Home

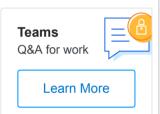
PUBLIC



Tags

Users

Jobs





Fuzzy search for keyword in a string with a dataframe

Ask Question

I have a string with keywords in it.

Example: "iphone71 is awesome"

I have a dataframe with the product varieties present.

Variation1 Variation2

IPHONE IPHONE 6S

IPHONE X IPHONE 71

IPHONE 7 IPHONE 6

I want to identify if the product present in the string given is present in the dataframe or not. If present I need to provide the name out.

My attempt:

Since the product in the string is **iphone71**, with no space in between and is not in line with what is present in the dataframe, I removed the spaces in the string.

```
df=pd.DataFrame({"Variation1":("iphone","iphone x","iphone 7"),"Variation2":
("iphone7","iphone 71","iphone 6")})
df=df.apply(lambda x: x.astype(str).str.upper())

question="iphone71 is awesome"
question=question.upper()
question=question.replace(" ","")
question
```

'IPHONE71ISAVAILABLE'

I thought of checking the **iphone71 pattern** in stripped dataframe and if match is found provide the unstripped value out from dataframe

```
def remove whitespace(x):
    try:
        x = "".join(x.split())
    except:
        pass
    return x
df.applymap(remove_whitespace)
    Variation1 Variation2
   IPHONE
                IPHONE6S
   IPHONEX
                IPHONE71
   IPHONE7
                IPHONE6
# fuzz is used to compare TWO strings
from fuzzywuzzy import fuzz
# process is used to compare a string to MULTIPLE other strings
from fuzzywuzzy import process
fuzz.partial_ratio("IPHONE7", "IPHONE71ISAVAILABLE")
100
```

I was thinking to use partial_ratio from fuzzywuzzy package to get IPHONE 71 as output, but even IPHONE 7 matches that condition.

How to accomplish this?

Expected output:

IPHONE 71

python string dataframe fuzzywuzzy

edited yesterday

asked yesterday



1 Answer

You can use fuzz.token sort ration instead.

edited yesterday

answered yesterday



This logic only works if my string has keyword of exact same format which is present in dataframe. It fails if s="iphone71 is awesome" instead of giving iphone 71 it will give other result – Sam yesterday

sorry it should work for the any strings like: "IPHONE 71 string", "IPHONE 71"
 "IPHONE 7"" – CSMaverick yesterday

Yes, it will work only if the keyword in string is SAME as the value of cells in dataframe – Sam yesterday

- yeah in that case you need to set the matching limit percentage more than
- 90% in fuzz.token_sort_ration(s,i) >= 90. it should work ! CSMaverick yesterday

Answer Your Question