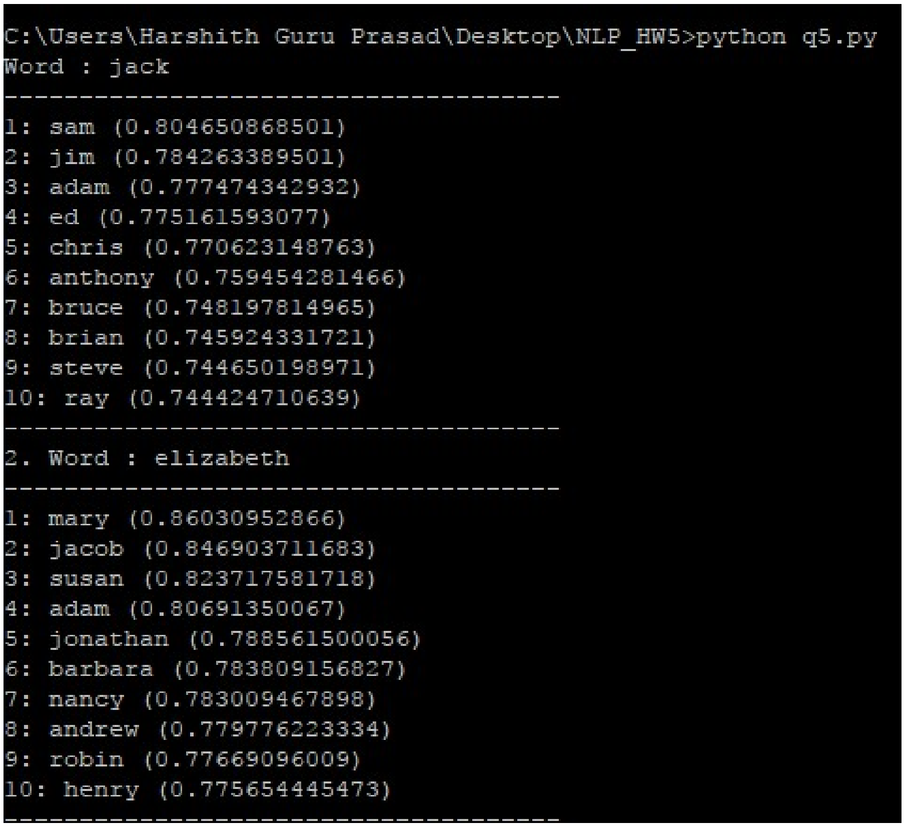
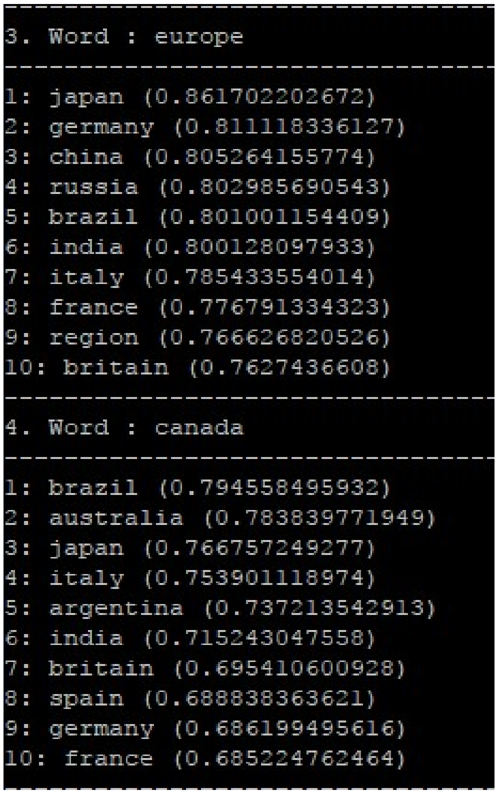
Answer 5

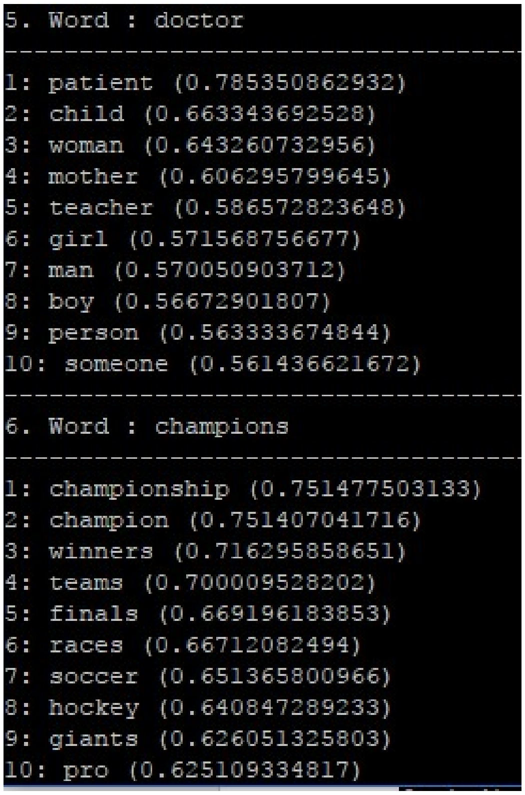
Top 10 results for query words using Dense word embeddings Vectors



