AIWIR ASSIGNMENT-1

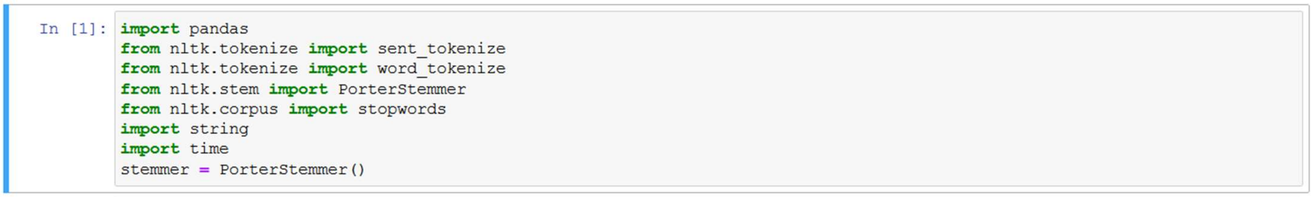
SEC: 6E

Submitted by:

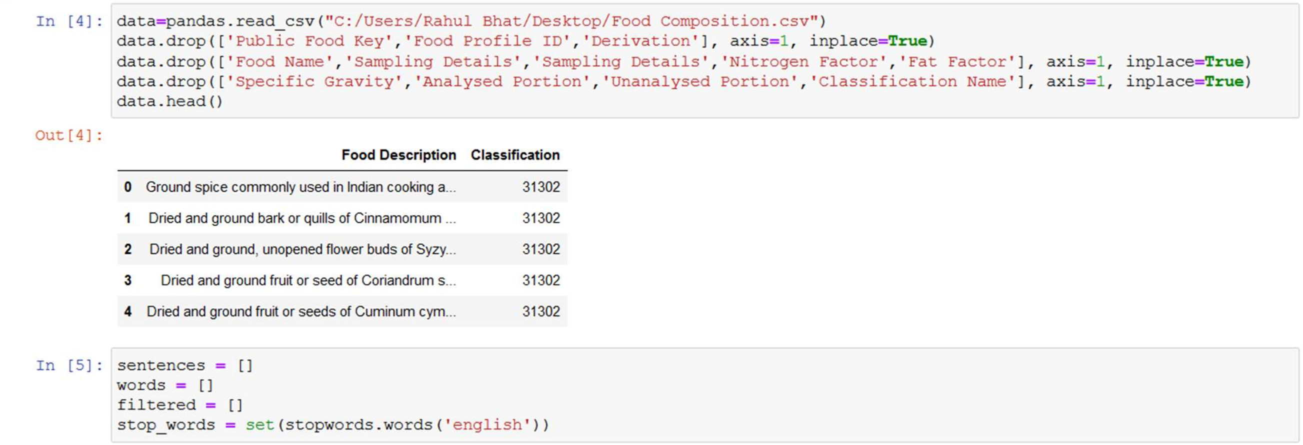
Name	SRN
Rishab Kshatri	PES2UG19CS327
Rahul S Bhat	PES2UG19CS315
Rahul Radhakrishna	PES2UG19CS314

