CloudAssignment1

October 26, 2021

1 Cloud Assignment 1

1.0.1 1. The top 10 posts by score

[2]:
```
#importing libraries
from pyhive import hive
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.feature_extraction.text import CountVectorizer
import pandas as pd

#creating connection instance to set up connectiion with hive
conn = hive.Connection(host="127.0.0.1", port=10000, username="hdoop")

dataframe = pd.read_sql("SELECT OWNERDISPLAYNAME,TITLE,SCORE,BODY FROM␣
 ↪STACKEXDATA ORDER BY SCORE DESC LIMIT 10", conn)
display(dataframe)
```

```
  ownerdisplayname                                              title  score  \
0        GManNickG  Why is processing a sorted array faster than p...  25933
1  Hamza Yerlikaya  How do I undo the most recent local commits in...  23348
2   Matthew Rankin  How do I delete a Git branch locally and remot...  18514
3           pupeno  What is the difference between 'git pull' and ...  12834
4         Alex. S.              "What does the ""yield"" keyword do?"  11551
5              Oli             What is the correct JSON content type?  10921
6        paxos1977             How do I undo 'git add' before commit?  10079
7           Walker    How can I remove a specific item from an array?   9931
8          Forrest                How do I rename a local Git branch?   9792
9        GManNickG           "What is the ""-->"" operator in C/C++?"   9560

                                                 body
0  Here is a piece of C code that shows some very...
1  I accidentally committed the wrong files to Gi...
2  I want to delete a branch both locally and rem...
3  What are the differences between git pull and ...
4  What is the use of the yield keyword in Python...
5  Ive been messing around with JSON for some tim...
6  I mistakenly added files to Git using the comm...
```

```
7  I have an array of numbers and Im using the pu...
8  I dont want to rename a remote branch as descr...
9  After reading Hidden Features and Dark Corners...
```

1.0.2 2. The top 10 users by post score

[3]:
```
df2 = pd.read_sql("SELECT OWNERUSERID,OWNERDISPLAYNAME,SUM(SCORE) AS␣
 ↪TOTAL_SCORE FROM STACKEXDATA GROUP BY OWNERUSERID,OWNERDISPLAYNAME ORDER BY␣
 ↪TOTAL_SCORE DESC LIMIT 10", conn)
display(df2)
```

	owneruserid	ownerdisplayname	total_score
0	87234	GManNickG	37672
1	4883	readonly	28817
2	9951	e-satis	26878
3	6068	pupeno	25944
4	89904	Hamza Yerlikaya	24024
5	51816	Joan Venge	23763
6	49153	Ali	20203
7	179736	TIMEX	19603
8	95592	Matthew Rankin	19479
9	63051	flybywire	19362

1.0.3 3.The number of distinct users, who used the word “cloud” in one of their posts

[4]:
```
df3 = pd.read_sql('SELECT COUNT(DISTINCT(OwnerUserId)) AS NO_OF_USERS FROM␣
 ↪STACKEXDATA WHERE TITLE LIKE "% cloud %" OR BODY LIKE "% cloud %"', conn)
display(df3)
```

	no_of_users
0	163

Making a data frame ‘df4’ of the titles and body of the top 10 users.

[15]:
```
df4 = pd.read_sql('''
SELECT OWNERUSERID, OWNERDISPLAYNAME,TITLE,BODY FROM STACKEXDATA WHERE␣
 ↪OWNERUSERID IN (87234,4883,9951,6068,89904,51816,49153,179736,95592,63051)
''', conn)
display(df4)
```

	owneruserid	ownerdisplayname	\
0	89904	Hamza Yerlikaya	
1	95592	Matthew Rankin	

```
2          4883         readonly
3         49153              Ali
4          6068            pupeno
..          ...               ...
420       51816       Joan Venge
421        9951          e-satis
422       51816       Joan Venge
423      179736            TIMEX
424       63051        flybywire

                                                 title  \
0    How do I undo the most recent local commits in...
1    How do I delete a Git branch locally and remot...
2    How can you find out which process is listenin...
3    How to replace all occurrences of a string in ...
4            How do you get a timestamp in JavaScript?
..                                                 ...
420              Clamping floating numbers in Python?
421  Is there any Python-like interactive console f...
422                        Checking delegates for null
423  How can I remove all nil elements in a Swift a...
424  What are the advantages and disadvantages of j...

                                                  body
0    I accidentally committed the wrong files to Gi...
1    I want to delete a branch both locally and rem...
2    How can you find out which process is listenin...
3    I have this string in my JavaScript code quotT...
4    Something similar to Unixs timestamp that is a...
..                                                 ...
420  Is there a builtin function for this in Python...
421  I spent a lot of time programming in Java rece...
422  I was reading the Essential C    book and am w...
423  Basic way doesnt work for index in   lt listco...
424  What are the advantages and disadvantages of j...

[425 rows x 4 columns]
```