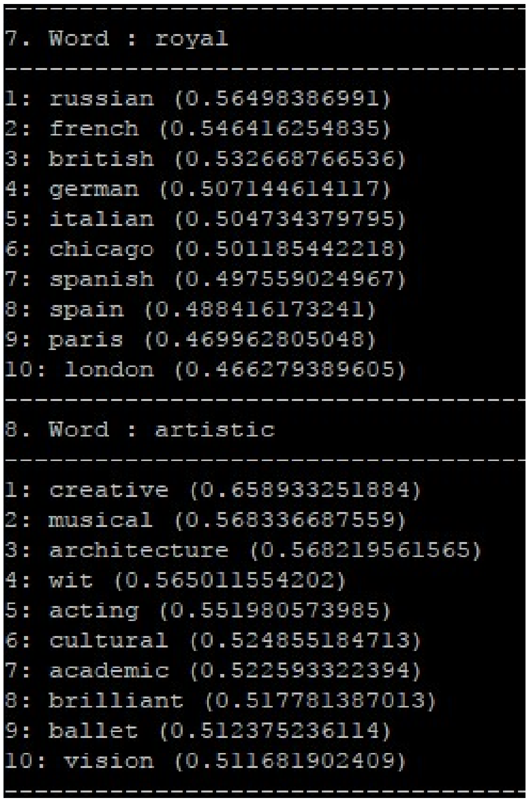
```
C:\Users\Harshith Guru Prasad\Desktop\NLP_HW5>python q5.py
Word : jack
-------------------------------------
1: sam (0.804650868501)
2: jim (0.784263389501)
3: adam (0.777474342932)
4: ed (0.775161593077)
5: chris (0.770623148763)
6: anthony (0.759454281466)
7: bruce (0.748197814965)
8: brian (0.745924331721)
9: steve (0.744650198971)
10: ray (0.744424710639)
-------------------------------------
2. Word : elizabeth
-------------------------------------
1: mary (0.86030952866)
2: jacob (0.846903711683)
3: susan (0.823717581718)
4: adam (0.80691350067)
5: jonathan (0.788561500056)
6: barbara (0.783809156827)
7: nancy (0.783009467898)
8: andrew (0.779776223334)
9: robin (0.77669096009)
10: henry (0.775654445473)
-------------------------------------
```