Boolean 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()
```

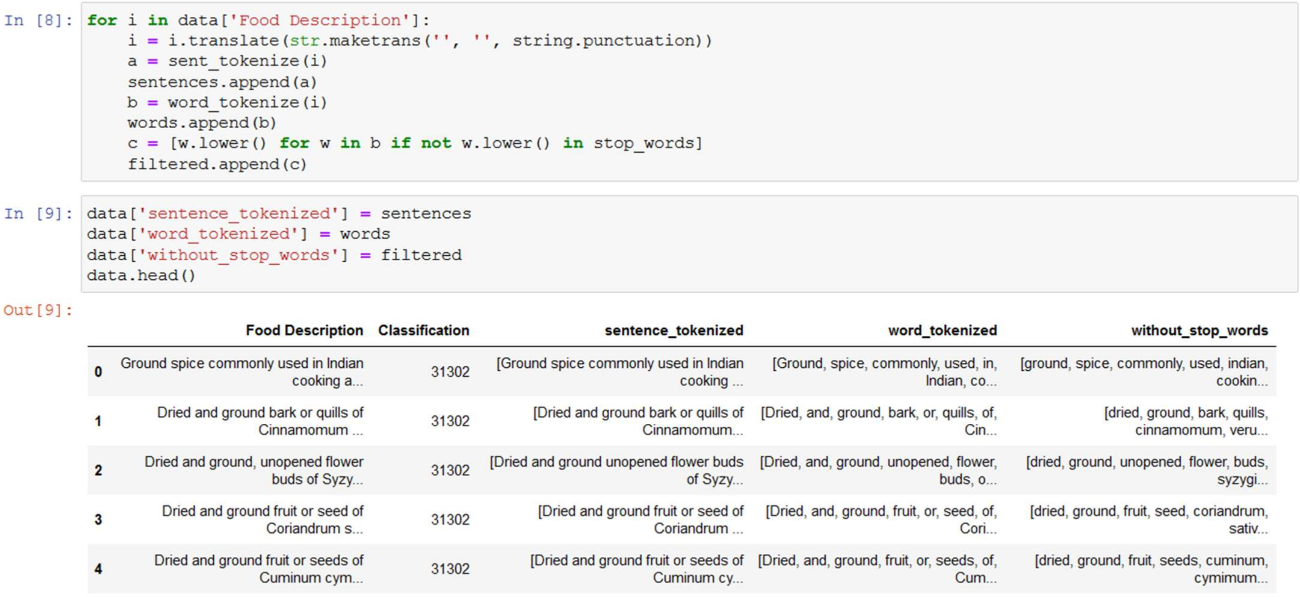


```
In [4]: data=pandas.read_csv("C:/Users/Rahul Bhat/Desktop/Food Composition.csv")
        data.drop(['Public Food Key','Food Profile ID','Derivation'], axis=1, inplace=True)
        data.drop(['Food Name','Sampling Details','Sampling Details','Nitrogen Factor','Fat Factor'], axis=1, inplace=True)
        data.drop(['Specific Gravity','Analysed Portion','Unanalysed Portion','Classification Name'], axis=1, inplace=True)
        data.head()
```

Out[4]:

	Food Description	Classification
0	Ground spice commonly used in Indian cooking a...	31302
1	Dried and ground bark or quills of Cinnamomum ...	31302
2	Dried and ground, unopened flower buds of Syzy...	31302
3	Dried and ground fruit or seed of Coriandrum s...	31302
4	Dried and ground fruit or seeds of Cuminum cym...	31302

```
In [5]: sentences = []
        words = []
        filtered = []
        stop_words = set(stopwords.words('english'))
```

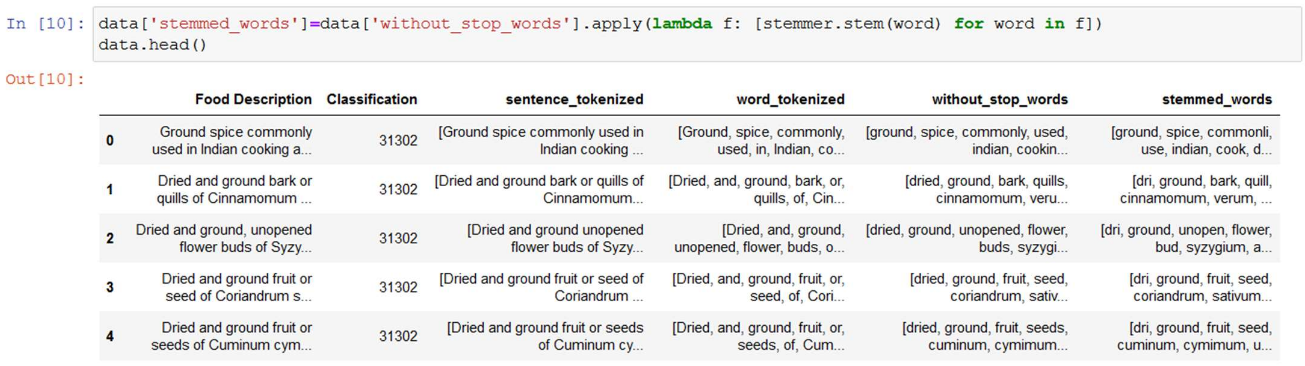


