**Vector space model test cases**

Although the best method to test your Vector space model would be to copy 1 or 2 lines of any document which seems very specific to only that document and use it as query.

Here are some test cases you can try:

Q1: ‘Developing your Zomato business account and profile is a great way to boost your restaurant’s online reputation’

Output:

1. [('zomato.txt', 0.5812028782957865),
2. ('instagram.txt', 0.22620471057834346),
3. ('swiggy.txt', 0.16784774094251997),
4. ('messenger.txt', 0.15425230061695694),
5. ('whatsapp.txt', 0.1341637571856167),

Text

Description automatically generated

Q2: Warwickshire, came from an ancient family and was the heiress to some land

1. ('shakespeare.txt', 0.16959924501725457),
2. ('nike.txt', 0.11737054067636613),
3. ('volkswagen.txt', 0.10497419581645412),
4. ('motorola.txt', 0.10287494862982087),
5. ('paypal.txt', 0.09783026605008277),
6. ('operating.txt', 0.09641094135740791),

Text

Description automatically generated

PS. You might get a little different order of matching documents which have similarity lower than 0.1

YOU ARE NOT ALLOWED TO USE SCIKIT-LEARN LIBRARY IN PYTHON

You can use this code to do comparative testing of your model.

Steps:

1. Download corpus from BB.
2. Create folder named “Corpus” on your drive.
3. Open https://colab.research.google.com/drive/16KrVxl4wKPcUk5xM9xk4dz9iDdxFbxv?usp=sharing
4. Run code, give drive access and test for any query.