

Readme:

In this project we are implementing a recommender system, which recommends articles related to the text given by the user as an input(along with the weight of relation of the article with the input).

A dataset of articles is first fed to the code. The program runs on the dataset to retrieve the keywords by eliminating unnecessary words like prepositions, blank or white spaces etc. Once the keywords are retrieved the code performs topic extraction using LDA model. The topics created are then related to each other using heatmaps.

The input from the user is preprocessed to make it into a collection of keywords. The keywords are then sorted into their most relevant topics, based on which the relevant articles are selected from the dataset.

Software:

This program is written using python in jupyter notebook. To use the jupyter notebook, anaconda is required.

>>Install anaconda and activate the anaconda base on your terminal to use the jupyter notebook. Open the code in jupyter notebook.

The first module of the code involved many python libraries out of which some needs to be installed using npm.

>>Install numpy using pip install numpy.

>>Similarly install spacy, scipy, wordcloud and pickle.

>>Install en_core_web_sm from spacy using pip. Upgrade before installing. >>Install plydavis and gensim.

How to run it??

1)open cmd

2)open jupyter using command "jupyter-lab "

3)open the Folder

4)run main.ipynb file

Steps:

>>The program imports the dataset and splits the dataset into articles and stores it in a python list. >>The list is fed into the 'Vocab' class where the data is preprocessed to remove way too frequently or rarely occurring words.

>>The preprocessed data now contains keywords. The data is fed into the gensim model to convert into a topic space.

>>Once converted into a topic space, the relation between the topics is calculated using heatmaps. If the heat map value between two topics is greater than 0.3 it is graphically represented using a line between the topics, else the topics are more or less not related.

>>The user is prompted to enter the input text.

>>The input text is sorted into topics from which the articles are retrieved from the dataset.