```
In [8]: for i in data['Food Description']:
            i = i.translate(str.maketrans('', '', string.punctuation))
            a = sent_tokenize(i)
            sentences.append(a)
            b = word_tokenize(i)
            words.append(b)
            c = [w.lower() for w in b if not w.lower() in stop_words]
            filtered.append(c)
```

```
In [9]: data['sentence_tokenized'] = sentences
        data['word_tokenized'] = words
        data['without_stop_words'] = filtered
        data.head()
```

Out[9]:

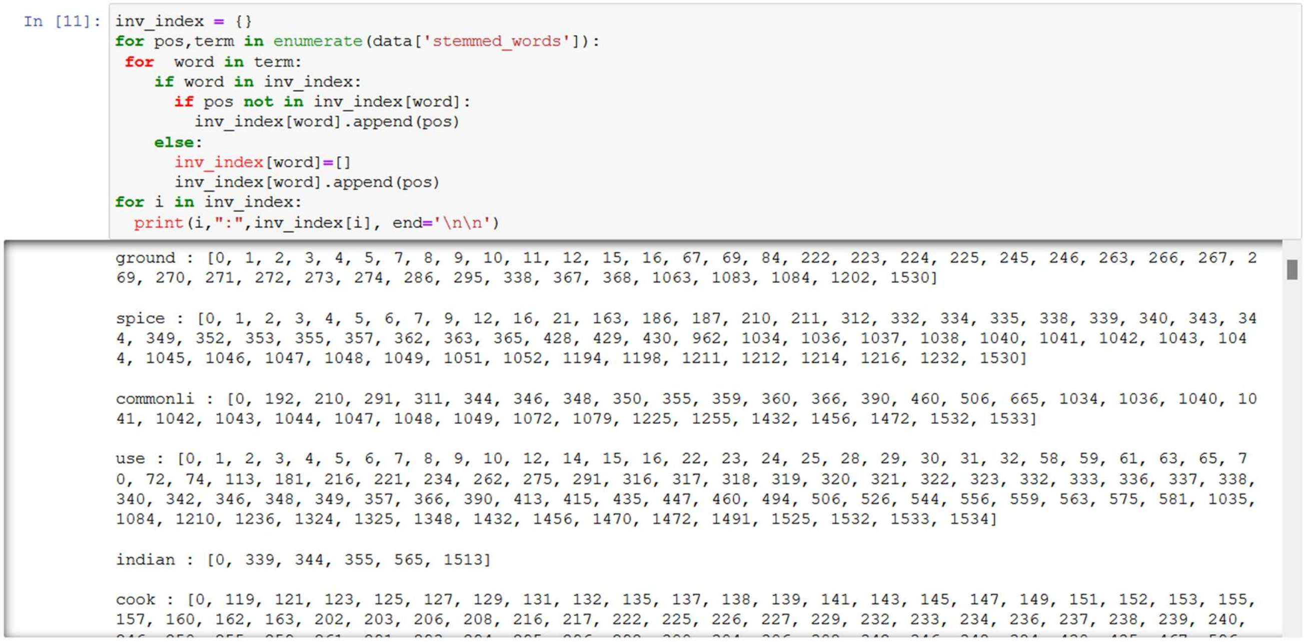
	Food Description	Classification	sentence_tokenized	word_tokenized	without_stop_words
0	Ground spice commonly used in Indian cooking a...	31302	[Ground spice commonly used in Indian cooking ...	[Ground, spice, commonly, used, in, Indian, co...	[ground, spice, commonly, used, indian, cookin...
1	Dried and ground bark or quills of Cinnamomum ...	31302	[Dried and ground bark or quills of Cinnamomum...	[Dried, and, ground, bark, or, quills, of, Cin...	[dried, ground, bark, quills, cinnamomum, veru...
2	Dried and ground, unopened flower buds of Syzy...	31302	[Dried and ground unopened flower buds of Syzy...	[Dried, and, ground, unopened, flower, buds, o...	[dried, ground, unopened, flower, buds, syzygi...
3	Dried and ground fruit or seed of Coriandrum s...	31302	[Dried and ground fruit or seed of Coriandrum ...	[Dried, and, ground, fruit, or, seed, of, Cori...	[dried, ground, fruit, seed, coriandrum, sativ...
4	Dried and ground fruit or seeds of Cuminum cym...	31302	[Dried and ground fruit or seeds of Cuminum cy...	[Dried, and, ground, fruit, or, seeds, of, Cum...	[dried, ground, fruit, seeds, cuminum, cymimum...



