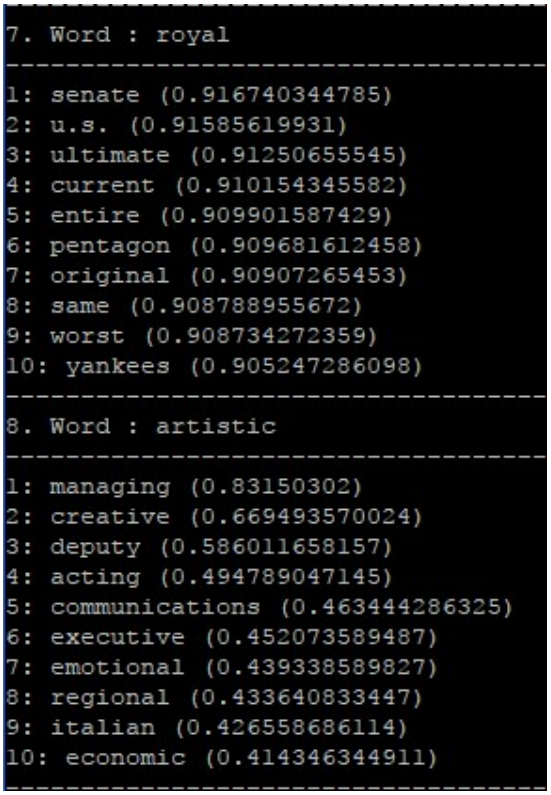
Answer 3

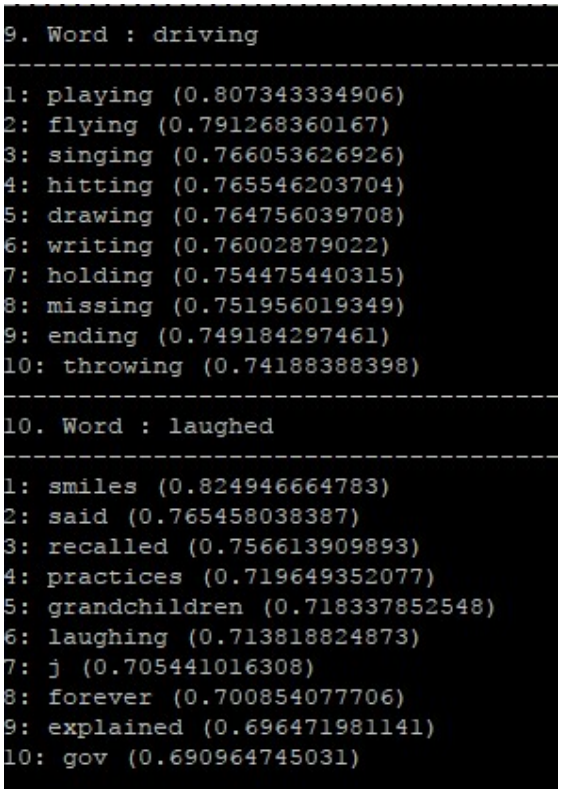
Top 10 results for query words using sparse context count vectors

```
C:\Users\Harshith Guru Prasad\Desktop\NLP_HW5>python q3.py
file nytcounts.4k has contexts for 3648 words
1. Word : jack
-----------------------------------
1: adam (0.879731849466)
2: james (0.859791406534)
3: susan (0.856596258738)
4: daniel (0.847600400729)
5: jonathan (0.847532173876)
6: peter (0.844088466095)
7: eric (0.843254808498)
8: elizabeth (0.843212500388)
9: andrew (0.837502134837)
10: max (0.837313367421)
-----------------------------------
2. Word : elizabeth
-----------------------------------
1: adam (0.882771436888)
2: jonathan (0.868338323453)
3: eric (0.86455850788)
4: daniel (0.863217379325)
5: justin (0.856777947573)
6: peter (0.855567100466)
7: james (0.855382456656)
8: susan (0.851058094611)
9: andrew (0.849156460579)
10: nancy (0.84080228964)
```

```
-----------------------------------
3. Word : europe
-----------------------------------
1: britain (0.95603891122)
2: baghdad (0.954041633678)
3: japan (0.953400171463)
4: paris (0.952949847717)
5: afghanistan (0.951884399275)
6: iraq (0.95124429183)
7: germany (0.945281419977)
8: london (0.945073821088)
9: canada (0.937266768476)
10: france (0.93687244673)
-----------------------------------
4. Word : canada
-----------------------------------
1: france (0.969709142992)
2: japan (0.969410517391)
3: italy (0.966082801486)
4: india (0.965625613461)
5: spain (0.964191539677)
6: germany (0.963245343674)
7: brazil (0.960173412413)
8: britain (0.955476011936)
9: argentina (0.954673744638)
10: china (0.94555092242)
-----------------------------------
```

```
-----------------------------------
5. Word : doctor
-----------------------------------
1: woman (0.94954084661)
2: man (0.936455930487)
3: child (0.9335985004)
4: patient (0.929597458444)
5: kid (0.923367230249)
6: boy (0.918106983379)
7: person (0.914950736293)
8: car (0.905498285702)
9: computer (0.896804544601)
10: giant (0.892801146927)
-----------------------------------
6. Word : champions
-----------------------------------
1: streets (0.732283566787)
2: scenes (0.730724322506)
3: capital (0.728545683851)
4: floor (0.726160279639)
5: pain (0.72386994167)
6: experience (0.723702721305)
7: road (0.723428445906)
8: words (0.723159923239)
9: foundation (0.721219256995)
10: skin (0.719563737015)
-----------------------------------
```





Analysis:

For the first two **proper nouns** – **jack** and **elizabeth**, the most similar words returned makes sense as they are all proper nouns of **names**. The word 'elizabeth' appears in the top ten results of the word 'jack' but the converse is not true as the top ten results for Elizabeth have higher cosine similarity scores than jack. This is on account of the relative positions of the two words while traversing the semantics graph.

For the **locations** such as **europe** and **canada**, the most similar words are all locations (names of countries) and hence the similarity metric used makes sense. Canada appears in the list of top ten most similar words for Europe. However, Europe does not feature in Canada's top ten list of semantically similar words as canada has more semantically similar words(countries) than Europe(continent).

However, for the **common noun doctor**, the words returned are generic and not completely relevant to the semantic meaning of a doctor. The most relevant result returned for doctor is patient. Similarly, for **champions**, the results returned have less similarity with respect to the query word and most of the results are more irrelevant.

For **adjectives** as well, like **'royal'** and **'artistic'**, the output makes less sense on account of context ambiguity and hence makes less sense. Adjectives did not provide relevant or semantically meaningful results. 'Royal' yields results from a different context than what was expected.

Verbs like **'driving'** are similar to other verbs in the same tense. The verb **'laughed'** also fetches other verbs or nouns that it is commonly used with and makes comparatively more sense than adjectives.

Overall, proper nouns have good results while common nouns have a more diverse context. Verbs are semantically similar to other verbs especially the ones in the same tense. Adjectives yield results that are not entirely relevant to the context of the word.