

## tfidf-4th task

October 31, 2021

```
[2]: import pandas as pd
```

```
[4]: pip install pyhive
```

```
Collecting pyhive
  Downloading https://files.pythonhosted.org/packages/7a/3b/379563ead1d431
b946d5d20b8c3b960c318581926702040931aaa2d5cf28/PyHive-0.6.4.tar.gz (44kB)
    |                               | 51kB 5.0MB/s eta 0:00:011
Requirement already satisfied: future in
/opt/conda/anaconda/lib/python3.7/site-packages (from pyhive) (0.17.1)
Requirement already satisfied: python-dateutil in
/opt/conda/anaconda/lib/python3.7/site-packages (from pyhive) (2.8.0)
Requirement already satisfied: six>=1.5 in
/opt/conda/anaconda/lib/python3.7/site-packages (from python-dateutil->pyhive)
(1.15.0)
Building wheels for collected packages: pyhive
  Building wheel for pyhive (setup.py) ... done
  Created wheel for pyhive: filename=PyHive-0.6.4-cp37-none-any.whl
size=51495
sha256=3989ce7cecc9ce578f8a630f77f6ff384c81ceee7b0982810b415ac874af71bd
  Stored in directory: /root/.cache/pip/wheels/0b/8d/51/4f4d0799c9281b8f6d742795
733395638b5c2f490cf49dc09a
Successfully built pyhive
Installing collected packages: pyhive
Successfully installed pyhive-0.6.4
Note: you may need to restart the kernel to use updated packages.
```

```
[5]: pip install thrift
```

```
Collecting thrift
  Downloading https://files.pythonhosted.org/packages/6e/97/a73a1a62f62375
b21464fa45a0093ef0b653cb14f7599cffce35d51c9161/thrift-0.15.0.tar.gz (59kB)
    |                               | 61kB 5.1MB/s eta 0:00:011
Requirement already satisfied: six>=1.7.2 in
/opt/conda/anaconda/lib/python3.7/site-packages (from thrift) (1.15.0)
Building wheels for collected packages: thrift
  Building wheel for thrift (setup.py) ... done
  Created wheel for thrift:
```

```
filename=thrift-0.15.0-cp37-cp37m-linux_x86_64.whl size=411305
sha256=e73303df16bdb07c67fffcacc560b00c14c01ff6c71f54cb81fe23f2101eb587
  Stored in directory: /root/.cache/pip/wheels/ed/98/a6/f324d326f5ebc20cf4aa06f0
a1cffc29f0c31ed34830db24be
Successfully built thrift
Installing collected packages: thrift
Successfully installed thrift-0.15.0
Note: you may need to restart the kernel to use updated packages.
```

```
[6]: pip install thrift-sasl
```

```
Collecting thrift-sasl
  Downloading https://files.pythonhosted.org/packages/c3/9e/636c24ce1c0d46ce3020
c5836c5a375d8e862fa81a240e0e352cc991dcf8/thrift_sasl-0.4.3-py2.py3-none-any.whl
Requirement already satisfied: six>=1.13.0 in
/opt/conda/anaconda/lib/python3.7/site-packages (from thrift-sasl) (1.15.0)
Requirement already satisfied: thrift>=0.10.0; python_version >= "3.0" in
/opt/conda/anaconda/lib/python3.7/site-packages (from thrift-sasl) (0.15.0)
Collecting pure-sasl>=0.6.2 (from thrift-sasl)
  Downloading https://files.pythonhosted.org/packages/83/b7/a0d688f86c869073cc28
c0640899394a1cf68a6d87ee78a09565e9037da6/pure-sasl-0.6.2.tar.gz
Building wheels for collected packages: pure-sasl
  Building wheel for pure-sasl (setup.py) ... done
  Created wheel for pure-sasl: filename=pure_sasl-0.6.2-cp37-none-any.whl
size=11429
sha256=f61fd8e87789aa8e81fc160422a3af3b39496a220a56f5d84c166051911d9365
  Stored in directory: /root/.cache/pip/wheels/8d/34/73/6ed9e69a7ffd204e8d71bf42
2f9fa028f551d0dac431dc1082
Successfully built pure-sasl
Installing collected packages: pure-sasl, thrift-sasl
Successfully installed pure-sasl-0.6.2 thrift-sasl-0.4.3
Note: you may need to restart the kernel to use updated packages.
```

Importing Packages

```
[7]: from pyhive import hive
      from sklearn.feature_extraction.text import TfidfVectorizer
      from sklearn.feature_extraction.text import CountVectorizer
```

```
[ ]:
```

```
[10]: pip install sasl
```

```
Collecting sasl
  Downloading https://files.pythonhosted.org/packages/df/ae/d8dda9ef1636f5
48935c271910d3b35afbf1782df582fda88a13ea48de53/sasl-0.3.1.tar.gz (44kB)
    |                                     | 51kB 5.0MB/s eta 0:00:011
Requirement already satisfied: six in
```

```

/opt/conda/anaconda/lib/python3.7/site-packages (from sasl) (1.15.0)
Building wheels for collected packages: sasl
  Building wheel for sasl (setup.py) ... done
  Created wheel for sasl: filename=sasl-0.3.1-cp37-cp37m-linux_x86_64.whl
size=232743
sha256=61c026e44334044c1647063c09e061354928418c24bc03da6a9d86bd77e05656
  Stored in directory: /root/.cache/pip/wheels/a3/2e/2f/d341ce73b59f464dd4c03e2b
833712c0392a2bed0b7502a5bb
Successfully built sasl
Installing collected packages: sasl
Successfully installed sasl-0.3.1
Note: you may need to restart the kernel to use updated packages.

```

[ ]:

```

[19]: dft2 = pd.read_sql("SELECT OWNERUSERID,OWNERDISPNAME,SUM(SCORE) AS TOTAL_SCORE_
↳FROM AARTI GROUP BY OWNERUSERID,OWNERDISPNAME ORDER BY TOTAL_SCORE DESC_
↳LIMIT 10", conn)
display(dft2)

```

	owneruserid	ownerdispname	total_score
0	87234	GManNickG	37672.0
1	4883	readonly	28817.0
2	9951	e-satis	26969.0
3	6068	pupeno	25944.0
4	51816	Joan Venge	24420.0
5	89904	Hamza Yerlikaya	24024.0
6	49153	Ali	20203.0
7	179736	TIMEX	19603.0
8	95592	Matthew Rankin	19479.0
9	63051	flybywire	19362.0

```

[22]: dft4 = pd.read_sql('''SELECT OWNERUSERID, OWNERDISPNAME,TITLE,BODY FROM AARTI_
↳WHERE OWNERUSERID IN_
↳(87234,4883,9951,6068,89904,51816,49153,179736,95592,63051)
'', conn)
display(dft4)

```

	owneruserid	ownerdispname	\
0	63051		
1	89904		
2	4883	Readonly	
3	49153		
4	179736		
..	...	...	
691	51816	Joan Venge	
692	9951	e-satis	

```

693      51816      Joan Venge
694      179736      TIMEX
695      63051      flybywire

```

```

                                title \
0          Include JSTL dependency with Maven
1          Java Iterate Bits in Byte Array
2          List comprehension in Ruby
3      Validating alphabetic only string in javascript
4      How do I use a dictionary to update fields in ...
..
691          Clamping floating numbers in Python?
692      Is there any Python-like interactive console f...
693          Checking delegates for null
694      How can I remove all nil elements in a Swift a...
695      What are the advantages and disadvantages of j...

```

```

                                body
0      I am using maven how do I add a dependency to...
1          How can i iterate bits in a byte array
2      To do the equivalent of Python list comprehens...
3      How can I quickly validate if a string is alph...
4      Suppose I have a model like this class Bookmo...
..
691      Is there a built in function for this in Pytho...
692      I spent a lot of time programming in Java rece...
693      I was reading the Essential C book and am...
694      Basic way doesn t work for index in lt ...
695      What are the advantages and disadvantages of j...