[16]:
```
df4['title_body'] = df4[['title', 'body']].agg(' '.join, axis=1)
del df4['title']
del df4['body']
display(df4)
```

```
    owneruserid ownerdisplayname  \
0         89904  Hamza Yerlikaya
1         95592   Matthew Rankin
2          4883         readonly
```

```
3         49153              Ali
4          6068           pupeno
..          ...              ...
420       51816       Joan Venge
421        9951          e-satis
422       51816       Joan Venge
423      179736             TIMEX
424       63051        flybywire

                                            title_body
0    How do I undo the most recent local commits in...
1    How do I delete a Git branch locally and remot...
2    How can you find out which process is listenin...
3    How to replace all occurrences of a string in ...
4    How do you get a timestamp in JavaScript? Some...
..                                                 ...
420  Clamping floating numbers in Python? Is there ...
421  Is there any Python-like interactive console f...
422  Checking delegates for null I was reading the ...
423  How can I remove all nil elements in a Swift a...
424  What are the advantages and disadvantages of j...

[425 rows x 3 columns]
```

[18]:
```
vectorizer = TfidfVectorizer(stop_words='english', lowercase=True)
#extracting the top 10 users in a list
userlist = list(df4["ownerdisplayname"].unique())

#iterating through the list of users
for user in userlist:

    #storing the posts per top 10 users in a new dataframe
    user_posts_df=df4[df4["ownerdisplayname"] == user]

    response = vectorizer.fit_transform(user_posts_df["title_body"])
    tf_idf_df= pd.DataFrame(response.toarray(),columns=vectorizer.
 ↪get_feature_names())

    #taking the sum of tf-idf's of each term
    sums = tf_idf_df.sum(axis=0)

    #finding the top 10 terms from the list of terms based on sums
    topTen= sums.nlargest(10)

    #making a list of the top 10 terms
    topTenWordsList = list(topTen.index)
```

```
    #storing the tf-idfs in a new dataframe
    tf_idf_topTen=tf_idf_df[topTenWordsList]

    #displaying the tf-idf's of top 10 terms of top 10 users
    display("TF-IDF of the top 10 terms for the user : "+user)
    display(tf_idf_topTen)
```

'TF-IDF of the top 10 terms for the user : Hamza Yerlikaya'

	file	timer	new	java	setpreferredwidth	\
0	0.000000	0.0000	0.000000	0.000000	0.000000	
1	0.000000	0.0000	0.000000	0.040077	0.606105	
2	0.351468	0.0000	0.000000	0.000000	0.000000	
3	0.000000	0.0000	0.000000	0.000000	0.000000	
4	0.463315	0.0000	0.000000	0.000000	0.000000	
5	0.000000	0.0000	0.000000	0.313903	0.000000	
6	0.000000	0.6912	0.235033	0.091408	0.000000	
7	0.077501	0.0000	0.442922	0.000000	0.000000	
8	0.000000	0.0000	0.000000	0.000000	0.000000	
9	0.000000	0.0000	0.000000	0.163094	0.000000	

	tablegetcolumnmodelgetcolumn	files	application	jar	array
0	0.000000	0.085641	0.000000	0.000000	0.000000
1	0.606105	0.000000	0.000000	0.000000	0.000000
2	0.000000	0.312480	0.000000	0.000000	0.000000
3	0.000000	0.000000	0.000000	0.000000	0.000000
4	0.000000	0.102980	0.264787	0.264787	0.000000
5	0.000000	0.000000	0.000000	0.000000	0.000000
6	0.000000	0.000000	0.000000	0.000000	0.000000
7	0.000000	0.000000	0.000000	0.000000	0.000000
8	0.000000	0.097851	0.251599	0.251599	0.000000
9	0.000000	0.000000	0.000000	0.000000	0.493305

