

###Lab: 02

Date:11.02.2022

###Program No.:02

###Program Name: Write a Python Program to find Term Document Incidence Matrix.

#ANINDYA NAG

ROLL : UG/02/BTCSE/2018/005

#CODE:2

```

import numpy as np
import pandas as pd
d1 = "I love to play cricket".split()
d2 = "I love play cricket and football everyday".split()
d3 = "Dhoni was captain of cricket team".split()
d4 = "Dhoni won the world cup".split()
d5 = "I love a cup of tea everyday".split()
documents = [d1, d2, d3, d4, d5]
word_count = set(d1 + d2 + d3 + d4 + d5)

small_word = []
for i in word_count:
    small_word.append(i.lower())
terms = sorted(small_word)

tdi = np.zeros(shape = (len(terms),len(documents)))
print(tdi)

```

```

↳ [[0. 0. 0. 0. 0.]
    [0. 0. 0. 0. 0.]
    [0. 0. 0. 0. 0.]
    [0. 0. 0. 0. 0.]
    [0. 0. 0. 0. 0.]
    [0. 0. 0. 0. 0.]
    [0. 0. 0. 0. 0.]
    [0. 0. 0. 0. 0.]
    [0. 0. 0. 0. 0.]
    [0. 0. 0. 0. 0.]
    [0. 0. 0. 0. 0.]
    [0. 0. 0. 0. 0.]
    [0. 0. 0. 0. 0.]
    [0. 0. 0. 0. 0.]
    [0. 0. 0. 0. 0.]
    [0. 0. 0. 0. 0.]
    [0. 0. 0. 0. 0.]
    [0. 0. 0. 0. 0.]
    [0. 0. 0. 0. 0.]]

```

```

for i, t in enumerate(terms):
    for j, d in enumerate(documents):
        if t in d:
            tdi[i][j] = 1

```

```

dic = {}
for i in terms:
    for j in range(len(tdi)):
        dic[i] = tdi[j]

```

```

for i, j in dic.items():
    print(f"{i}:{j}")

```

```

a:[0. 0. 0. 1. 0.]
and:[0. 0. 0. 1. 0.]
captain:[0. 0. 0. 1. 0.]
cricket:[0. 0. 0. 1. 0.]
cup:[0. 0. 0. 1. 0.]
dhoni:[0. 0. 0. 1. 0.]
everyday:[0. 0. 0. 1. 0.]
football:[0. 0. 0. 1. 0.]
i:[0. 0. 0. 1. 0.]
love:[0. 0. 0. 1. 0.]
of:[0. 0. 0. 1. 0.]
play:[0. 0. 0. 1. 0.]
tea:[0. 0. 0. 1. 0.]
team:[0. 0. 0. 1. 0.]
the:[0. 0. 0. 1. 0.]
to:[0. 0. 0. 1. 0.]
was:[0. 0. 0. 1. 0.]
won:[0. 0. 0. 1. 0.]
world:[0. 0. 0. 1. 0.]

```

```

df = pd.DataFrame(tdi, index=dic.keys(), columns = ['d1', 'd2', 'd3', 'd4', 'd5'])
df = df.applymap(lambda x: int(x))
print(df)

```

	d1	d2	d3	d4	d5
a	0	0	0	0	1
and	0	1	0	0	0
captain	0	0	1	0	0
cricket	1	1	1	0	0
cup	0	0	0	1	1
dhoni	0	0	0	0	0

everyday	0	1	0	0	1
football	0	1	0	0	0
i	0	0	0	0	0
love	1	1	0	0	1
of	0	0	1	0	1
play	1	1	0	0	0
tea	0	0	0	0	1
team	0	0	1	0	0
the	0	0	0	1	0
to	1	0	0	0	0
was	0	0	1	0	0
won	0	0	0	1	0
world	0	0	0	1	0

✓ 0s completed at 3:22 PM

