1. Used basic prefix filter in the first mapper to reduce the number of emitted pairs.
2. In the first reducer:
   2.1 Used a length filter to reduce further computation.
   2.2 In the nested for loop, only compare half of pairs because the similarity computation is symmetric. (i.e. sim(pair1, pair2)==sim(pair2,pair1))
   2.3 Remove all intermediate variables if possible and directly emit result if possible.
   2.4 when calculating the similarity, check the key term in the reducer equals to the smallest term of the intersection set. If they are equal, then compute similarity. If not, then continue next loop.