Programming Assignment: Word Vectors

In [2]:
```
import numpy as np
```

In [4]:
```
filename = 'glove.6B.300d.txt'
with open (filename) as f :
    content = f.read().splitlines()
```

In [5]:
```
n = len(content)
vecs = np.zeros((n,300))
words = []
index = 0
```

In [6]:
```
for rawline in content :
    line = rawline.split()
    words.append(line[0])
    vecs[index] = np.genfromtxt(line[1:])
    index = index+1
```

In [22]:
```
k = 6 # Number of neighbors to find (including the word itself)
nn = NearestNeighbors(n_neighbors=k, metric='euclidean')
nn.fit(vecs)
```

Out[22]:
```
NearestNeighbors(metric='euclidean', n_neighbors=6)
```

In [23]:
```
# Dictionary to store nearest neighbors for each word
nn_dict = {}

# Words to find neighbors for
target_words = ['communism', 'africa', 'happy', 'sad', 'upset', 'computer',

for i in target_words:
    word_ind = words.index(i) #find index of current word (i)
    vector = vecs[word_ind].reshape(1, -1)  # get word V, and reshape for in
    dist, ind = nn.kneighbors(vector, k)  # get dist+ind of k neighbords of
    n_ind = ind[0][1:]  # exclude word by starting at 1
    n_words = [words[i] for i in n_ind]  # for each word i, we will get the
    nn_dict[i] = n_words
```

In [24]:
```
for w, n in nn_dict.items():
    print(f'Word: {w}, Nearest Neighbors: {n}')
```

```
Word: communism, Nearest Neighbors: ['fascism', 'capitalism', 'nazism', 'sta
linism', 'socialism']
Word: africa, Nearest Neighbors: ['african', 'continent', 'south', 'africans
', 'zimbabwe']
Word: happy, Nearest Neighbors: ['glad', 'pleased', 'always', 'everyone', 's
ure']
Word: sad, Nearest Neighbors: ['sorry', 'tragic', 'happy', 'pathetic', 'awfu
l']
Word: upset, Nearest Neighbors: ['upsetting', 'surprised', 'upsets', 'stunne
d', 'shocked']
Word: computer, Nearest Neighbors: ['computers', 'software', 'technology', '
laptop', 'computing']
Word: cat, Nearest Neighbors: ['cats', 'dog', 'pet', 'feline', 'dogs']
Word: dollar, Nearest Neighbors: ['currency', 'dollars', 'euro', 'multibilli
on', 'weaker']
```