```
In [10]: data['stemmed_words']=data['without_stop_words'].apply(lambda f: [stemmer.stem(word) for word in f])
         data.head()
```

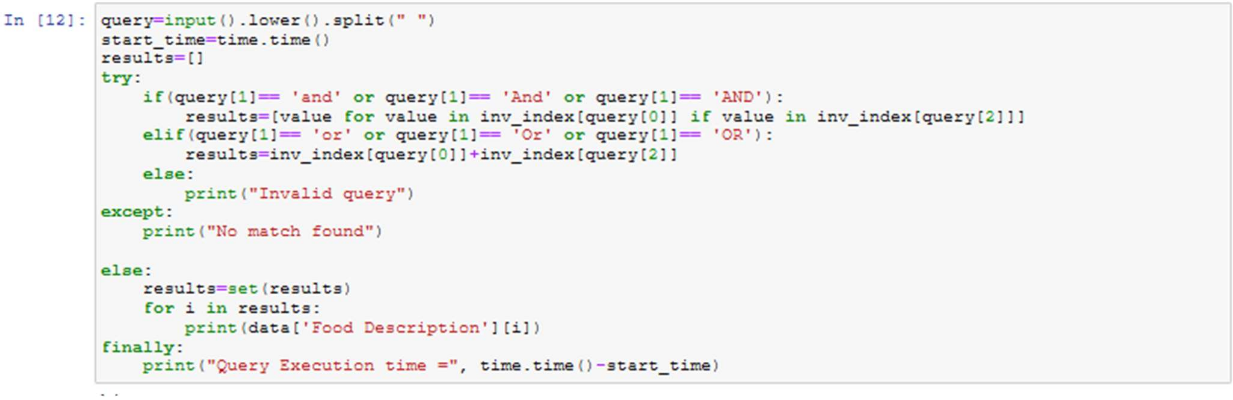
Out[10]:

	Food Description	Classification	sentence_tokenized	word_tokenized	without_stop_words	stemmed_words
0	Ground spice commonly used in Indian cooking a...	31302	[Ground spice commonly used in Indian cooking ...	[Ground, spice, commonly, used, in, Indian, co...	[ground, spice, commonly, used, indian, cookin...	[ground, spice, commonli, use, indian, cook, d...
1	Dried and ground bark or quills of Cinnamomum ...	31302	[Dried and ground bark or quills of Cinnamomum...	[Dried, and, ground, bark, or, quills, of, Cin...	[dried, ground, bark, quills, cinnamomum, veru...	[dri, ground, bark, quill, cinnamomum, verum, ...
2	Dried and ground, unopened flower buds of Syzy...	31302	[Dried and ground unopened flower buds of Syzy...	[Dried, and, ground, unopened, flower, buds, o...	[dried, ground, unopened, flower, buds, syzygi...	[dri, ground, unopen, flower, bud, syzygium, a...
3	Dried and ground fruit or seed of Coriandrum s...	31302	[Dried and ground fruit or seed of Coriandrum ...	[Dried, and, ground, fruit, or, seed, of, Cori...	[dried, ground, fruit, seed, coriandrum, sativ...	[dri, ground, fruit, seed, coriandrum, sativum...
4	Dried and ground fruit or seeds of Cuminum cym...	31302	[Dried and ground fruit or seeds of Cuminum cy...	[Dried, and, ground, fruit, or, seeds, of, Cum...	[dried, ground, fruit, seeds, cuminum, cymimum...	[dri, ground, fruit, seed, cuminum, cymimum, u...