```

[696 rows x 4 columns]

```

[24]: dft4['title_body'] = dft4[['title', 'body']].agg(' '.join, axis=1)
del dft4['title']
del dft4['body']
display(dft4)

```

```

owneruserid ownerdispname \
0      63051
1      89904
2      4883      Readonly
3      49153
4      179736
..      ...
691      51816      Joan Venge
692      9951      e-satis
693      51816      Joan Venge

```

```
694      179736      TIMEX
695      63051      flybywire
```

```

                                title_body
0   Include JSTL dependency with Maven I am using ...
1   Java Iterate Bits in Byte Array How can i iter...
2   List comprehension in Ruby To do the equivalen...
3   Validating alphabetic only string in javascrip...
4   How do I use a dictionary to update fields in ...
..
691 Clamping floating numbers in Python? Is there ...
692 Is there any Python-like interactive console f...
693 Checking delegates for null I was reading the ...
694 How can I remove all nil elements in a Swift a...
695 What are the advantages and disadvantages of j...
```

```
[696 rows x 3 columns]
```

```
[29]: tfidf_vectorizer = TfidfVectorizer(stop_words='english', lowercase=True)
#extracting users
userlist = list(dft4["ownerdispname"].unique())

for user in userlist:
    user_posts_df=dft4[dft4["ownerdispname"] == user]
    response = tfidf_vectorizer.fit_transform(user_posts_df["title_body"])
    tf_idf_df= pd.DataFrame(response.toarray(),columns=tfidf_vectorizer.
    ↪get_feature_names())
    sums = tf_idf_df.sum(axis=0)
    topTen= sums.nlargest(10)
    topTenWordsList = list(topTen.index)
    tf_idf_topTen=tf_idf_df[topTenWordsList]
    display("TF-IDF of the top 10 terms usre : "+user)
    display(tf_idf_topTen)
```

```
'TF-IDF of the top 10 terms usre : '
```

	python	file	way	using	use	user	want	like \
0	0.0	0.000000	0.000000	0.124140	0.000000	0.0	0.0	0.000000
1	0.0	0.000000	0.000000	0.000000	0.000000	0.0	0.0	0.000000
2	0.0	0.000000	0.000000	0.000000	0.000000	0.0	0.0	0.077719
3	0.0	0.000000	0.000000	0.096050	0.100336	0.0	0.0	0.100336
4	0.0	0.000000	0.03861	0.040861	0.042685	0.0	0.0	0.000000
..	...	...	...	...	...	...	...	...
209	0.0	0.000000	0.000000	0.000000	0.000000	0.0	0.0	0.123180
210	0.0	0.000000	0.000000	0.000000	0.291836	0.0	0.0	0.000000
211	0.0	0.000000	0.000000	0.000000	0.000000	0.0	0.0	0.000000

212	0.0	0.000000	0.000000	0.000000	0.109830	0.0	0.0	0.000000
213	0.0	0.661956	0.000000	0.000000	0.186654	0.0	0.0	0.000000

	array	class
0	0.000000	0.000000
1	0.402003	0.000000
2	0.000000	0.000000
3	0.000000	0.118611
4	0.000000	0.000000
..	...	...
209	0.000000	0.000000
210	0.000000	0.000000
211	0.000000	0.000000
212	0.000000	0.129835
213	0.000000	0.000000

[214 rows x 10 columns]

'TF-IDF of the top 10 terms usre : Readonly'

	blob	list	way	commits	commit	following	rename	\
0	0.000000	0.398472	0.176358	0.000000	0.000000	0.176358	0.000000	
1	0.787562	0.198620	0.175813	0.226534	0.226534	0.000000	0.000000	
2	0.000000	0.398472	0.176358	0.000000	0.000000	0.176358	0.000000	
3	0.000000	0.000000	0.109294	0.352060	0.140824	0.054647	0.000000	
4	0.000000	0.000000	0.000000	0.000000	0.000000	0.054274	0.000000	
5	0.000000	0.000000	0.000000	0.000000	0.000000	0.117753	0.000000	
6	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.57735	
7	0.787562	0.198620	0.175813	0.226534	0.226534	0.000000	0.000000	

	sqlite	table	arrayselectx
0	0.000000	0.000000	0.263335
1	0.000000	0.000000	0.000000
2	0.000000	0.000000	0.263335
3	0.000000	0.000000	0.000000
4	0.000000	0.000000	0.000000
5	0.000000	0.000000	0.000000
6	0.57735	0.57735	0.000000
7	0.000000	0.000000	0.000000

'TF-IDF of the top 10 terms usre : Joan Venge'

	python	class	like	list	string	gt	want	function	\
0	0.000000	0.000000	0.000000	0.000000	0.000000	0.0	0.000000	0.000000	
1	0.000000	0.000000	0.000000	0.148323	0.037936	0.0	0.000000	0.000000	
2	0.000000	0.728811	0.000000	0.000000	0.000000	0.0	0.103430	0.000000	

