CSE 419L: LAB 04 Assignment: BSBI Algorithm

You have been given a dataset **(D1),** which contains 1000 documents, distributed in to 10 folders.

https://www.kaggle.com/datasets/jensenbaxter/10dataset-text-document-classification

BSBI Algorithm implementation id given at:

https://github.com/benymaxparsa/Inverted-Index-Construction

[DO not Use the dataset provided in this repository]

Execute the BSBI algorithm on **D1** dataset.

1. Identify the minimum block size for which Algorithm can execute 100%.

2. Gradually increase the block size and repeat the experiment.

   For example: If minimum block size is **X** unit, Try **2X** block size for the second time, **4X** for the third time..... Keep repeating your experiment until your find a block size **B (2^nX),** such that all the documents can be stored in that one Block.

   [Note: Algorithm takes Block size in Bytes]

3. For each experiment log the time taken by the algorithm. Draw a line plot to show the relation b/w block size vs time. [Add your RollNo. In the title of the plot].

Do not clear the outputs of the Notebook file. Output traces will be required for evaluation.