In [1]:

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

In [12]:

```python
#Helping functions used later
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 [5]:

```python
#Read CSV file
df = pd.read_csv('movie_dataset.csv')
df.columns
```

Out[5]:

```
Index(['index', 'budget', 'genres', 'homepage', 'id', 'keywords',
       'original_language', 'original_title', 'overview', 'popularity',
       'production_companies', 'production_countries', 'release_date',
       'revenue', 'runtime', 'spoken_languages', 'status', 'tagline', 'tit
le',
       'vote_average', 'vote_count', 'cast', 'crew', 'director'],
      dtype='object')
```

In [6]:

```python
#Select features
features = ['keywords','cast','genres','director']
```

In [9]:

```python
#Create a column in df which combines all the features
for feature in features:
    df[feature]=df[feature].fillna("")

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

df["combined_features"]=df.apply(combine_features,axis=1)
df["combined_features"].head()
```

Out[9]:

```
0    culture clash future space war space colony so...
1    ocean drug abuse exotic island east india trad...
2    spy based on novel secret agent sequel mi6 Dan...
3    dc comics crime fighter terrorist secret ident...
4    based on novel mars medallion space travel pri...
Name: combined_features, dtype: object
```

In [10]:

```python
#Create count matrix from this new combined column
cv = CountVectorizer()
count_matrix= cv.fit_transform(df['combined_features'])
```

In [11]:

```python
#Compute cosine similarity based on count_matrix
cosine_sim =  cosine_similarity(count_matrix)
```

In [32]:

```python
#Get the index of the movie based on its title
movieUserLikes='Avatar'
movie_index=get_index_from_title(movieUserLikes)

similar_movies= list(enumerate(cosine_sim[movie_index]))

#Get a list of similar movies in a list of similar movies in descending order
sorted_similar_movies= sorted(similar_movies,key=lambda x:x[1],reverse=True)
```

In [33]:

```python
#Print tites of first 50 movies
i=0
for movie in sorted_similar_movies:
    print(get_title_from_index(movie[0]))
    i=i+1
    if i>50:
        break
```

```python
#Print tites of first 50 movies
i=0
for movie in sorted_similar_movies:
    print(get_title_from_index(movie[0]))
```

Avatar
Guardians of the Galaxy
Aliens
Star Wars: Clone Wars: Volume 1
Star Trek Into Darkness
Star Trek Beyond
Alien
Lockout
Jason X
The Helix... Loaded
Moonraker
Planet of the Apes
Galaxy Quest
Gravity
Alien³
Jupiter Ascending
The Wolverine
Silent Running
Zathura: A Space Adventure
Trekkies
Cargo
Wing Commander
Star Trek
Lost in Space
Babylon A.D.
The Fifth Element
Oblivion
Titan A.E.
AVP: Alien vs. Predator
The Empire Strikes Back
Dragonball Evolution
Superman Returns
Divergent
John Carter
The Black Hole
The Ice Pirates
Memoirs of an Invisible Man
Starship Troopers
The Astronaut's Wife
Machete Kills
Soldier
The Abyss
Damnation Alley
Men in Black
Space Cowboys
Space Dogs
The Time Machine
Sheena
Captain America: Civil War
Star Trek: Insurrection
Oz: The Great and Powerful