3	0.000000	0.000000	0.000000	0.000000	0.000000	0.0	0.000000	0.000000
4	0.000000	0.000000	0.000000	0.148323	0.037936	0.0	0.000000	0.000000
..	...	...	...	...	...	...	...	...
73	0.103065	0.000000	0.000000	0.000000	0.000000	0.0	0.000000	0.000000
74	0.000000	0.000000	0.090704	0.000000	0.000000	0.0	0.061809	0.000000
75	0.000000	0.728811	0.000000	0.000000	0.000000	0.0	0.103430	0.000000
76	0.257971	0.000000	0.123205	0.000000	0.000000	0.0	0.000000	0.214615
77	0.000000	0.060923	0.000000	0.000000	0.000000	0.0	0.000000	0.000000

	map	possible
0	0.000000	0.157678
1	0.714935	0.000000
2	0.000000	0.000000
3	0.000000	0.157678
4	0.714935	0.000000
..	...	...
73	0.000000	0.000000
74	0.000000	0.000000
75	0.000000	0.000000
76	0.000000	0.000000
77	0.000000	0.000000

[78 rows x 10 columns]

'TF-IDF of the top 10 terms usre : Tim'

	dictionary	inserted	items	order	retrieve	python	possible	\
0	0.437617	0.437617	0.437617	0.437617	0.437617	0.138177	0.152846	
1	0.437617	0.437617	0.437617	0.437617	0.437617	0.138177	0.152846	
2	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
3	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.118586	
4	0.000000	0.000000	0.000000	0.000000	0.000000	0.081021	0.000000	
5	0.000000	0.000000	0.000000	0.000000	0.000000	0.138558	0.000000	
6	0.000000	0.000000	0.000000	0.000000	0.000000	0.208313	0.000000	
7	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.082907	
8	0.000000	0.000000	0.000000	0.000000	0.000000	0.068789	0.000000	
9	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.262791	

	directories	interface	method
0	0.000000	0.000000	0.000000
1	0.000000	0.000000	0.000000
2	0.000000	0.000000	0.000000
3	0.599099	0.000000	0.000000
4	0.000000	0.000000	0.000000
5	0.000000	0.000000	0.000000
6	0.000000	0.000000	0.439831
7	0.000000	0.558462	0.000000

8	0.000000	0.000000	0.108931
9	0.000000	0.000000	0.000000

'TF-IDF of the top 10 terms usre : andy'

	reboot	file	advantages	ajax	disadvantages	json \
0	0.507093	0.000000	0.000000	0.000000	0.000000	0.000000
1	0.000000	0.000000	0.342687	0.342687	0.342687	0.342687
2	0.507093	0.000000	0.000000	0.000000	0.000000	0.000000
3	0.000000	0.000000	0.342687	0.342687	0.342687	0.342687
4	0.000000	0.356118	0.000000	0.000000	0.000000	0.000000
5	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
6	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
7	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
8	0.000000	0.228787	0.000000	0.000000	0.000000	0.000000
9	0.000000	0.310595	0.000000	0.000000	0.000000	0.000000

	requests	vs	xml	happens
0	0.000000	0.000000	0.000000	0.338062
1	0.342687	0.342687	0.299810	0.000000
2	0.000000	0.000000	0.000000	0.338062
3	0.342687	0.342687	0.299810	0.000000
4	0.000000	0.000000	0.000000	0.000000
5	0.000000	0.000000	0.000000	0.000000
6	0.000000	0.000000	0.000000	0.000000
7	0.000000	0.000000	0.084697	0.000000
8	0.000000	0.000000	0.000000	0.000000
9	0.000000	0.000000	0.000000	0.000000

'TF-IDF of the top 10 terms usre : TIMEX'

	python	want	user	string	gt	dictionary	lt \
0	0.000000	0.000000	0.537874	0.0	0.00000	0.000000	0.000000
1	0.000000	0.000000	0.000000	0.0	0.00000	0.000000	0.000000
2	0.000000	0.000000	0.000000	0.0	0.00000	0.000000	0.000000
3	0.097745	0.092713	0.000000	0.0	0.00000	0.279298	0.000000
4	0.045531	0.000000	0.000000	0.0	0.35595	0.000000	0.371895
..	...	...	...	...	...	...	...
110	0.071606	0.000000	0.000000	0.0	0.00000	0.000000	0.000000
111	0.053461	0.000000	0.000000	0.0	0.00000	0.076380	0.000000
112	0.000000	0.032024	0.000000	0.0	0.00000	0.000000	0.000000
113	0.000000	0.000000	0.132608	0.0	0.00000	0.000000	0.063177
114	0.000000	0.000000	0.000000	0.0	0.00000	0.000000	0.113867