'TF-IDF of the top 10 terms for the user : Matthew Rankin'

	install	python	pip	branch	installed	navigation	flask	\
0	0.000000	0.000000	0.000000	0.697465	0.000000	0.000000	0.000000	
1	0.000000	0.000000	0.000000	0.000000	0.000000	0.551288	0.000000	
2	0.464973	0.290143	0.286137	0.000000	0.000000	0.000000	0.000000	
3	0.339622	0.000000	0.135849	0.000000	0.157430	0.000000	0.000000	
4	0.000000	0.000000	0.000000	0.000000	0.436263	0.000000	0.520551	
5	0.427251	0.000000	0.284834	0.000000	0.000000	0.000000	0.000000	
6	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
7	0.000000	0.432970	0.000000	0.000000	0.000000	0.000000	0.000000	

	version	started	psycopg

```
0  0.000000  0.000000  0.000000
1  0.000000  0.000000  0.000000
2  0.000000  0.000000  0.000000
3  0.000000  0.000000  0.516576
4  0.520551  0.000000  0.000000
5  0.000000  0.000000  0.000000
6  0.000000  0.000000  0.000000
7  0.000000  0.516622  0.000000
```

'TF-IDF of the top 10 terms for the user : readonly'

```
      python       use      ruby      list     table  difference    branch  \
0   0.000000  0.000000  0.000000  0.000000  0.000000    0.000000  0.000000
1   0.000000  0.000000  0.000000  0.000000  0.000000    0.000000  0.000000
2   0.000000  0.000000  0.000000  0.000000  0.000000    0.000000  0.000000
3   0.000000  0.000000  0.413988  0.000000  0.000000    0.000000  0.000000
4   0.000000  0.000000  0.000000  0.000000  0.000000    0.000000  0.636117
5   0.146760  0.168717  0.000000  0.000000  0.000000    0.000000  0.000000
6   0.000000  0.191672  0.000000  0.000000  0.000000    0.491935  0.000000
7   0.000000  0.000000  0.000000  0.000000  0.000000    0.000000  0.000000
8   0.000000  0.000000  0.209780  0.000000  0.000000    0.000000  0.000000
9   0.251336  0.000000  0.000000  0.000000  0.000000    0.000000  0.000000
10  0.195328  0.112276  0.000000  0.000000  0.000000    0.288162  0.000000
11  0.000000  0.000000  0.000000  0.142349  0.000000    0.000000  0.000000
12  0.000000  0.000000  0.000000  0.000000  0.000000    0.000000  0.000000
13  0.000000  0.000000  0.404077  0.000000  0.000000    0.000000  0.000000
14  0.101808  0.234079  0.000000  0.418562  0.000000    0.000000  0.000000
15  0.000000  0.119743  0.000000  0.000000  0.307326    0.000000  0.000000
16  0.000000  0.000000  0.000000  0.000000  0.000000    0.000000  0.000000
17  0.000000  0.231741  0.000000  0.000000  0.000000    0.000000  0.000000
18  0.000000  0.000000  0.000000  0.000000  0.000000    0.000000  0.000000
19  0.087249  0.000000  0.000000  0.000000  0.000000    0.000000  0.000000
20  0.149597  0.000000  0.000000  0.000000  0.000000    0.220697  0.000000
21  0.231989  0.000000  0.000000  0.000000  0.000000    0.000000  0.000000
22  0.000000  0.000000  0.000000  0.000000  0.000000    0.000000  0.000000
23  0.000000  0.000000  0.000000  0.000000  0.000000    0.000000  0.000000
24  0.000000  0.000000  0.000000  0.000000  0.000000    0.000000  0.350427
25  0.000000  0.000000  0.000000  0.000000  0.000000    0.000000  0.000000
26  0.000000  0.000000  0.000000  0.000000  0.000000    0.000000  0.000000
27  0.073720  0.000000  0.000000  0.000000  0.000000    0.000000  0.000000
28  0.000000  0.198917  0.000000  0.000000  0.255265    0.000000  0.000000
29  0.000000  0.000000  0.000000  0.000000  0.521156    0.000000  0.000000
30  0.000000  0.000000  0.000000  0.202044  0.000000    0.000000  0.000000
31  0.159297  0.000000  0.218306  0.436612  0.000000    0.000000  0.000000
32  0.121700  0.000000  0.000000  0.000000  0.000000    0.000000  0.000000

         way    method   process
```

```
0   0.000000  0.000000  0.350231
1   0.000000  0.000000  0.000000
2   0.000000  0.000000  0.000000
3   0.000000  0.000000  0.000000
4   0.000000  0.000000  0.000000
5   0.000000  0.000000  0.000000
6   0.000000  0.000000  0.000000
7   0.000000  0.000000  0.000000
8   0.000000  0.209780  0.000000
9   0.144470  0.000000  0.000000
10  0.000000  0.000000  0.000000
11  0.000000  0.000000  0.000000
12  0.000000  0.000000  0.000000
13  0.000000  0.000000  0.000000
14  0.000000  0.000000  0.000000
15  0.119743  0.000000  0.000000
16  0.000000  0.000000  0.187004
17  0.000000  0.000000  0.000000
18  0.000000  0.000000  0.000000
19  0.000000  0.000000  0.000000
20  0.000000  0.000000  0.000000
21  0.000000  0.423901  0.000000
22  0.181739  0.000000  0.000000
23  0.000000  0.000000  0.000000
24  0.062539  0.000000  0.000000
25  0.000000  0.000000  0.000000
26  0.000000  0.000000  0.000000
27  0.000000  0.101028  0.000000
28  0.099459  0.000000  0.382898
29  0.000000  0.000000  0.000000
30  0.169489  0.000000  0.000000
31  0.183130  0.218306  0.000000
32  0.000000  0.000000  0.000000
```

