COMP9313

Project 3

z5089358

Fengting YANG

Optimization:

In frequency part, first count the tokens in R, and then count the tokens in S and in R. Thus, the frequency size will be smaller than count all tokens in R or in S.

Then, using prefix to get the probably similar sets. Then calculate the similarity then remove redundancy and those tuples which has similarity less than threshold.

Finally, save to file.