In [16]:

```
pip install rank_bm25
```

```
Looking in indexes: https://pypi.org/simple, https://us-python.pkg.dev/colab-wheels/publi
c/simple/
Collecting rank_bm25
  Downloading rank_bm25-0.2.2-py3-none-any.whl (8.6 kB)
Requirement already satisfied: numpy in /usr/local/lib/python3.8/dist-packages (from rank
_bm25) (1.21.6)
Installing collected packages: rank-bm25
Successfully installed rank-bm25-0.2.2
```

In [24]:

```
print(file_subset[2001])
```

```
/content/drive/MyDrive/tweet_files/22916.txt
```

In [3]:

```
from google.colab import drive
drive.mount('/content/drive')
```

```
Mounted at /content/drive
```

In [34]:

```
# f=file_subset[2000]
f = open(file_subset[2000], 'r')

content = f.read()
tokenized_corpus = content.split(" ")
print(tokenized_corpus)
f.close()
```

```
["'#WorldPrematurityDay", 'my', 'little', 'miracle', 'born', 'at', '36', 'weeks', 'due',
'to', 'me', 'having', '#preeclampsia', '\\xf0\\x9f\\x92\\x99', "\n'SWACHNGO'"]
```

In [39]:

```
import os
from nltk.tokenize import word_tokenize

# List the files in the directory
# file_list = os.listdir('/content/drive/MyDrive/tweet_files')
data_dir = '/content/drive/MyDrive/tweet_files/'
DATA_SET_DIR = data_dir
print('\nGetting List of text files from' + DATA_SET_DIR)
files = os.listdir(DATA_SET_DIR)
print((files[11]))
print('\nFile list retrieved from ' + DATA_SET_DIR)
```

```
Getting List of text files from/content/drive/MyDrive/tweet_files/
414.txt

File list retrieved from /content/drive/MyDrive/tweet_files/
```

In [70]:

```
from typing_extensions import Concatenate
corpus = []
for f in files:
  strm = open(DATA_SET_DIR + f,'r')
  content = strm.read()
  words = content.split(" ")
  corpus.append(' '.join(words))
  strm.close()
```

In [71]:

```
print(corpus[11])
```

```
'This will be the best and i really love this i want this one so badly \n#JeepGrandCherok
ee #OIIIIIIIO
'PatelDepika00'
```

In [83]:

```
from rank_bm25 import BM25Okapi #imp copy from here
def bm25_precision10(query, corpus):
    tokenized_corpus = [doc.split(" ") for doc in corpus]
    bm25 = BM25Okapi(tokenized_corpus)
    tokenized_query = query.split(" ")
    doc_scores = bm25.get_scores(tokenized_query)
    print("Top relevant docs are :")
    print(bm25.get_top_n(tokenized_query, corpus, n=5))
    print("Top 10 relevant docs in the query :")
    docscores = list(doc_scores)
    topten = sorted(range(len(docscores)), key=lambda i: docscores[i], reverse=True)[:10
]
    print(topten)

bm25_precision10("football",corpus)
```

```
Top relevant docs are :
['" only inter-school rink football tournament for girls U-16... Friday, Nov 18, 11am\\xe
2\\x80\\xa6 \n\'mridula2c\'', '"Roads turned into football arena during morning leg of th
e Maharashtra leg.\\nA fresh start to today\'s Padyatra.\\nOur\\xe2\\x80\\xa6 \n\'DrMusta
faAli81\'', '"Roads turned into football arena during morning leg of the Maharashtra leg.
\\nA fresh start to today\'s Padyatra.\\nOur\\xe2\\x80\\xa6 \n\'galshikla\'', '"#INDvsNZ\
\nMan\'s IPL has come, Women\'s IPL has come, Now football IPL should come.\\n
\\xe2\\x9a\\xbd\\xef\\xb8\\x8f \\xe2\\x9a\\xbd\\xef\\xb8\\x8f \\xe2\\x9a\\xbd\\xef\\xb8\\
x8f \\xe2\\x9a\\xbd\\xef\\xb8\\x8f"\n\'Sportskeeda\'', "'Congrulate @ShemarooEnt fr  cele
brating ur 60th Anniversary today! #Shemaroo60YearsYoung \n'isakshi12'"]
Top 10 relevant docs in the query :
[7663, 6048, 6052, 1475, 0, 1, 2, 3, 4, 5]
```

In [84]:

```
pip install PySimpleGUI
```

```
Looking in indexes: https://pypi.org/simple, https://us-python.pkg.dev/colab-wheels/publi
c/simple/
Collecting PySimpleGUI
  Downloading PySimpleGUI-4.60.4-py3-none-any.whl (509 kB)
     |████████████████████████████████| 509 kB 8.9 MB/s
Installing collected packages: PySimpleGUI
Successfully installed PySimpleGUI-4.60.4
```

In [89]:

```
import matplotlib
matplotlib.use('Agg')
```

In [ ]:

```
import PySimpleGUI as sg

layout = [
    [sg.VPush()],
    [sg.Text("Search: "), sg.Input(key='INPUT')],
    [sg.Ok()],
    [sg.Text("", size=(0, 1), key='OUTPUT'),  ],
    [sg.VPush()],
]

window = sg.Window("Tweets Database", layout, size=(1400, 600), element_justification='c
',background_color='white')
```

```
while True:
    event, values = window.read()
    if event == sg.WINDOW_CLOSED:
        break
    elif event == 'Ok':
        name = values['INPUT']
        tokenized_query = name.split(" ")
        doc_scores = bm25.get_scores(tokenized_query)
        window['OUTPUT'].update(value=bm25.get_top_n(tokenized_query, corpus, n=1))

window.close()
```

In [92]:

```
pip install aiml
```

```
Looking in indexes: https://pypi.org/simple, https://us-python.pkg.dev/colab-wheels/publi
c/simple/
Collecting aiml
  Downloading aiml-0.9.2-py2.py3-none-any.whl (2.1 MB)
     |████████████████████████████████| 2.1 MB 7.5 MB/s
Requirement already satisfied: setuptools in /usr/local/lib/python3.8/dist-packages (from
aiml) (57.4.0)
Installing collected packages: aiml
Successfully installed aiml-0.9.2
```

In [97]:

```
bm25 = BM25Okapi(corpus)
scores = bm25.get_scores("football")
ranked_files = sorted(zip(file_list, scores), key=lambda x: x[1], reverse=True)
for file, score in ranked_files[:10]:
  print(f'{file}: {score}')
```

```
/content/drive/MyDrive/tweet_files/9049.txt: 3.5242713777647525
/content/drive/MyDrive/tweet_files/9044.txt: 3.52244957956026
/content/drive/MyDrive/tweet_files/8858.txt: 3.5214510814521356
/content/drive/MyDrive/tweet_files/6610.txt: 3.519589095901893
/content/drive/MyDrive/tweet_files/7523.txt: 3.513014478679762
/content/drive/MyDrive/tweet_files/22944.txt: 3.5117797323135287
/content/drive/MyDrive/tweet_files/18506.txt: 3.511468465610529
/content/drive/MyDrive/tweet_files/19431.txt: 3.5047138096368933
/content/drive/MyDrive/tweet_files/18728.txt: 3.5034764789201676
/content/drive/MyDrive/tweet_files/8904.txt: 3.503263648604668
```

In [ ]:

```

```