```
In [11]: inv_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)
         for i in inv_index:
           print(i,":",inv_index[i], end='\n\n')
```

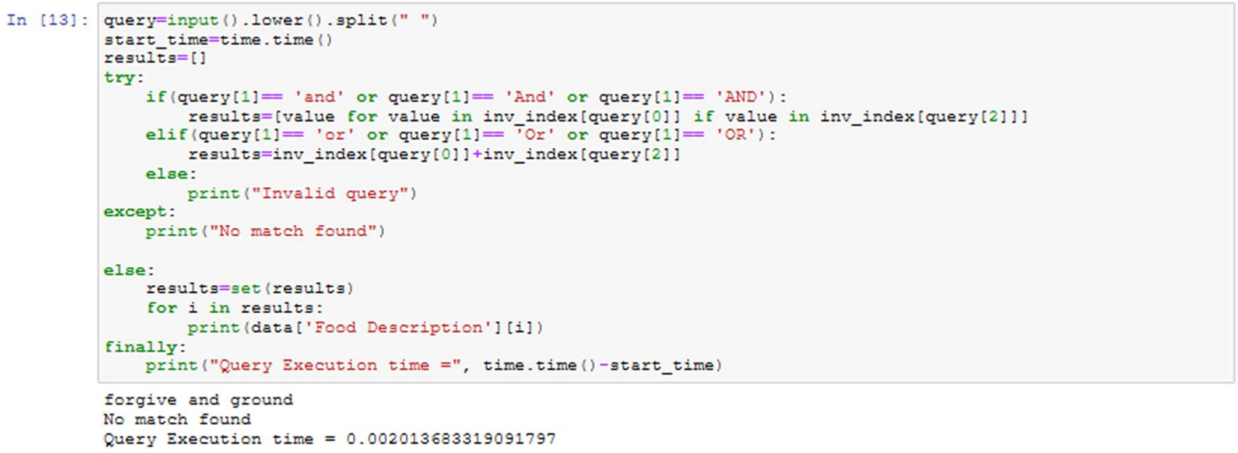
```
ground : [0, 1, 2, 3, 4, 5, 7, 8, 9, 10, 11, 12, 15, 16, 67, 69, 84, 222, 223, 224, 225, 245, 246, 263, 266, 267, 2
69, 270, 271, 272, 273, 274, 286, 295, 338, 367, 368, 1063, 1083, 1084, 1202, 1530]

spice : [0, 1, 2, 3, 4, 5, 6, 7, 9, 12, 16, 21, 163, 186, 187, 210, 211, 312, 332, 334, 335, 338, 339, 340, 343, 34
4, 349, 352, 353, 355, 357, 362, 363, 365, 428, 429, 430, 962, 1034, 1036, 1037, 1038, 1040, 1041, 1042, 1043, 104
4, 1045, 1046, 1047, 1048, 1049, 1051, 1052, 1194, 1198, 1211, 1212, 1214, 1216, 1232, 1530]

commonli : [0, 192, 210, 291, 311, 344, 346, 348, 350, 355, 359, 360, 366, 390, 460, 506, 665, 1034, 1036, 1040, 10
41, 1042, 1043, 1044, 1047, 1048, 1049, 1072, 1079, 1225, 1255, 1432, 1456, 1472, 1532, 1533]

use : [0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 12, 14, 15, 16, 22, 23, 24, 25, 28, 29, 30, 31, 32, 58, 59, 61, 63, 65, 7
0, 72, 74, 113, 181, 216, 221, 234, 262, 275, 291, 316, 317, 318, 319, 320, 321, 322, 323, 332, 333, 336, 337, 338,
340, 342, 346, 348, 349, 357, 366, 390, 413, 415, 435, 447, 460, 494, 506, 526, 544, 556, 559, 563, 575, 581, 1035,
1084, 1210, 1236, 1324, 1325, 1348, 1432, 1456, 1470, 1472, 1491, 1525, 1532, 1533, 1534]

indian : [0, 339, 344, 355, 565, 1513]

cook : [0, 119, 121, 123, 125, 127, 129, 131, 132, 135, 137, 138, 139, 141, 143, 145, 147, 149, 151, 152, 153, 155,
157, 160, 162, 163, 202, 203, 206, 208, 216, 217, 222, 225, 226, 227, 229, 232, 233, 234, 236, 237, 238, 239, 240,
```



```
In [12]: query=input().lower().split(" ")
         start_time=time.time()
         results=[]
         try:
             if(query[1]== 'and' or query[1]== 'And' or query[1]== 'AND'):
                 results=[value for value in inv_index[query[0]] if value in inv_index[query[2]]]
             elif(query[1]== 'or' or query[1]== 'Or' or query[1]== 'OR'):
                 results=inv_index[query[0]]+inv_index[query[2]]
             else:
                 print("Invalid query")
         except:
             print("No match found")

         else:
             results=set(results)
             for i in results:
                 print(data['Food Description'][i])
         finally:
             print("Query Execution time =", time.time()-start_time)
```

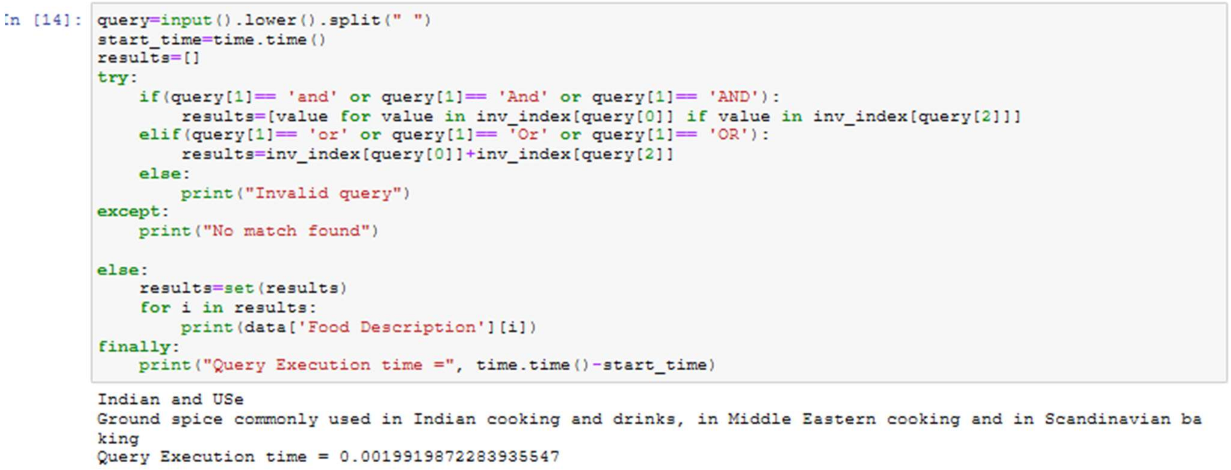
OUTPUTS:

1. No matches found – spelling mismatch of forgive



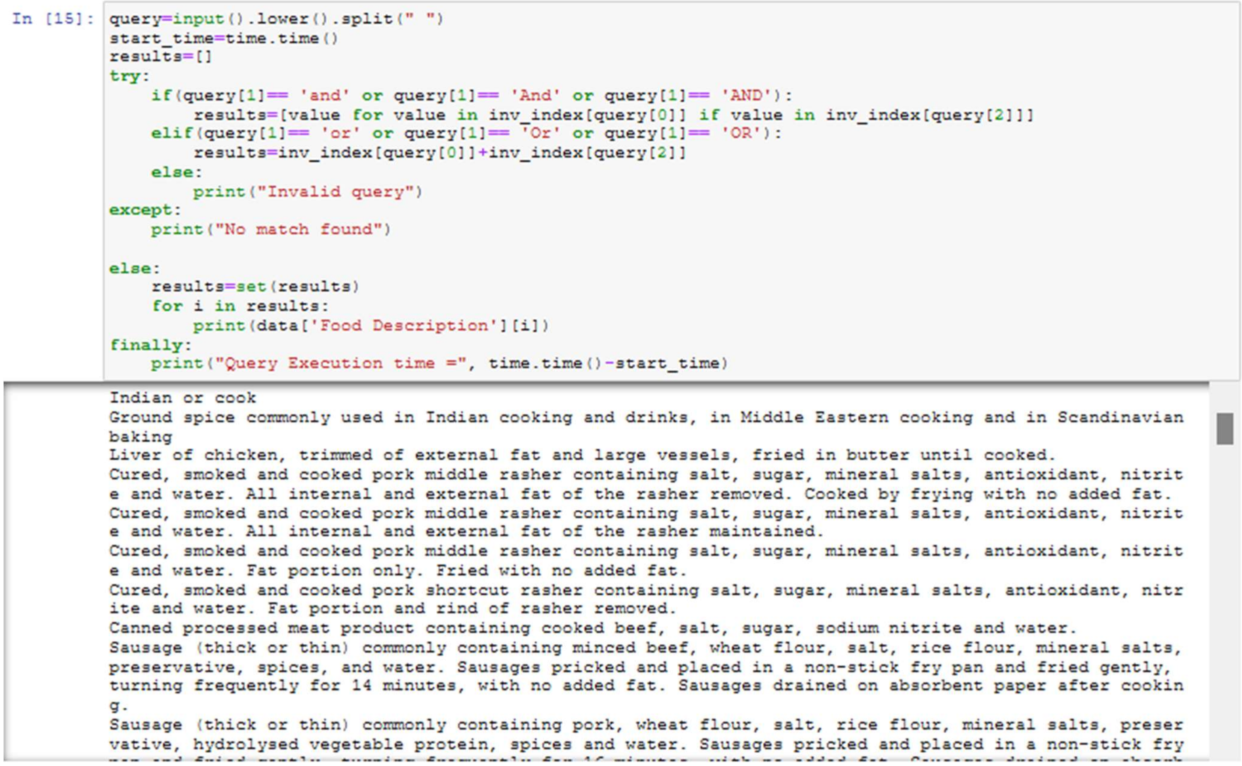
```
In [13]: query=input().lower().split(" ")
         start_time=time.time()
         results=[]
         try:
             if(query[1]== 'and' or query[1]== 'And' or query[1]== 'AND'):
                 results=[value for value in inv_index[query[0]] if value in inv_index[query[2]]]
             elif(query[1]== 'or' or query[1]== 'Or' or query[1]== 'OR'):
                 results=inv_index[query[0]]+inv_index[query[2]]
             else:
                 print("Invalid query")
         except:
             print("No match found")

         else:
             results=set(results)
             for i in results:
                 print(data['Food Description'][i])
         finally:
             print("Query Execution time =", time.time()-start_time)

         forgive and ground
         No match found
         Query Execution time = 0.002013683319091797
```

2. Mixture of upper and lower case in query



```
In [14]: query=input().lower().split(" ")
         start_time=time.time()
         results=[]
         try:
             if(query[1]== 'and' or query[1]== 'And' or query[1]== 'AND'):
                 results=[value for value in inv_index[query[0]] if value in inv_index[query[2]]]
             elif(query[1]== 'or' or query[1]== 'Or' or query[1]== 'OR'):
                 results=inv_index[query[0]]+inv_index[query[2]]
             else:
                 print("Invalid query")
         except:
             print("No match found")

         else:
             results=set(results)
             for i in results:
                 print(data['Food Description'][i])
         finally:
             print("Query Execution time =", time.time()-start_time)

         Indian and USe
         Ground spice commonly used in Indian cooking and drinks, in Middle Eastern cooking and in Scandinavian ba
         king
         Query Execution time = 0.0019919872283935547
```

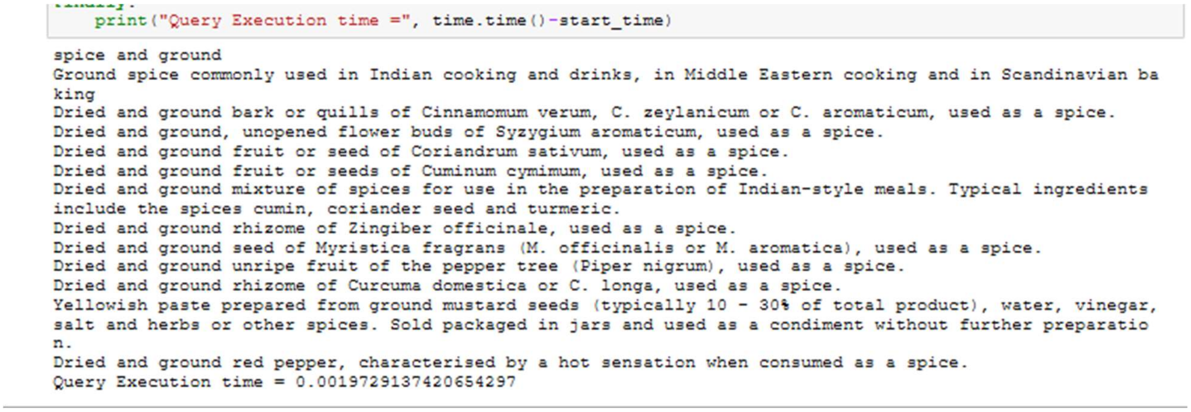
3. Or



