

Name: Akash Kumar Gautam
Roll Number: 2015011

Q1:

1. In order to find the way in which the query vector is created, the model for the given task is Google News Vector Model. The stated model is pre-trained on the most frequently used news data articles.

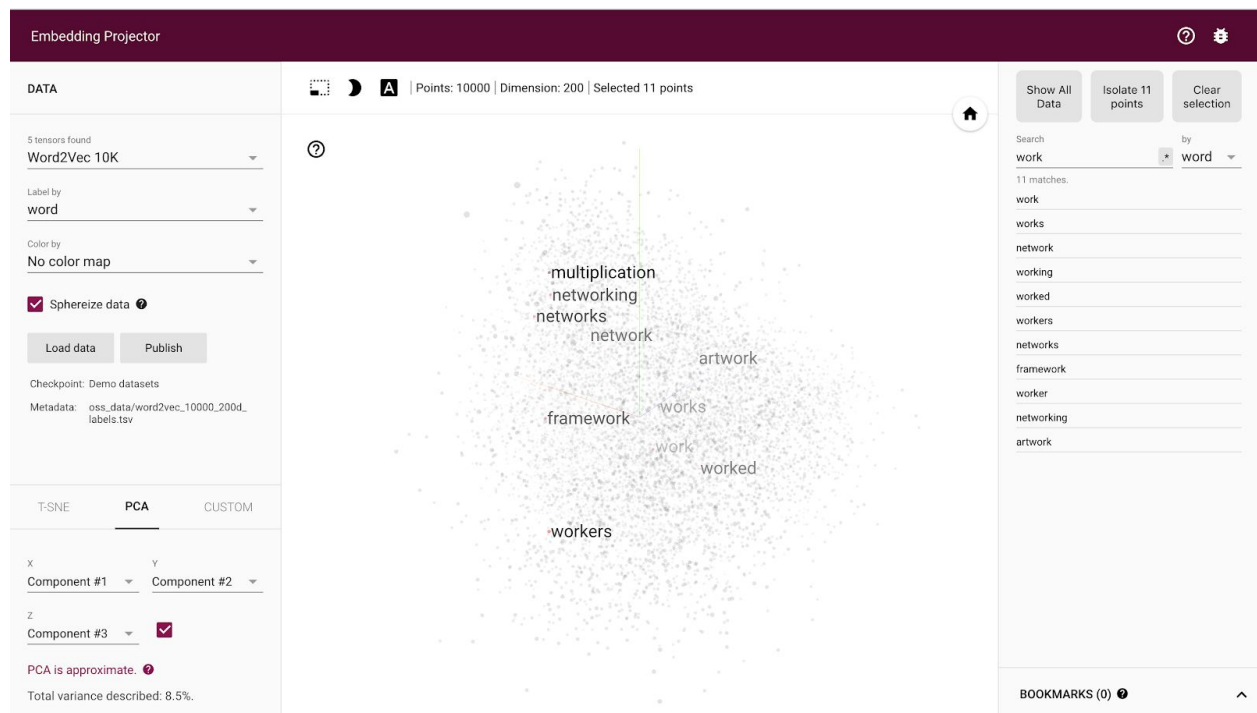
i) The first part asks to find the capital of China by
'China+Delhi'-'India'

Answer produced \Rightarrow Beijing

ii) The second part asks us to find the word related to USA in the given context by
'USA+ ISRO'-'India'

Answer produced \Rightarrow STScl

2. The second part of the question asks us to visit the given URL and set the stated parameters and evaluate the results. Below image shows the same



Q2:

We perform the same set of queries as asked in the question. There are two sets of similarity score vectors the first one finds similarity of doc1 with one doc from each of the classes. The second similarity score vector contains similarity of doc1 with 19 others from the same class(comp.graphics)

The first similarity score vector

```
[('0', 0.6869927644729614), ('8', 0.685661792755127), ('14', 0.6821522116661072), ('3', 0.6763232350349426), ('13', 0.6760851144790649), ('12', 0.6755436062812805), ('9', 0.6750540733337402), ('1', 0.6725870370864868), ('7', 0.6714857220649719), ('6', 0.6693177819252014), ('11', 0.6691198348999023), ('10', 0.6657466292381287), ('4', 0.6614934802055359), ('2', 0.6614078879356384), ('5', 0.6609190702438354), ('15', 0.658626139163971), ('17', 0.6506545543670654), ('16', 0.6490045785903931), ('18', 0.610402524471283)]
```

The second similarity score vector

```
[('2', 0.8181167840957642), ('1', 0.8088515996932983), ('5', 0.8077327013015747), ('0', 0.8067080974578857), ('8', 0.8049415946006775), ('11', 0.8036221265792847), ('13', 0.8012580871582031), ('3', 0.799538791179657), ('4', 0.7967931628227234), ('12', 0.7949236035346985), ('10', 0.7939199209213257), ('9', 0.7916457653045654), ('6', 0.7910410165786743), ('14', 0.7899700403213501), ('17', 0.7887645959854126), ('7', 0.7831411957740784), ('15', 0.7802196741104126), ('18', 0.7797093391418457), ('16', 0.7746587991714478)]
```

It can be seen clearly that there is an improvement of similarity of doc1 with docs of same class. Hence the results are as expected.

Q3: On running the code on the terminal, the output will be obtained