

# AIWIR ASSIGNMENT-3

SEC: 6E

Submitted by:

Name	SRN
Rishab Kshatri	PES2UG19CS327
Rahul S Bhat	PES2UG19CS315
Rahul Radhakrishna	PES2UG19CS314

## Wildcard query retrieval using inverted index

CODE:

```
In [1]: import pandas
from nltk.tokenize import sent_tokenize
from nltk.tokenize import word_tokenize
from nltk.stem import PorterStemmer
from nltk.corpus import stopwords
import string
import time
stemmer = PorterStemmer()
```

```
In [2]: data=pandas.read_csv("C:/Users/Rahul Bhat/Desktop/data.csv")
data.head()
```

```
Out[2]:
```

	Sentence	Sentiment
0	The GeoSolutions technology will leverage Bene...	positive
1	ESION lows, down1.50 to \$2.50 BK a real po...	negative
2	For the last quarter of 2010 , Componenta 's n...	positive
3	According to the Finnish-Russian Chamber of Co...	neutral
4	The Swedish buyout firm has sold its remaining...	neutral

```
In [3]: sentences = []
words = []
filtered = []
stop_words = set(stopwords.words('english'))
print(stop_words)

{'haven', 'it's', 'ain', 'can', 'an', 'having', 'a', 'ourselves', 'by', 'against', 'wasn', 'mightn't', 'about', 'furt
her', 'no', 'where', 'doing', 'yourself', 'are', 'me', 'that'll', 'did', 'as', 'before', 'most', 'nor', 'd', 'aren'
t', 'not', 'his', 'has', 'is', 'or', 'only', 'this', 'you'd', 'when', 'own', 'do', 'how', 'been', 'she', 'anv', 'betw
```

```
In [4]: for i in data.Sentence:
        i = i.translate(str.maketrans('', '', string.punctuation))
        a = sent_tokenize(i)
        sentences.append(a)
        b = word_tokenize(i)
        words.append(b)
        c = [w.lower() for w in b if not w.lower() in stop_words]
        filtered.append(c)
data.head()
```

```
Out[4]:
```

	Sentence	Sentiment
0	The GeoSolutions technology will leverage Bene...	positive
1	ESION lows, down1.50 to \$2.50 BK a real po...	negative
2	For the last quarter of 2010 , Componenta 's n...	positive
3	According to the Finnish-Russian Chamber of Co...	neutral
4	The Swedish buyout firm has sold its remaining...	neutral

```
In [5]: data['sentence_tokenized'] = sentences
data['word_tokenized'] = words
data['filtered'] = filtered
data.head()
```

Out[5]:

	Sentence	Sentiment	sentence_tokenized	word_tokenized	filtered
0	The GeoSolutions technology will leverage Bene...	positive	[The GeoSolutions technology will leverage Ben...	[The, GeoSolutions, technology, will, leverage...	[geosolutions, technology, leverage, benefon, ...
1	<i>ESI</i> on lows, down1.50 to \$2.50 BK a real po...	negative	[ESI on lows down 150 to 250 BK a real possibi...	[ESI, on, lows, down, 150, to, 250, BK, a, rea...	[esi, lows, 150, 250, bk, real, possibility]
2	For the last quarter of 2010 , Componenta's n...	positive	[For the last quarter of 2010 Componenta s ne...	[For, the, last, quarter, of, 2010, Componenta...	[last, quarter, 2010, componenta, net, sales, ...
3	According to the Finnish-Russian Chamber of Co...	neutral	[According to the FinnishRussian Chamber of Co...	[According, to, the, FinnishRussian, Chamber, ...	[according, finnishrussian, chamber, commerce,...
4	The Swedish buyout firm has sold its remaining...	neutral	[The Swedish buyout firm has sold its remainin...	[The, Swedish, buyout, firm, has, sold, its, r...	[swedish, buyout, firm, sold, remaining, 224, ...

```
In [6]: data['stemmed_words']=data['filtered'].apply(lambda f: [stemmer.stem(word) for word in f])
data.head()
```

Out[6]:

	Sentence	Sentiment	sentence_tokenized	word_tokenized	filtered	stemmed_words
0	The GeoSolutions technology will leverage Bene...	positive	[The GeoSolutions technology will leverage Ben...	[The, GeoSolutions, technology, will, leverage...	[geosolutions, technology, leverage, benefon, ...	[geosolut, technolog, leverag, benefon, gp, so...
1	<i>ESI</i> on lows, down1.50 to \$2.50 BK a real po...	negative	[ESI on lows down 150 to 250 BK a real possibi...	[ESI, on, lows, down, 150, to, 250, BK, a, rea...	[esi, lows, 150, 250, bk, real, possibility]	[esi, low, 150, 250, bk, real, possibl]
2	For the last quarter of 2010 , Componenta's n...	positive	[For the last quarter of 2010 Componenta s ne...	[For, the, last, quarter, of, 2010, Componenta...	[last, quarter, 2010, componenta, net, sales, ...	[last, quarter, 2010, componenta, net, sale, d...

```
In [9]: inv_index = {}
permuterm_index={}

for pos,term in enumerate(data['stemmed_words']):
    for word in term:
        if word in inv_index:
            if pos not in inv_index[word]:
                inv_index[word].append(pos)
        else:
            inv_index[word]=[]
            inv_index[word].append(pos)
            permuterm=set()
            w=word+"$"
            for i in range(len(word)):
                w=w[1:] + w[0:1]
                permuterm.add(w)
            permuterm_index[word]=[permuterm]

for i in inv_index:
    print(i,":",inv_index[i], end='\n\n')
    print(i,":",permuterm_index[i], end='\n\n')
```

