1 :
Calculate the frequency of tokens in (R union S)
Ordering the tokens in R and S separately by frequency (for the tokens with same frequency , sort by their element id in increasing order)

2:
Map the records in R and S separately with prefix tokens as the keys and records as the values. In the reducer, as the values get grouped by prefix tokens, all the values passed in a reduce call share the same prefix token. => (prefix token,((Rid,recordR),(Sid,recordS)))
Calculate the similarity between two records and generate ((Rid,Sid),similarity)