In [1]:
```
import numpy as np
import pandas as pd
import nltk
from sklearn.feature_extraction.text import CountVectorizer, TfidfTransformer
from sklearn.neighbors import KDTree
person = pd.read_csv('famous_people.csv')
print(person.head())
```

```
   Sno                                          URI               Name  \
0    1      https://en.wikipedia.org/wiki/Narendra_Modi      Narendra Modi
1    2        https://en.wikipedia.org/wiki/Sania_Mirza        Sania Mirza
2    3        https://en.wikipedia.org/wiki/Virat_Kohli        Virat Kohli
3    4        https://en.wikipedia.org/wiki/Miley_Cyrus        Miley Cyrus
4    5  https://en.wikipedia.org/wiki/Leonardo_DiCaprio  Leonardo DiCaprio

                                                Text
0  Narendra Damodardas Modi (Gujarati pronunciati...
1  Sania Mirza (Hindustani pronunciation: ['saːnɪ...
2  Virat Kohli (About this soundpronunciation (he...
3  Miley Ray Cyrus (born Destiny Hope Cyrus; Nove...
4  Leonardo Wilhelm DiCaprio (/dɪˈkæpriːoʊ/, Itali...
```

In [2]:
```
count_vector = CountVectorizer()
train_counts = count_vector.fit_transform(person.Text)
tfidf_transform = TfidfTransformer()
train_tfidf = tfidf_transform.fit_transform(train_counts)
a = np.array(train_tfidf.toarray())
kdtree = KDTree(a ,leaf_size=3)
person_name=input("Enter the name of the Person:- ")
person['tfidf']=list(train_tfidf.toarray())
distance, idx = kdtree.query(person['tfidf'][person['Name']== person_name].tolist(
for i, value in list(enumerate(idx[0])):
    print("Name : {}".format(person['Name'][value]))
    print("Distance : {}".format(distance[0][i]))
    print("URI : {}".format(person['URI'][value]))
```

```
Enter the name of the Person:- MS Dhoni
Name : MS Dhoni
Distance : 0.0
URI : https://en.wikipedia.org/wiki/MS_Dhoni
Name : Virat Kohli
Distance : 0.8497822049958499
URI : https://en.wikipedia.org/wiki/Virat_Kohli
Name : Yuvraj Singh
Distance : 0.9684791160176293
URI : https://en.wikipedia.org/wiki/Yuvraj_Singh
```

In [ ]: