2 different indices:

idf index:

similar with stage2:

<word, idf>, inverse document frequency

tfidf index:

in order to compute BM25

we need txt-> length ration, (word:tf)

<txt, length\_Ratio, Map(word,tf)>