	use	javascript	return
0	0.000000	0.000000	0.000000



1	0.000000	0.000000	0.000000
2	0.000000	0.275821	0.000000
3	0.120016	0.000000	0.000000
4	0.055905	0.000000	0.000000
..	...	...	...
110	0.000000	0.099799	0.210055
111	0.000000	0.000000	0.000000
112	0.000000	0.000000	0.000000
113	0.000000	0.000000	0.068069
114	0.000000	0.000000	0.000000

[115 rows x 10 columns]

'TF-IDF of the top 10 terms usre : pupeno'

	file	java	android	sql	git	way	like \
0	0.000000	0.000000	0.000000	0.000000	0.143193	0.088019	0.091283
1	0.000000	0.071220	0.000000	0.000000	0.000000	0.050015	0.000000
2	0.000000	0.000000	0.000000	0.000000	0.000000	0.081616	0.000000
3	0.000000	0.531605	0.000000	0.000000	0.000000	0.000000	0.064529
4	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.089201
5	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
6	0.000000	0.000000	0.000000	0.000000	0.000000	0.081609	0.000000
7	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
8	0.000000	0.000000	0.000000	0.000000	0.728914	0.000000	0.000000
9	0.000000	0.000000	0.130150	0.000000	0.000000	0.000000	0.000000
10	0.000000	0.000000	0.174881	0.000000	0.000000	0.000000	0.000000
11	0.000000	0.000000	0.000000	0.000000	0.000000	0.140540	0.000000
12	0.000000	0.362108	0.000000	0.000000	0.000000	0.000000	0.000000
13	0.000000	0.000000	0.275010	0.000000	0.000000	0.000000	0.000000
14	0.000000	0.000000	0.000000	0.219137	0.000000	0.000000	0.000000
15	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.032758
16	0.000000	0.000000	0.151689	0.000000	0.000000	0.000000	0.000000
17	0.000000	0.000000	0.337085	0.000000	0.000000	0.125140	0.000000
18	0.000000	0.000000	0.000000	0.000000	0.000000	0.079078	0.164020
19	0.000000	0.000000	0.000000	0.105631	0.000000	0.000000	0.072300
20	0.656317	0.000000	0.000000	0.000000	0.417075	0.085457	0.000000
21	0.108337	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	0.371197	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
24	0.000000	0.000000	0.352152	0.000000	0.000000	0.000000	0.000000
25	0.000000	0.000000	0.000000	0.000000	0.000000	0.084689	0.263487
26	0.000000	0.364640	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	0.312392	0.000000	0.000000	0.000000	0.000000	0.122027	0.126552
29	0.082427	0.366784	0.000000	0.000000	0.000000	0.000000	0.066783
30	0.000000	0.000000	0.000000	0.000000	0.000000	0.083640	0.000000

31	0.000000	0.000000	0.000000	0.000000	0.000000	0.026801	0.055589
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	0.187244	0.000000	0.000000	0.277057	0.000000	0.109712	0.000000
36	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.021477

	application	data	dump
0	0.000000	0.000000	0.000000
1	0.000000	0.000000	0.000000
2	0.000000	0.000000	0.000000
3	0.177202	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	0.000000	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	0.436139	0.261367	0.000000
14	0.000000	0.000000	0.000000
15	0.000000	0.040432	0.000000
16	0.000000	0.000000	0.000000
17	0.356390	0.000000	0.000000
18	0.000000	0.000000	0.000000
19	0.000000	0.267708	0.617252
20	0.000000	0.000000	0.000000
21	0.000000	0.123813	0.000000
22	0.000000	0.000000	0.000000
23	0.000000	0.000000	0.000000
24	0.000000	0.000000	0.000000
25	0.000000	0.000000	0.000000
26	0.000000	0.000000	0.000000
27	0.000000	0.095551	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	0.116987	0.323682
33	0.026204	0.000000	0.000000
34	0.000000	0.000000	0.000000
35	0.000000	0.093622	0.000000
36	0.029489	0.000000	0.000000

