In [41]:

```
import pandas as pd
import numpy as np
from sklearn.feature_extraction.text import CountVectorizer
from sklearn.metrics.pairwise import cosine_similarity

df = pd.read_csv("mo.csv")
```

In [52]:

```
df.head()
```

Out[52]:

	index	budget	genres	homepage	id	keywords	original_language	original_title	overview
0	0	237000000	Action Adventure Fantasy Science Fiction	http://www.avatarmovie.com/	19995	culture clash future space war space colony so...	en	Avatar	In the 22nd century, a paraplegic Marine is di...
1	1	300000000	Adventure Fantasy Action	http://disney.go.com/disneypictures/pirates/	285	ocean drug abuse exotic island east india trad...	en	Pirates of the Caribbean: At World's End	Captain Barbossa, long believed to be dead, ha...
2	2	245000000	Action Adventure Crime	http://www.sonypictures.com/movies/spectre/	206647	spy based on novel secret agent sequel mi6	en	Spectre	A cryptic message from Bond’s past sends him o...
3	3	250000000	Action Crime Drama Thriller	http://www.thedarkknightrises.com/	49026	dc comics crime fighter terrorist secret ident...	en	The Dark Knight Rises	Following the death of District Attorney Harve...
4	4	260000000	Action Adventure Science Fiction	http://movies.disney.com/john-carter	49529	based on novel mars medallion space travel pri...	en	John Carter	John Carter is a war-weary, former military ca...

5 rows × 25 columns

In [53]:

```
features = ['keywords','cast','genres','director']
```

In [54]:

```
def combine_features(row):
    return row['keywords']+" "+row['cast']+" "+row['genres']+" "+row['director']
```

In [44]:

```
for feature in features:
    df[feature] = df[feature].fillna('') #filling all NaNs with blank string

df["combined_features"] = df.apply(combine_features,axis=1) #applying combined_features() method
over each rows of dataframe and storing the combined string in "combined_features" column
```

In [45]:

```
df.iloc[0].combined_features
```

Out[45]:

```
'culture clash future space war space colony society Sam Worthington Zoe Saldana Sigourney Weaver
Stephen Lang Michelle Rodriguez Action Adventure Fantasy Science Fiction James Cameron'
```

In [46]:

```
cv = CountVectorizer() #creating new CountVectorizer() object
count_matrix = cv.fit_transform(df["combined_features"])
```

In [47]:

```
cosine_sim = cosine_similarity(count_matrix)
```

In [48]:

```
def get_title_from_index(index):
    return df[df.index == index]["title"].values[0]
def get_index_from_title(title):
    return df[df.title == title]["index"].values[0]
```

In [55]:

```
movie_user_likes = "Pirates of the Caribbean: At World's End"
movie_index = get_index_from_title(movie_user_likes)
similar_movies = list(enumerate(cosine_sim[movie_index]))
```

In [56]:

```
sorted_similar_movies = sorted(similar_movies,key=lambda x:x[1],reverse=True)[1:]
```

In [57]:

```
i=0
print("Top 5 similar movies to "+movie_user_likes+" are:\n")
for element in sorted_similar_movies:
    print(get_title_from_index(element[0]))
    i=i+1
    if i>5:
        break
```

```
Top 5 similar movies to Pirates of the Caribbean: At World's End are:

Pirates of the Caribbean: The Curse of the Black Pearl
Pirates of the Caribbean: Dead Man's Chest
Spider-Man 3
Hancock
Spider-Man 2
Anna and the King
```

In [ ]:

```

```

In [ ]:

```

```

In [ ]:

```

```

In [ ]:

```
```