'TF-IDF of the top 10 terms for the user : Ali'

```
    javascript     using     array       php    jquery  file      want  \
0     0.116053  0.000000  0.000000  0.000000  0.000000   0.0  0.000000
1     0.000000  0.000000  0.000000  0.000000  0.000000   0.0  0.000000
2     0.000000  0.237201  0.000000  0.000000  0.343019   0.0  0.000000
3     0.144416  0.000000  0.393527  0.000000  0.000000   0.0  0.000000
4     0.000000  0.000000  0.000000  0.102097  0.000000   0.0  0.000000
..         ...       ...       ...       ...       ...   ...       ...
74    0.000000  0.000000  0.613903  0.112644  0.000000   0.0  0.000000
75    0.000000  0.034228  0.000000  0.000000  0.000000   0.0  0.000000
76    0.000000  0.000000  0.000000  0.000000  0.000000   0.0  0.000000
77    0.000000  0.000000  0.000000  0.000000  0.000000   0.0  0.000000
```

```
78    0.000000  0.077236  0.000000  0.000000  0.000000   0.0  0.096646

       class    string       way
0   0.000000  0.194331  0.000000
1   0.000000  0.000000  0.000000
2   0.000000  0.000000  0.148405
3   0.000000  0.000000  0.000000
4   0.000000  0.113974  0.000000
..       ...       ...       ...
74  0.000000  0.000000  0.000000
75  0.305420  0.137546  0.042829
76  0.052248  0.000000  0.043961
77  0.000000  0.000000  0.000000
78  0.000000  0.206918  0.000000

[79 rows x 10 columns]


'TF-IDF of the top 10 terms for the user : pupeno'


        file      java   android       sql       git       way      like  \
0   0.000000  0.000000  0.000000  0.000000  0.000000  0.000000  0.000000
1   0.000000  0.000000  0.000000  0.000000  0.728914  0.000000  0.000000
2   0.000000  0.000000  0.130150  0.000000  0.000000  0.000000  0.000000
3   0.000000  0.000000  0.172570  0.000000  0.000000  0.000000  0.000000
4   0.000000  0.000000  0.000000  0.000000  0.000000  0.140540  0.000000
5   0.000000  0.386646  0.000000  0.000000  0.000000  0.000000  0.000000
6   0.000000  0.000000  0.271615  0.000000  0.000000  0.000000  0.000000
7   0.000000  0.000000  0.000000  0.217416  0.000000  0.000000  0.000000
8   0.000000  0.000000  0.000000  0.000000  0.000000  0.000000  0.046940
9   0.000000  0.000000  0.147495  0.000000  0.000000  0.000000  0.000000
10  0.000000  0.000000  0.323119  0.000000  0.000000  0.119955  0.000000
11  0.000000  0.000000  0.000000  0.000000  0.000000  0.076296  0.158251
12  0.000000  0.000000  0.000000  0.105312  0.000000  0.000000  0.072082
13  0.652980  0.000000  0.000000  0.000000  0.414954  0.085022  0.000000
14  0.102698  0.000000  0.000000  0.000000  0.000000  0.000000  0.000000
15  0.000000  0.000000  0.000000  0.000000  0.000000  0.000000  0.000000
16  0.371197  0.000000  0.000000  0.000000  0.000000  0.000000  0.000000
17  0.000000  0.000000  0.352152  0.000000  0.000000  0.000000  0.000000
18  0.000000  0.000000  0.000000  0.000000  0.000000  0.093310  0.290312
19  0.000000  0.364640  0.000000  0.000000  0.000000  0.000000  0.000000
20  0.000000  0.000000  0.000000  0.000000  0.000000  0.000000  0.000000
21  0.307285  0.000000  0.000000  0.000000  0.000000  0.120032  0.124483
22  0.083260  0.370491  0.000000  0.000000  0.000000  0.000000  0.067458
23  0.000000  0.000000  0.000000  0.000000  0.000000  0.088339  0.000000
24  0.000000  0.000000  0.000000  0.000000  0.000000  0.029366  0.060910
25  0.000000  0.000000  0.000000  0.000000  0.142314  0.087479  0.090723
26  0.000000  0.154458  0.000000  0.000000  0.000000  0.108471  0.000000
```

