

Edward Plesa

CS421

Assignment 3

Cosine similarity between “a” and “b”: 1.0

Cosine similarity between “c” and “d”: 0.8521420302013858

1. The similarity between “a” and “b” made the most sense to me. This is because when you filter the tokens of stop words you get a few words that are completely the same, the only difference being the order in which they are in.
2. Nothing really surprised me about this set of cosine similarities. These short sentences are very similar to each other even with texts “c” and “d” having a few different words, they still have a quite high similarity score as I would expect, given the sentence has nearly the same meaning in plain english.
3. A = "the man with the terrible smell",
B = "man that smells terrible"
C = "hey zain how are you doing"
D = "yo zain whats good fonem"
4. Cosine similarity between “a” and “b”: 0.804465953691378
Cosine similarity between “c” and “d”: 0.5898041787123603
5. The values do seem reasonable in my humble opinion. For text “c” and “d” they similarity is quite low seeing as slang words probably aren’t included in the

google news articles so there is not many words to vectorize. For “a” and “b” the sentences are, again, quite similar so I would expect a high cosine value here.