In [3]:
```
print_me("Avanish Singh")
print_me("191550022")
```

Avanish Singh

191550022

In [4]:
```
d1 = "New York Times"
d2 = "New York Post"
d3 = "Los Angeles Times"
Q = "New New Times"
```

In [5]:
```
d1 = d1.split()
d2 = d2.split()
d3 = d3.split()
Q = Q.split()
```

In [6]:
```
unique_words = list(set(d1+d2+d3))
```

In [7]:
```
alldocs = [d1,d2,d3]
freq_docWords = {}
```

In [9]:
```
import math
for words in unique_words:
    count = 0
    for doc in alldocs:
        if(words in doc):
            count = count+1
    freq_docWords[words]= count
idf_value = {}
for words in unique_words:
    idf_value[words] = math.log(len(alldocs)/freq_docWords[words],10)
d1_r = {}
d2_r = {}
d3_r = {}
Q_r = {}
for i in unique_words:
    d1_r[i]= d1.count(i)*idf_value[i]
    d2_r[i]= d2.count(i)*idf_value[i]
    d3_r[i]= d3.count(i)*idf_value[i]
    Q_r[i]= Q.count(i)*idf_value[i]
sim_d1 = 0
sim_d2 = 0
sim_d3 = 0
for i in unique_words:
    sim_d1= sim_d1 + d1_r[i]*Q_r[i]
    sim_d2= sim_d2 + d2_r[i]*Q_r[i]
    sim_d3= sim_d3 + d3_r[i]*Q_r[i]

print("Similarity Coefficient of d1 ",sim_d1)
print("Similarity Coefficient of d2 ",sim_d2)
print("Similarity Coefficient of d3 ",sim_d3)
```

```
Similarity Coefficient of d1  0.09302439454744511
Similarity Coefficient of d2  0.062016263031630076
Similarity Coefficient of d3  0.031008131515815038
```

In [10]:
```
#COSINE SIMILARITY
```

In [11]:
```
doc1_freq = {}
doc2_freq = {}
doc3_freq = {}
Q_freq = {}
all_docs = [d1,d2,d3,Q]

freq_docWords = {}
```

In [12]:
```
for words in unique_words:
    doc1_freq[words]=d1.count(words)
    doc2_freq[words]=d2.count(words)
    doc3_freq[words]=d3.count(words)
    Q_freq[words]=Q.count(words)
cosine_Similarity=[]
doc_freq = [doc1_freq, doc2_freq, doc3_freq]
for doc in doc_freq:
    a=0
    b=0
    c=0
    for words in unique_words:
        c = c+ doc[words]*Q_freq[words]
        a = a+ (doc[words])**2
        b = b+ (Q_freq[words])**2
    cosine_Similarity.append(c/(math.sqrt(a)*math.sqrt(b)))

print("Cosine Similarity of d1 and Q ",cosine_Similarity[0])
print("Cosine Similarity of d2 and Q ",cosine_Similarity[1])
print("Cosine Similarity of d3 and Q ",cosine_Similarity[2])
```

```
Cosine Similarity of d1 and Q  0.7745966692414834
Cosine Similarity of d2 and Q  0.5163977794943222
Cosine Similarity of d3 and Q  0.2581988897471611
```

In [ ]:
```
```