27	0.000000	0.000000	0.000000	0.000000	0.000000	0.083163	0.000000
28	0.000000	0.389127	0.000000	0.000000	0.000000	0.000000	0.070851
29	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.106695
30	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
31	0.000000	0.000000	0.000000	0.000000	0.000000	0.080477	0.000000
32	0.116987	0.000000	0.000000	0.692403	0.000000	0.000000	0.094784
33	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
34	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
35	0.173716	0.000000	0.000000	0.274177	0.000000	0.135714	0.000000
36	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.049869

	time	tables	code
0	0.198534	0.000000	0.000000
1	0.000000	0.000000	0.000000
2	0.000000	0.000000	0.000000
3	0.388284	0.000000	0.000000
4	0.000000	0.000000	0.000000
5	0.000000	0.000000	0.000000
6	0.000000	0.000000	0.000000
7	0.000000	0.233437	0.000000
8	0.000000	0.000000	0.000000
9	0.000000	0.000000	0.000000
10	0.000000	0.000000	0.341624
11	0.000000	0.000000	0.325928
12	0.000000	0.452290	0.000000
13	0.000000	0.000000	0.000000
14	0.000000	0.000000	0.038082
15	0.000000	0.000000	0.000000
16	0.000000	0.000000	0.000000
17	0.000000	0.000000	0.000000
18	0.000000	0.000000	0.000000
19	0.000000	0.000000	0.000000
20	0.000000	0.277321	0.000000
21	0.000000	0.000000	0.000000
22	0.000000	0.000000	0.000000
23	0.133851	0.000000	0.000000
24	0.000000	0.000000	0.000000
25	0.000000	0.000000	0.000000
26	0.328708	0.000000	0.000000
27	0.000000	0.000000	0.000000
28	0.000000	0.000000	0.000000
29	0.000000	0.000000	0.000000
30	0.000000	0.000000	0.000000
31	0.000000	0.000000	0.000000
32	0.000000	0.000000	0.000000
33	0.000000	0.000000	0.178511
34	0.000000	0.000000	0.000000
35	0.000000	0.000000	0.064417

```
36  0.000000  0.000000  0.000000
```

```
'TF-IDF of the top 10 terms for the user : Joan Venge'
```

```
      python      like      list    string      want     class     index  \
0   0.144668  0.073637  0.000000  0.447124  0.000000  0.000000  0.000000
1   0.067989  0.069213  0.315199  0.000000  0.184228  0.000000  0.248781
2   0.000000  0.057536  0.000000  0.000000  0.076573  0.000000  0.310212
3   0.088854  0.090455  0.000000  0.549244  0.000000  0.000000  0.000000
4   0.210693  0.000000  0.000000  0.000000  0.000000  0.000000  0.000000
..       ...       ...       ...       ...       ...       ...       ...
61  0.120300  0.000000  0.000000  0.000000  0.000000  0.000000  0.000000
62  0.000000  0.047863  0.000000  0.000000  0.063700  0.000000  0.000000
63  0.000000  0.000000  0.000000  0.000000  0.096343  0.701068  0.000000
64  0.273733  0.000000  0.000000  0.000000  0.000000  0.000000  0.000000
65  0.000000  0.000000  0.000000  0.000000  0.000000  0.067684  0.000000

       value       use  function
0   0.000000  0.000000  0.000000
1   0.000000  0.000000  0.000000
2   0.000000  0.000000  0.000000
3   0.000000  0.000000  0.000000
4   0.000000  0.000000  0.000000
..       ...       ...       ...
61  0.197671  0.000000  0.000000
62  0.000000  0.072656  0.000000
63  0.000000  0.000000  0.000000
64  0.000000  0.000000  0.240971
65  0.203053  0.000000  0.000000

[66 rows x 10 columns]
```

```
'TF-IDF of the top 10 terms for the user : e-satis'
```

```
      python      like       git      head      does     using       way  \
0   0.000000  0.066630  0.078782  0.000000  0.000000  0.000000  0.000000
1   0.000000  0.101237  0.000000  0.000000  0.119702  0.000000  0.000000
2   0.525068  0.000000  0.000000  0.000000  0.000000  0.000000  0.000000
3   0.441995  0.000000  0.000000  0.000000  0.000000  0.000000  0.277646
4   0.000000  0.000000  0.000000  0.000000  0.000000  0.000000  0.083623
5   0.000000  0.000000  0.000000  0.000000  0.421029  0.000000  0.158685
6   0.000000  0.000000  0.000000  0.000000  0.000000  0.119729  0.000000
7   0.000000  0.000000  0.000000  0.000000  0.227740  0.000000  0.000000
8   0.000000  0.000000  0.376139  0.411180  0.075228  0.000000  0.000000
9   0.155490  0.328761  0.000000  0.000000  0.000000  0.000000  0.146510
10  0.000000  0.000000  0.000000  0.000000  0.000000  0.000000  0.000000
```

```
11  0.000000  0.156069  0.000000  0.000000  0.000000  0.000000  0.173878
12  0.000000  0.000000  0.075446  0.000000  0.000000  0.150893  0.000000
13  0.000000  0.188069  0.000000  0.000000  0.000000  0.000000  0.000000
14  0.000000  0.000000  0.000000  0.000000  0.000000  0.236845  0.000000
