**Approaches for proceeding with the Duplicate Detection:**

* TF-IDF statistic measure to list out the important words in the document and compare the test data if they have those words or not
* Okapi BM25 – can give us the score of how much the document is matching with the Plot-Genie