

NLP Write Up

Aasish Dhungel

May 11, 2021

Question 1) Report similarity scores for the folloing pairs: (dirty, clean), (big, dirty), (big, large), (big, small)

Similarity between ['dirty', 'clean'] \rightarrow 0.36177364

Similarity between ['big', 'dirty'] \rightarrow 0.27424753

Similarity between ['big', 'large'] \rightarrow 0.4694664

Similarity between ['big', 'small'] \rightarrow 0.49960312

Question 2) Report 5 most similar items and the scores to 'polite', 'orange'

5 Most similar words to 'polite'

[('respectful', 0.7213218212127686),
(('courteous', 0.6826084852218628),
(('gracious', 0.6555663347244263),
(('thoughtful', 0.6383759379386902),
(('timid', 0.6307592391967773)]

5 Most similar words to 'orange'

[('emerald', 0.5249948501586914),
(('yellow', 0.5149500966072083),
(('poinsettia', 0.512459933757782),
(('rainbow', 0.511077344417572),
(('upturned', 0.5064201951026917)]

Question 3) Now change the parameters of your model as follows: window=2, size=50. Answer the 2 questions above for this new model.

Similarity between ['dirty', 'clean'] \rightarrow 0.51589286
Similarity between ['big', 'dirty'] \rightarrow 0.3590638
Similarity between ['big', 'large'] \rightarrow 0.68734884
Similarity between ['big', 'small'] \rightarrow 0.7412376

5 most similar words to 'polite'

[('courteous', 0.8280003070831299),
(('timid', 0.7914358973503113),
(('respectful', 0.7899367809295654),
(('forthright', 0.7803406715393066),
(('candid', 0.7640137672424316)]

5 most similar words to 'orange'

[('poinsettia', 0.6837872862815857),
(('alamo', 0.6731881499290466),
(('fringed', 0.6637517809867859),
(('emerald', 0.6446513533592224),
(('blue', 0.6411835551261902)]