In [1]:

```
import pandas as pd
```

In [2]:

```
hashtags_frequency = pd.read_csv(r'C:\Users\daksh\Desktop\SMA Tweets\wordcloud.csv')
hashtags_frequency
```

Out[2]:

	word	count
0	ukraineconflict	1184
1	ukraine	629
2	russia	416
3	ukrainecrisis	299
4	russiawar	225
5	ukrainerussia	205
6	ukrainewar	176
7	kyiv	171
8	ukrainerussiacrisis	168
9	putin	101
10	ukraineunderattack	79
11	russiaukrainewar	75
12	nato	74
13	russiaukraine	69
14	ukraineinvasion	66
15	eu	61
16	europe	58
17	russianinvasion	46
18	us	42

In [3]:

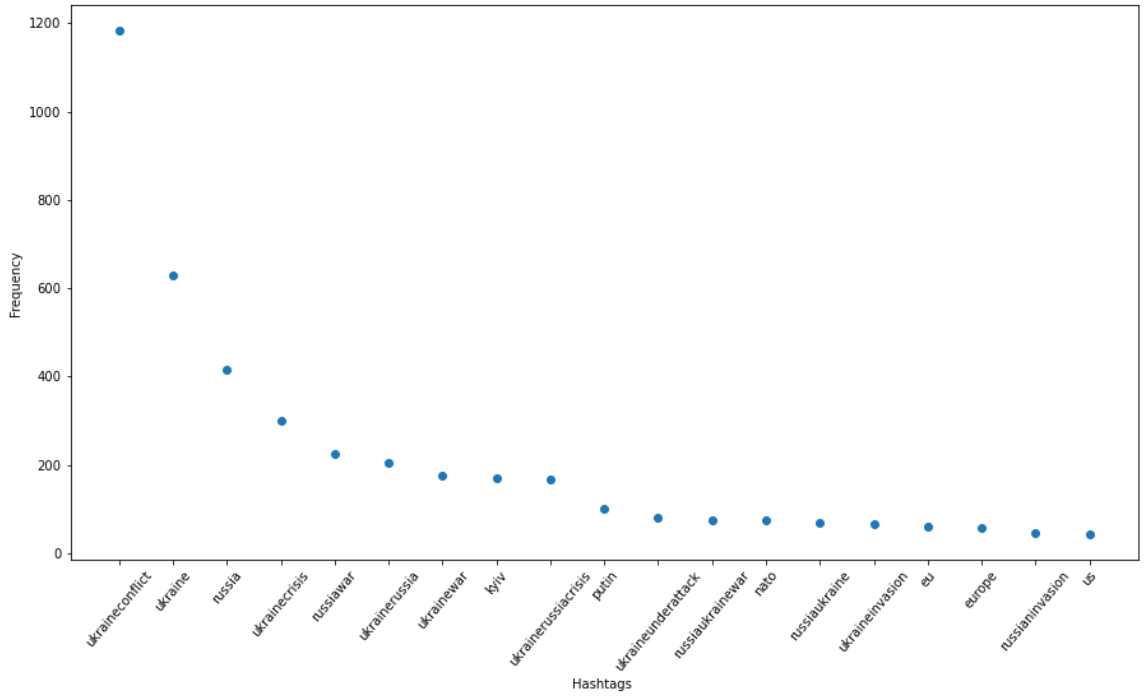
```
import matplotlib.pyplot as plt
from sklearn.cluster import KMeans
```

```
C:\Users\daksh\anaconda3\lib\site-packages\scipy\__init__.py:138: UserWarn
ing: A NumPy version >=1.16.5 and <1.23.0 is required for this version of
SciPy (detected version 1.24.2)
  warnings.warn(f"A NumPy version >={np_minversion} and <{np_maxversion} i
s required for this version of "
```

In [4]:

```
x = hashtags_frequency['word']
y = hashtags_frequency['count']

plt.figure(figsize=(15,8))
plt.xlabel("Hashtags ", rotation='horizontal')
plt.ylabel("Frequency ")
plt.xticks(rotation = 50)
plt.scatter(x, y)
plt.show()
```



In [5]:

```
count = []
for i in range(1,20):
    count.append(i)
hashtags_frequency['Index'] = count
```

In [6]:

```
hashtags_frequency
```

Out[6]:

	word	count	Index
0	ukraineconflict	1184	1
1	ukraine	629	2
2	russia	416	3
3	ukrainecrisis	299	4
4	russiawar	225	5
5	ukrainerussia	205	6
6	ukrainewar	176	7
7	kyiv	171	8
8	ukrainerussiacrisis	168	9
9	putin	101	10
10	ukraineunderattack	79	11
11	russiaukrainewar	75	12
12	nato	74	13
13	russiaukraine	69	14
14	ukraineinvasion	66	15
15	eu	61	16
16	europe	58	17
17	russianinvasion	46	18
18	us	42	19

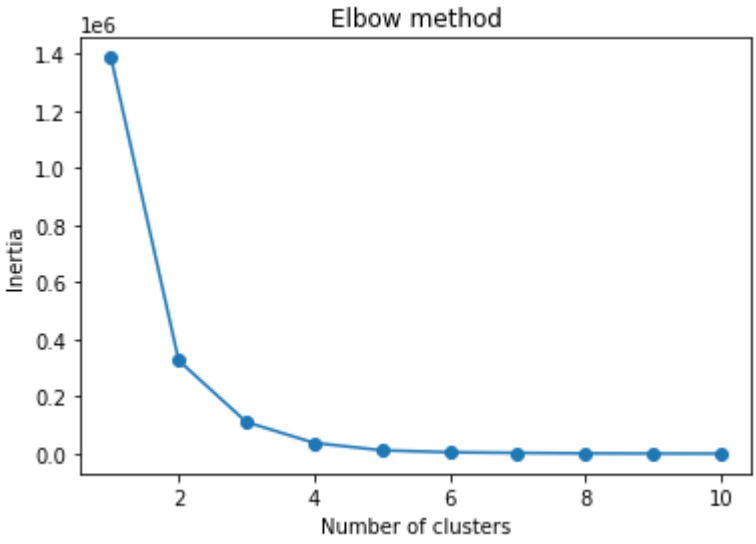
In [12]:

```
x = hashtags_frequency['Index']
y = hashtags_frequency['count']
data = list(zip(x, y))
inertias = []
# print(data)

for i in range(1,11):
    kmeans = KMeans(n_clusters=i)
    kmeans.fit(data)
    inertias.append(kmeans.inertia_)

plt.plot(range(1,11), inertias, marker='o')
plt.title('Elbow method')
plt.xlabel('Number of clusters')
plt.ylabel('Inertia')
plt.show()
```

```
C:\Users\daksh\anaconda3\lib\site-packages\sklearn\cluster\_kmeans.py:881:
UserWarning: KMeans is known to have a memory leak on Windows with MKL, wh
en there are less chunks than available threads. You can avoid it by setti
ng the environment variable OMP_NUM_THREADS=1.
  warnings.warn(
```



In [18]:

```
kmeans = KMeans(n_clusters=3)
kmeans.fit(data)

x = hashtags_frequency['word']
plt.figure(figsize=(15,8))
plt.xlabel("Hashtags ", rotation='horizontal')
plt.ylabel("Frequency ")
plt.xticks(rotation = 50)
plt.scatter(x, y, c=kmeans.labels_)
plt.show()
```