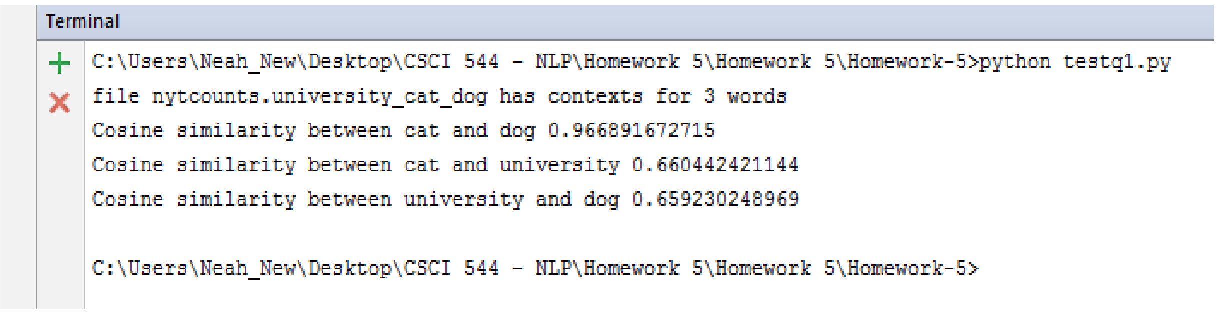
Q1

After implementing the cossim_sparse () function for the file nytcounts.university_cat_dog, the following results are obtained:

file nytcounts.university_cat_dog has contexts for 3 word
Cosine similarity between cat and dog 0.966891672715
Cosine similarity between cat and university 0.660442421144
Cosine similarity between university and dog 0.659230248969



```
C:\Users\Neah_New\Desktop\CSCI 544 - NLP\Homework 5\Homework 5\Homework-5>python testq1.py
file nytcounts.university_cat_dog has contexts for 3 words
Cosine similarity between cat and dog 0.966891672715
Cosine similarity between cat and university 0.660442421144
Cosine similarity between university and dog 0.659230248969

C:\Users\Neah_New\Desktop\CSCI 544 - NLP\Homework 5\Homework 5\Homework-5>
```

Here, it is evident that the cosine similarity of the pair (cat, dog) is high compared to that of the pairs (cat, university) and (university, dog).

The pair (cat, dog) has a high cosine similarity of 0.966891672715 because as per world knowledge (explained in lecture), the concepts ‘cat’ and ‘dog’ appear in similar contexts and they have similar properties. Also, it is worthwhile to restate that the cosine similarity between the vectors of two words is high when both the vectors have high values in same dimensions. Here, each dimension of the vector in nytcounts.university_cat_dog represents the frequency in which the word appears in a particular context. Hence, the cosine similarity is high because there are several contexts in which both ‘cat’ and ‘dog’ are found.
If we were posed the question, ‘How similar are the words ‘cat’ and ‘dog’?’, I would say they are pretty similar. They can be used in several similar contexts, for instance, ‘the large cat..’, ‘the large dog..’, ‘the pet dog..’, ‘the pet cat..’, and so on. Hence, I agree with the cosine similarity metric giving a high value of 0.966 for the pair (cat, dog).

The cosine similarity of the pairs (cat, university) and (university, dog) is low as the two concepts do not have close meanings, they do not occur in similar contexts, they don’t occur together, they are not found together and we do not think of the two concepts together. The vectors for these concepts do not have high values for the same dimensions, hence their cosine similarity is quite low. In real world, there are rarely any contexts where the pair of words (cat, university) and (dog, university) would occur together.

Hence, the high cosine similarity for the pair (cat, dog) and the low cosine similarity for the pairs (cat, university) and (university, dog) makes sense.