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

	install	pip	python	lt	gt	branch	installed \
0	0.427251	0.284834	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	0.338392	0.000000	0.000000	0.000000	0.000000
3	0.000000	0.000000	0.000000	0.000000	0.043561	0.571753	0.000000
4	0.000000	0.000000	0.000000	0.650278	0.544984	0.000000	0.000000
5	0.349045	0.319958	0.319958	0.000000	0.000000	0.000000	0.000000
6	0.235219	0.211697	0.094087	0.000000	0.000000	0.000000	0.109034
7	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.447027

	flask	version	django
0	0.000000	0.000000	0.000000
1	0.000000	0.000000	0.425746
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	0.054517
7	0.533396	0.533396	0.000000

'TF-IDF of the top 10 terms usre : readonly'

	python	use	ruby	list	table	branch	way \
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	0.070218	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
3	0.000000	0.202521	0.000000	0.000000	0.259890	0.000000	0.101261
4	0.000000	0.000000	0.000000	0.000000	0.521156	0.000000	0.000000
5	0.000000	0.000000	0.000000	0.202044	0.000000	0.000000	0.169489
6	0.163253	0.000000	0.223728	0.447455	0.000000	0.000000	0.187678
7	0.121700	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
8	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
9	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
10	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
11	0.000000	0.000000	0.413988	0.000000	0.000000	0.000000	0.000000
12	0.000000	0.000000	0.000000	0.000000	0.000000	0.636117	0.000000
13	0.152171	0.174938	0.000000	0.000000	0.000000	0.000000	0.000000
14	0.000000	0.191672	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	0.204627	0.000000	0.000000	0.000000	0.000000
17	0.251336	0.000000	0.000000	0.000000	0.000000	0.000000	0.144470
18	0.195328	0.112276	0.000000	0.000000	0.000000	0.000000	0.000000
19	0.000000	0.000000	0.000000	0.135482	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	0.404077	0.000000	0.000000	0.000000	0.000000

22	0.101808	0.234079	0.000000	0.418562	0.000000	0.000000	0.000000
23	0.000000	0.120988	0.000000	0.000000	0.310520	0.000000	0.120988
24	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
25	0.000000	0.213146	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	0.088347	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
28	0.149597	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
29	0.194472	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	0.179925
31	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
32	0.000000	0.000000	0.000000	0.000000	0.000000	0.402695	0.114986

	difference	git	process
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	0.389835
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	0.000000	0.350231
9	0.000000	0.203342	0.000000
10	0.000000	0.149441	0.000000
11	0.000000	0.000000	0.000000
12	0.000000	0.202994	0.000000
13	0.000000	0.000000	0.000000
14	0.491935	0.000000	0.000000
15	0.000000	0.000000	0.000000
16	0.000000	0.000000	0.000000
17	0.000000	0.000000	0.000000
18	0.288162	0.000000	0.000000
19	0.000000	0.000000	0.000000
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	0.000000
24	0.000000	0.000000	0.187558
25	0.000000	0.238205	0.000000
26	0.000000	0.000000	0.000000
27	0.000000	0.000000	0.000000
28	0.220697	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	0.192758	0.000000

'TF-IDF of the top 10 terms usre : flybywire'