15  0.000000  0.000000  0.179128  0.734307  0.000000  0.000000  0.000000
16  0.081895  0.138523  0.000000  0.000000  0.163789  0.000000  0.000000
17  0.000000  0.033726  0.000000  0.000000  0.000000  0.159508  0.112722
18  0.000000  0.000000  0.338695  0.000000  0.000000  0.000000  0.000000
19  0.109076  0.092250  0.000000  0.000000  0.000000  0.000000  0.000000
20  0.000000  0.000000  0.187925  0.000000  0.062642  0.000000  0.059024
21  0.000000  0.000000  0.000000  0.000000  0.000000  0.000000  0.000000
22  0.000000  0.000000  0.000000  0.000000  0.000000  0.300439  0.000000
23  0.000000  0.000000  0.000000  0.000000  0.000000  0.000000  0.000000
24  0.079642  0.202069  0.000000  0.000000  0.000000  0.079642  0.000000

         use      know        id
0   0.000000  0.070219  0.000000
1   0.000000  0.000000  0.000000
2   0.262534  0.000000  0.000000
3   0.000000  0.000000  0.000000
4   0.000000  0.000000  0.000000
5   0.000000  0.225160  0.000000
6   0.119729  0.000000  0.000000
7   0.113870  0.000000  0.000000
8   0.150456  0.000000  0.000000
9   0.000000  0.069295  0.178117
10  0.000000  0.000000  0.000000
11  0.000000  0.082239  0.000000
12  0.000000  0.134492  0.000000
13  0.222372  0.000000  0.000000
14  0.000000  0.000000  0.000000
15  0.000000  0.000000  0.000000
16  0.000000  0.072993  0.000000
17  0.119631  0.142171  0.045680
18  0.000000  0.000000  0.000000
19  0.000000  0.000000  0.124948
20  0.000000  0.000000  0.000000
21  0.000000  0.000000  0.000000
22  0.000000  0.000000  0.000000
23  0.000000  0.000000  0.488716
24  0.000000  0.141971  0.000000

'TF-IDF of the top 10 terms for the user : TIMEX'

       python     want    string      user  dictionary      file    return  \
0    0.296315  0.00000  0.000000  0.000000    0.000000  0.000000  0.000000
1    0.168368  0.00000  0.000000  0.000000    0.000000  0.000000  0.000000
```

```
2    0.000000  0.00000  0.000000  0.000000    0.000000  0.214901  0.000000
3    0.000000  0.00000  0.188953  0.000000    0.000000  0.000000  0.000000
4    0.000000  0.00000  0.000000  0.000000    0.000000  0.000000  0.000000
..        ...      ...       ...       ...         ...       ...       ...
110  0.000000  0.00000  0.000000  0.000000    0.000000  0.000000  0.283947
111  0.061815  0.00000  0.000000  0.000000    0.086926  0.000000  0.000000
112  0.000000  0.04225  0.000000  0.000000    0.000000  0.000000  0.000000
113  0.000000  0.00000  0.000000  0.084033    0.000000  0.000000  0.092008
114  0.000000  0.00000  0.000000  0.000000    0.000000  0.000000  0.000000

          use  just  list
0    0.000000   0.0   0.0
1    0.000000   0.0   0.0
2    0.086508   0.0   0.0
3    0.000000   0.0   0.0
4    0.000000   0.0   0.0
..        ...   ...   ...
110  0.000000   0.0   0.0
111  0.000000   0.0   0.0
112  0.000000   0.0   0.0
113  0.000000   0.0   0.0
114  0.000000   0.0   0.0

[115 rows x 10 columns]


'TF-IDF of the top 10 terms for the user : flybywire'

        file    python      want        vs  standard      bash   command  \
0   0.000000  0.000000  0.000000  0.000000  0.000000  0.000000  0.000000
1   0.000000  0.000000  0.112601  0.000000  0.000000  0.000000  0.000000
2   0.000000  0.000000  0.000000  0.000000  0.000000  0.000000  0.000000
3   0.000000  0.113048  0.000000  0.000000  0.000000  0.000000  0.000000
4   0.243946  0.000000  0.000000  0.000000  0.547364  0.181656  0.000000
5   0.166264  0.000000  0.155217  0.000000  0.466327  0.412698  0.000000
6   0.000000  0.000000  0.000000  0.000000  0.000000  0.000000  0.000000
7   0.112736  0.127016  0.000000  0.000000  0.000000  0.000000  0.000000
8   0.000000  0.537229  0.000000  0.000000  0.000000  0.000000  0.000000
9   0.000000  0.000000  0.122889  0.000000  0.000000  0.000000  0.000000
10  0.000000  0.000000  0.000000  0.000000  0.000000  0.000000  0.000000
11  0.125574  0.000000  0.117230  0.000000  0.000000  0.000000  0.164955
12  0.000000  0.151165  0.000000  0.158291  0.000000  0.000000  0.000000
13  0.077215  0.000000  0.288337  0.000000  0.000000  0.000000  0.000000
14  0.000000  0.150920  0.000000  0.000000  0.000000  0.000000  0.000000
15  0.000000  0.000000  0.000000  0.000000  0.000000  0.000000  0.000000
16  0.000000  0.000000  0.000000  0.386070  0.000000  0.000000  0.000000
17  0.000000  0.000000  0.000000  0.000000  0.000000  0.000000  0.000000
18  0.000000  0.000000  0.000000  0.000000  0.000000  0.000000  0.000000
```

