

Algorithm Design :

Goal:

- Figure all the common phrases within texts
- Find at how many occurrences for each phrase

What do have to do:

1. Clean Document (Takes Away punctuation / Non- Informative text)
2. Use a suffix tree to find candidate phrase
3. Merge candidate duster and sat. Then print the final clusters

