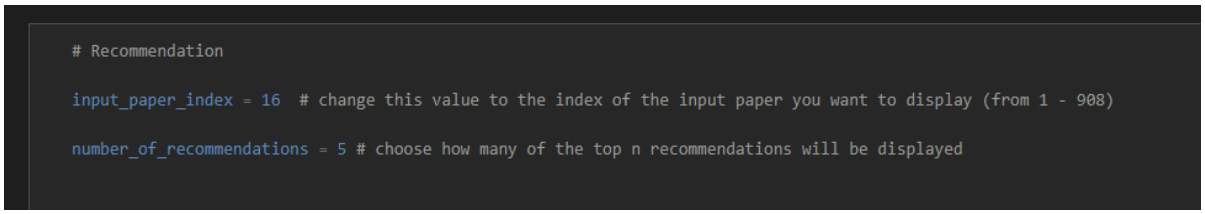
```
# Package importing and dataset loading

import pandas as pd
from sentence_transformers import SentenceTransformer
from sklearn.metrics.pairwise import cosine_similarity
from IPython.display import display

pd.set_option('display.max_colwidth', None)

# load dataset as a pandas DataFrame
data = pd.read_csv('cleaned_JASSS_dataset.csv', delimiter='|')

# set dataframe index to correspond with 'New_ID'
data.set_index('New_ID', inplace=True)

# concatenate title, abstract, and keywords
data['concatenated'] = data['Title'] + ' ' + data['Abstract'] + ' ' + da
```

✓ 6.5s

3. Run the second cell. This will use the ‘sentence-transformers/paraphrase-MiniLM-L6-v2’ embedder model to embed the concatenated titles + abstracts + keywords and create the cosine similarity matrix. This process may take more than half a minute depending on your system performance



4. Select the paper you wish to be the input paper by choosing its index number. Choose the number of recommendations you wish to see. (it is recommended for testing purposes to choose a lower number first as the display tables per paper are large)



5. The recommendations will appear below. First a table showing the details of the input paper will be show. This is followed by the top n most similar papers from the corpus. Below this will be the recommendation by authors in the input paper.

Known Issues:

- The names in the dataset are not fully cleaned as there are papers with different formatting variations for names. This may mean that not all the publications of an author will be taken into account for recommendation if the author's name has been written in a different variation (e.g. Nigel Gilbert, and Gilbert, Nigel;)

- The recommender system is still prone to making ineffective recommendations. The author is aware of this shortcoming and has addressed it in his dissertation.