geosolut : [0, 412]

geosolut : [{'lut\$geoso', '\$geosolut', 'ut\$geosol', 'olut\$geos', 'eosolut\$g', 'solut\$geo', 'osolut\$ge', 't\$geosolu'}]

technolog : [0, 59, 109, 137, 300, 412, 427, 436, 473, 558, 657, 672, 679, 682, 712, 720, 757, 758, 823, 849, 941, 990, 1018, 1105, 1124, 1215, 1329, 1462, 1499, 1516, 1597, 1600, 1629, 1638, 1689, 1719, 1724, 1879, 1925, 1941, 1970, 2072, 2129, 2166, 2176, 2216, 2263, 2357, 2484, 2487, 2534, 2675, 2696, 2752, 2828, 2936, 3031, 3056, 3067, 3079, 3085, 3181, 3237, 3238, 3242, 3248, 3272, 3308, 3353, 3484, 3487, 3510, 3564, 3685, 3993, 4048, 4082, 4093, 4110, 4116, 4140, 4145, 4149, 4158, 4256, 4341, 4439, 4449, 4481, 4497, 4515, 4578, 4615, 4815, 4823, 4864, 4869, 5084, 5155, 5162, 5167, 5191, 5271, 5288, 5329, 5342, 5389, 5410, 5496, 5497, 5534, 5566, 5604, 5709, 5780, 5816]

```
In [10]: query=input().lower()
per_results=[]
results=[]
start_time=time.time()
try:
    if('*' in query):
        query=query+'$'
        while(query[-1]!='*'):
            query=query[1:]+query[0:1]
        if(query.count('*')>1):
            query=query[:query.index('*')]+'*'

        for i in permuterm_index:
            match=[]
            string=""
            for j in permuterm_index[i][0]:
                if(query[-2] in j):
                    per_results.append(i)
                    break
        else:
            print(inv_index[query])
except:
    print("No Match Found")

else:
    for i in per_results:
        results+=inv_index[i]
    for i in set(results):
        print(data.Sentence[i])
finally:
    print("Query Processing time= ", time.time()-start_time)
```

## OUTPUT

### 1. Beginning in loca

loca\*  
The GeoSolutions technology will leverage Benefon 's GPS solutions by providing Location Based Search Technology , a Communities Platform , location relevant multimedia content and a new and powerful commercial model .  
As much biomass will be used as can be sourced locally , with the rest of the plant 's fuel needs met by peat .  
Since the association 's data do not cover sales figures from about 100 small local breweries and sales of imported beer products , the actual market shares of its members are smaller than those given in the report .  
In the Baltic countries , sales fell by 40.2 % , and in Russia , by 23.2 % in terms of euros , and by 10.7 % in terms of local currency .  
Their offering also covers localisation services related to production transfer to the Finnish company 's customers that already have production in Asian market or have made the decision to transfer production there .  
According to Finnish insurance companies Tapiola , Local Insurance Group ( L+Ähivakuutus ) , and Pohjois the two fierce storms at the end of July and in the beginning of August 2010 that felled trees around Finland have speeded up sales of forest insurance .  
The official opening of the office , located in Prague , will be celebrated on May 20 .  
The agreement will provide The Switch with double the converter capacity , whilst opening up further m

### 2. Beginning in loc

\$SBUX down PM, from \$DB downgrade.. PT cut from 70 to \$64  
Swedish , Finnish and Danish listed companies are organized in three market cap segments , Nordic Large Cap , Mid Cap and Small Cap .  
Finnish developer and manufacturer of mobile phone chargers Salcomp Plc OMX Helsinki : SAL1V on Wednesday 19 November lowered its full-year net sales estimate .  
Operating profit fell to EUR 38.1 mn from EUR 55.3 mn in 2007 .  
In 2008 , Kemira recorded revenue of approximately EUR 2.8 billion and had a staff of 9,400 .  
Investments in product development stood at 6.0 mln euro ( \$ 8.8 mln ) .  
HSBC Says Unit to Book \$585 Million Charge on Settlement  
Daily Mail parent company in talks with potential partners about joint bid for Yahoo's internet assets : <https://t.co/mK7MEMtADT> \$YHOO  
RISING costs have forced packaging producer Huhtamaki to axe 90 jobs at its Hampshire manufacturing plant .  
Nordic Walking was first used as a summer training method by cross-country skiers .  
According shipping company Viking Line , the EU decision will have a significant financial impact .  
In the building and home improvement trade , sales decreased by 22.5 % to EUR 201.4 mn .  
HELSINKI AFX - KCI Konecranes said it has won an order for four hot metal ladle cranes from Indian steel producer Bhushan Steel and Strips Ltd. .  
Query Processing time= 1.3742313385009766

f 1:

Dataset:

<https://www.kaggle.com/datasets/sbhatti/financial-sentiment-analysis>