Assignment 6, Deadline 18th Nov
In this assignment you will perform handson on some common NLP tools

1)**Word2Vec**: Word Similarity **[4 marks]**
LInk: https://github.com/mmihaltz/word2vec-GoogleNews-vectors
Load model: gensim.models.KeyedVectors.load_word2vec_format(path)
Perform the following task:

If Delhi is the capital of India then what is the capital of China?
If ISRO is related to India then what is related to USA.

Visit "projector.tensorflow.org" to visualize the vector space, show the following visualization :

Vector (Questions) = Vector (answers) - Vector (answer) + Vector (question)

Note your observation.

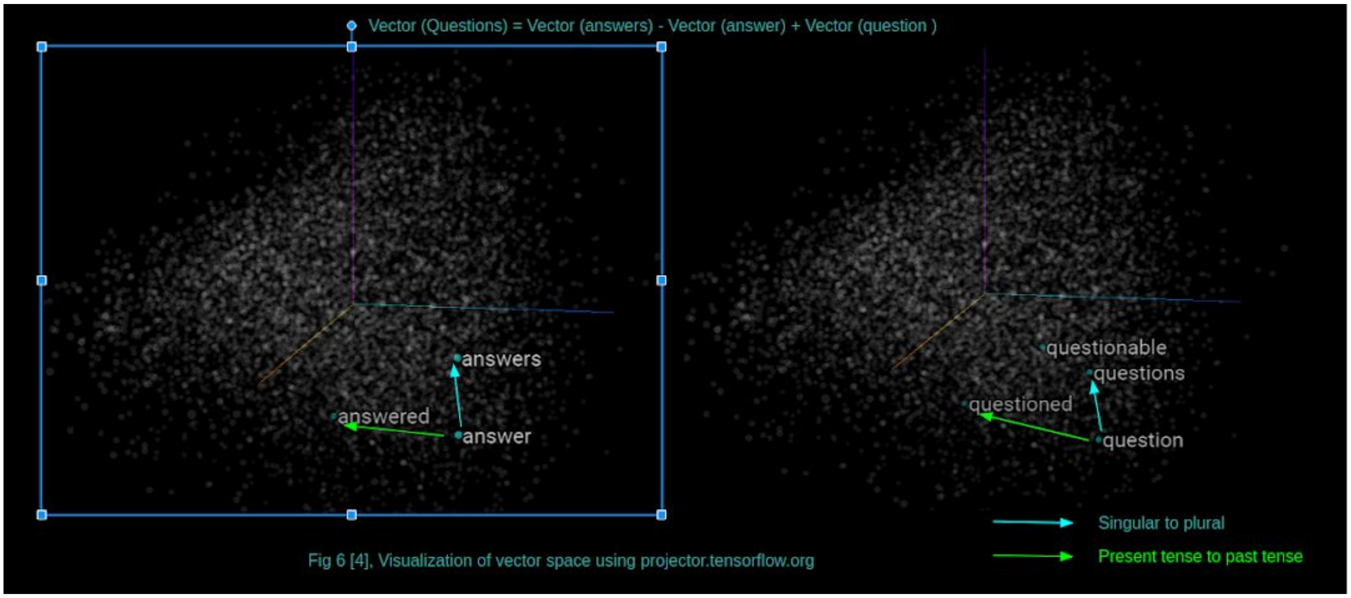

Fig 6 [4], Visualization of vector space using projector.tensorflow.org

2) Document similarity: **[4 marks]**
Use gensim Doc2Vec model to find the document similarity:
Data set link: https://archive.ics.uci.edu/ml/datasets/Twenty+Newsgroups

Take any document (doc1) from "comp.graphics" folder and find the document similarity with that document with one document each from other folders. You get 19 similarity scores. Now, take any 19 document from "comp.graphics" and find the similarity with doc1. Normalize the similarity scores and compare the values.

3) Working with tool Spacy: **[2 marks]**
a)Given any sentence or document as input, perform the following task:

Lemmatization , POS tagging

b)Given any sentence or document as input, perform Named Entity Recognition

.
c) Given two word as input measure the word similarity score .