```
In [15]: query=input().lower().split(" ")
         start_time=time.time()
         results=[]
         try:
             if(query[1]== 'and' or query[1]== 'And' or query[1]== 'AND'):
                 results=[value for value in inv_index[query[0]] if value in inv_index[query[2]]]
             elif(query[1]== 'or' or query[1]== 'Or' or query[1]== 'OR'):
                 results=inv_index[query[0]]+inv_index[query[2]]
             else:
                 print("Invalid query")
         except:
             print("No match found")

         else:
             results=set(results)
             for i in results:
                 print(data['Food Description'][i])
         finally:
             print("Query Execution time =", time.time()-start_time)
```

```
Indian or cook
Ground spice commonly used in Indian cooking and drinks, in Middle Eastern cooking and in Scandinavian
baking
Liver of chicken, trimmed of external fat and large vessels, fried in butter until cooked.
Cured, smoked and cooked pork middle rasher containing salt, sugar, mineral salts, antioxidant, nitrit
e and water. All internal and external fat of the rasher removed. Cooked by frying with no added fat.
Cured, smoked and cooked pork middle rasher containing salt, sugar, mineral salts, antioxidant, nitrit
e and water. All internal and external fat of the rasher maintained.
Cured, smoked and cooked pork middle rasher containing salt, sugar, mineral salts, antioxidant, nitrit
e and water. Fat portion only. Fried with no added fat.
Cured, smoked and cooked pork shortcut rasher containing salt, sugar, mineral salts, antioxidant, nitr
ite and water. Fat portion and rind of rasher removed.
Canned processed meat product containing cooked beef, salt, sugar, sodium nitrite and water.
Sausage (thick or thin) commonly containing minced beef, wheat flour, salt, rice flour, mineral salts,
preservative, spices, and water. Sausages pricked and placed in a non-stick fry pan and fried gently,
turning frequently for 14 minutes, with no added fat. Sausages drained on absorbent paper after cookin
g.
Sausage (thick or thin) commonly containing pork, wheat flour, salt, rice flour, mineral salts, preser
vative, hydrolysed vegetable protein, spices and water. Sausages pricked and placed in a non-stick fry
```

4. And



```
    print("Query Execution time =", time.time()-start_time)
```

```
spice and ground
Ground spice commonly used in Indian cooking and drinks, in Middle Eastern cooking and in Scandinavian ba
king
Dried and ground bark or quills of Cinnamomum verum, C. zeylanicum or C. aromaticum, used as a spice.
Dried and ground, unopened flower buds of Syzygium aromaticum, used as a spice.
Dried and ground fruit or seed of Coriandrum sativum, used as a spice.
Dried and ground fruit or seeds of Cuminum cymimum, used as a spice.
Dried and ground mixture of spices for use in the preparation of Indian-style meals. Typical ingredients
include the spices cumin, coriander seed and turmeric.
Dried and ground rhizome of Zingiber officinale, used as a spice.
Dried and ground seed of Myristica fragrans (M. officinalis or M. aromatica), used as a spice.
Dried and ground unripe fruit of the pepper tree (Piper nigrum), used as a spice.
Dried and ground rhizome of Curcuma domestica or C. longa, used as a spice.
Yellowish paste prepared from ground mustard seeds (typically 10 - 30% of total product), water, vinegar,
salt and herbs or other spices. Sold packaged in jars and used as a condiment without further preparatio
n.
Dried and ground red pepper, characterised by a hot sensation when consumed as a spice.
Query Execution time = 0.0019729137420654297
```

Link to the dataset:

https://www.kaggle.com/datasets/vinitshah0110/food-composition