19	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
20	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
21	0.000000	0.000000	0.085996	0.000000	0.000000	0.171488	0.302514
22	0.000000	0.000000	0.000000	0.227035	0.000000	0.000000	0.000000
23	0.071235	0.000000	0.066501	0.000000	0.000000	0.176817	0.374298
24	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
25	0.000000	0.000000	0.173412	0.000000	0.000000	0.000000	0.000000
26	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
27	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
28	0.323866	0.000000	0.100782	0.000000	0.000000	0.267965	0.000000
29	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
30	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
31	0.000000	0.328540	0.000000	0.000000	0.000000	0.000000	0.000000
32	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
33	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
34	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
35	0.000000	0.000000	0.068019	0.000000	0.000000	0.000000	0.000000
36	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
37	0.192038	0.000000	0.089639	0.000000	0.000000	0.000000	0.000000
38	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
39	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
40	0.278703	0.000000	0.000000	0.000000	0.000000	0.000000	0.366107
41	0.000000	0.000000	0.000000	0.194207	0.000000	0.000000	0.000000
42	0.000000	0.000000	0.000000	0.000000	0.207214	0.000000	0.000000
43	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
44	0.319440	0.000000	0.149107	0.000000	0.000000	0.000000	0.000000
45	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
46	0.000000	0.321788	0.000000	0.000000	0.000000	0.000000	0.000000
47	0.000000	0.000000	0.000000	0.232327	0.000000	0.000000	0.000000
48	0.000000	0.000000	0.000000	0.253717	0.000000	0.000000	0.000000