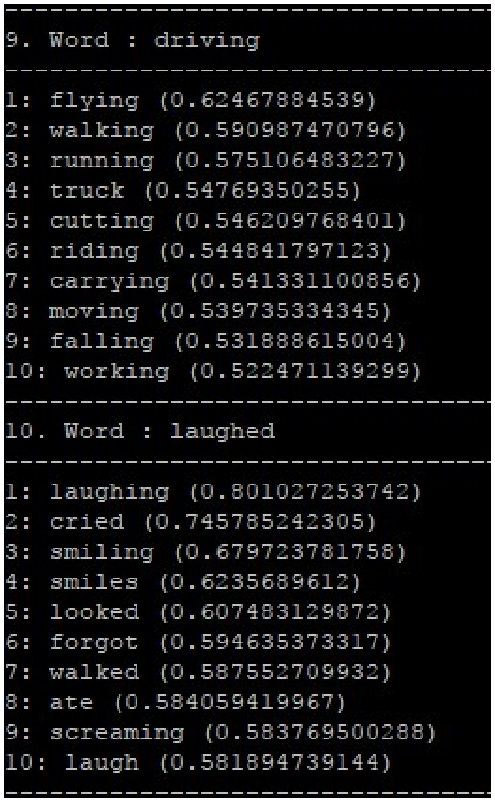
```
---------------------------------
3. Word : europe
---------------------------------
1: japan (0.861702202672)
2: germany (0.811118336127)
3: china (0.805264155774)
4: russia (0.802985690543)
5: brazil (0.801001154409)
6: india (0.800128097933)
7: italy (0.785433554014)
8: france (0.776791334323)
9: region (0.766626820526)
10: britain (0.7627436608)
---------------------------------
4. Word : canada
---------------------------------
1: brazil (0.794558495932)
2: australia (0.783839771949)
3: japan (0.766757249277)
4: italy (0.753901118974)
5: argentina (0.737213542913)
6: india (0.715243047558)
7: britain (0.695410600928)
8: spain (0.688838363621)
9: germany (0.686199495616)
10: france (0.685224762464)
---------------------------------
```



```
5. Word : doctor
-----------------------------------
1: patient (0.785350862932)
2: child (0.663343692528)
3: woman (0.643260732956)
4: mother (0.606295799645)
5: teacher (0.586572823648)
6: girl (0.571568756677)
7: man (0.570050903712)
8: boy (0.56672901807)
9: person (0.563333674844)
10: someone (0.561436621672)
-----------------------------------
6. Word : champions
-----------------------------------
1: championship (0.751477503133)
2: champion (0.751407041716)
3: winners (0.716295858651)
4: teams (0.700009528202)
5: finals (0.669196183853)
6: races (0.66712082494)
7: soccer (0.651365800966)
8: hockey (0.640847289233)
9: giants (0.626051325803)
10: pro (0.625109334817)
```



```
------------------------------------
7. Word : royal
------------------------------------
1: russian (0.56498386991)
2: french (0.546416254835)
3: british (0.532668766536)
4: german (0.507144614117)
5: italian (0.504734379795)
6: chicago (0.501185442218)
7: spanish (0.497559024967)
8: spain (0.488416173241)
9: paris (0.469962805048)
10: london (0.466279389605)
------------------------------------
8. Word : artistic
------------------------------------
1: creative (0.658933251884)
2: musical (0.568336687559)
3: architecture (0.568219561565)
4: wit (0.565011554202)
5: acting (0.551980573985)
6: cultural (0.524855184713)
7: academic (0.522593322394)
8: brilliant (0.517781387013)
9: ballet (0.512375236114)
10: vision (0.511681902409)
------------------------------------
```



```
------------------------------------
9. Word : driving
------------------------------------
1: flying (0.62467884539)
2: walking (0.590987470796)
3: running (0.575106483227)
4: truck (0.54769350255)
5: cutting (0.546209768401)
6: riding (0.544841797123)
7: carrying (0.541331100856)
8: moving (0.539735334345)
9: falling (0.531888615004)
10: working (0.522471139299)
------------------------------------
10. Word : laughed
------------------------------------
1: laughing (0.801027253742)
2: cried (0.745785242305)
3: smiling (0.679723781758)
4: smiles (0.6235689612)
5: looked (0.607483129872)
6: forgot (0.594635373317)
7: walked (0.587552709932)
8: ate (0.584059419967)
9: screaming (0.583769500288)
10: laugh (0.581894739144)
------------------------------------
```

Analysis

Proper Nouns: People names – The semantically similar results are more meaningful in the case of dense word embedding vectors as they are all names of males, while sparse context count vectors return names of both genders. Similarly for elizabeth, the top sematic result is mary(female) as per dense word embedding but adam in the case of sparse context count vectors. Locations – Both sparse context count and dense word embedding vectors return relevant results for names of locations(countries). For proper nouns likes names of people, both vector models provide good semantic results as names of people are used in the same context of an person entity. Different countries can be associated with different contexts.

Common Nouns: doctor – the dense word embedding vectors returns patient as the most similar word for the common noun doctor and both dense and context count vectors return similar results for the word doctor. However, for the common noun champions, dense word embedding technique provides far more relevant and semantically similar results than the context count vectors, which returns vague and highly irrelevant results. Common nouns are used more frequently, and the context count vectors yield poor results as a result of word frequencies with respect to various contexts. Dense vectors yield better results as they are language modeling and feature learning techniques where words or phrases from the vocabulary are mapped to vectors of real numbers.

Adjectives: For the adjective 'royal', context count vectors provide poor semantic results while dense word embedding vectors provide nationalities as semantically similar words. Similarly, for the adjective 'artistic' dense word embeddings provide relevant adjectives that are semantically more similar to 'artistic' than the results generated from the context count vectors. Both methods return adjectives however; the relevance between the results returned differs. Dense vectors perform better as they are prediction based embeddings and not frequency based like the context count vectors. They predict words that are most similar in context and relevance.

Verbs: The verb 'driving' has more relevant results generated from the dense word embedding vectors and are mostly verbs concerned with motion. The results are mostly verbs in the same tense. Context count vectors provide verbs that are semantically less similar and in the same tense as the query word. The verb laughed has more semantically similar and relevant words generated from the dense word embeddings as compared to context count vectors. Since context count vectors use context frequency to determine similarity, verbs which frequently used tend to be more similar but less relevant. However, dense vectors are prediction based and consider the weight of the context and its relevance and do not rely entirely on context frequency which can provide deceptive results based on the words with greatest frequencies in a text .