	java	instance	output
0	0.000000	0.000000	0.000000
1	0.000000	0.000000	0.000000
2	0.000000	0.000000	0.000000
3	0.000000	0.000000	0.000000
4	0.000000	0.000000	0.320451
5	0.000000	0.000000	0.218406
6	0.588666	0.000000	0.000000
7	0.000000	0.000000	0.000000
8	0.000000	0.000000	0.000000
9	0.000000	0.000000	0.000000
10	0.000000	0.000000	0.000000
11	0.000000	0.000000	0.000000
12	0.000000	0.000000	0.000000
13	0.000000	0.000000	0.000000
14	0.000000	0.000000	0.000000
15	0.000000	0.000000	0.000000

```
16  0.458910  0.000000  0.000000
17  0.000000  0.000000  0.000000
18  0.000000  0.000000  0.000000
19  0.000000  0.864911  0.000000
20  0.000000  0.000000  0.000000
21  0.000000  0.000000  0.363017
22  0.000000  0.000000  0.000000
23  0.000000  0.000000  0.280723
24  0.000000  0.000000  0.000000
25  0.000000  0.000000  0.000000
26  0.000000  0.000000  0.000000
27  0.000000  0.000000  0.000000
28  0.000000  0.000000  0.000000
29  0.000000  0.000000  0.000000
30  0.000000  0.000000  0.000000
31  0.000000  0.000000  0.000000
32  0.000000  0.000000  0.000000
33  0.000000  0.000000  0.000000
34  0.000000  0.000000  0.000000
35  0.000000  0.000000  0.000000
36  0.000000  0.000000  0.000000
37  0.000000  0.000000  0.000000
38  0.000000  0.000000  0.000000
39  0.160137  0.000000  0.000000
40  0.000000  0.000000  0.000000
41  0.000000  0.000000  0.000000
42  0.000000  0.000000  0.000000
43  0.000000  0.000000  0.000000
44  0.000000  0.000000  0.000000
45  0.000000  0.317777  0.000000
46  0.000000  0.000000  0.000000
47  0.000000  0.000000  0.000000
48  0.000000  0.000000  0.000000
```

'TF-IDF of the top 10 terms for the user : GManNickG'

```
      idiom      copy     array       int  arraysize      data  operator  \
0  0.000000  0.000000  0.260409  0.173606   0.304361  0.304361  0.000000
1  0.000000  0.000000  0.000000  0.149462   0.000000  0.000000  0.149462
2  0.480997  0.360748  0.091453  0.000000   0.000000  0.000000  0.091453

       does  mentioned    places
0  0.000000   0.000000  0.000000
1  0.000000   0.000000  0.000000
2  0.240499   0.240499  0.240499
```