	file	python	want	java	vs	gt	use \
0	0.237010	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
1	0.000000	0.000000	0.000000	0.000000	0.194207	0.000000	0.388413
2	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
3	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
4	0.297907	0.000000	0.139056	0.000000	0.000000	0.000000	0.175732
5	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
6	0.000000	0.000000	0.112601	0.000000	0.000000	0.000000	0.000000
7	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
8	0.000000	0.120377	0.000000	0.000000	0.000000	0.000000	0.000000
9	0.399837	0.000000	0.000000	0.000000	0.000000	0.300322	0.157240
10	0.166264	0.000000	0.155217	0.000000	0.000000	0.000000	0.000000
11	0.000000	0.000000	0.000000	0.657184	0.000000	0.000000	0.000000
12	0.099270	0.111844	0.000000	0.000000	0.000000	0.000000	0.000000
13	0.000000	0.461401	0.000000	0.000000	0.000000	0.000000	0.000000
14	0.000000	0.000000	0.123087	0.000000	0.000000	0.000000	0.000000
15	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
16	0.227342	0.000000	0.106118	0.000000	0.000000	0.000000	0.000000
17	0.000000	0.146583	0.000000	0.000000	0.153493	0.000000	0.306986
18	0.076246	0.000000	0.284718	0.000000	0.000000	0.000000	0.000000
19	0.000000	0.112790	0.000000	0.000000	0.000000	0.225579	0.000000
20	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
21	0.000000	0.000000	0.000000	0.544821	0.229172	0.000000	0.114586
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	0.000000
25	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
26	0.000000	0.000000	0.086403	0.000000	0.000000	0.052138	0.054596
27	0.000000	0.000000	0.000000	0.000000	0.228055	0.000000	0.000000
28	0.071389	0.000000	0.066645	0.000000	0.000000	0.000000	0.000000
29	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
30	0.000000	0.000000	0.116217	0.000000	0.000000	0.280514	0.000000
31	0.000000	0.000000	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	0.315679	0.000000	0.098234	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	0.000000	0.000000	0.000000	0.000000	0.000000
36	0.000000	0.316739	0.000000	0.000000	0.000000	0.000000	0.000000
37	0.000000	0.000000	0.000000	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	0.000000	0.000000	0.068348	0.000000	0.000000	0.082486	0.000000
41	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
42	0.175813	0.000000	0.082066	0.000000	0.000000	0.198083	0.000000
43	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000

44	0.000000	0.000000	0.000000	0.138204	0.000000	0.111033	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	0.000000	0.220426	0.000000	0.000000
48	0.000000	0.000000	0.000000	0.000000	0.255143	0.000000	0.000000

	bash	instance	standard
0	0.000000	0.000000	0.000000
1	0.000000	0.000000	0.000000
2	0.000000	0.000000	0.207214
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	0.000000	0.000000
9	0.165412	0.000000	0.498416
10	0.412698	0.000000	0.466327
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	0.000000	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	0.000000	0.000000
21	0.000000	0.000000	0.000000
22	0.000000	0.000000	0.000000
23	0.000000	0.000000	0.000000
24	0.000000	0.864911	0.000000
25	0.000000	0.000000	0.000000
26	0.172300	0.000000	0.000000
27	0.000000	0.000000	0.000000
28	0.177199	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	0.261191	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	0.000000	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 usre : e-satis'

	python	git	head	like	does	using	javascript \
0	0.000000	0.000000	0.145378	0.000000	0.000000	0.000000	0.268457
1	0.000000	0.000000	0.000000	0.000000	0.000000	0.313889	0.143533
2	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
3	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.319433
4	0.000000	0.070830	0.000000	0.060079	0.000000	0.000000	0.000000
5	0.000000	0.000000	0.000000	0.102714	0.121093	0.000000	0.276864
6	0.527016	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
7	0.445918	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
8	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
9	0.000000	0.000000	0.000000	0.000000	0.386554	0.000000	0.000000
10	0.000000	0.000000	0.000000	0.000000	0.000000	0.119441	0.000000
11	0.000000	0.000000	0.000000	0.000000	0.229084	0.000000	0.000000
12	0.000000	0.423709	0.349742	0.000000	0.141236	0.000000	0.000000
13	0.106858	0.000000	0.000000	0.226598	0.000000	0.000000	0.000000
14	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
15	0.000000	0.000000	0.000000	0.157317	0.000000	0.000000	0.000000
16	0.000000	0.076126	0.000000	0.000000	0.000000	0.152251	0.000000
17	0.000000	0.000000	0.000000	0.170587	0.000000	0.000000	0.000000
18	0.000000	0.000000	0.000000	0.000000	0.000000	0.237973	0.000000
19	0.000000	0.188855	0.701487	0.000000	0.000000	0.000000	0.000000
20	0.078006	0.000000	0.000000	0.132333	0.156013	0.000000	0.000000
21	0.000000	0.000000	0.000000	0.035120	0.000000	0.165618	0.000000
22	0.000000	0.339968	0.000000	0.000000	0.000000	0.000000	0.000000
23	0.110475	0.000000	0.000000	0.093707	0.000000	0.000000	0.000000
24	0.000000	0.189044	0.000000	0.000000	0.063015	0.000000	0.000000
25	0.080780	0.000000	0.000000	0.205558	0.000000	0.080780	0.000000

	gt	use	way
0	0.319505	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
5	0.000000	0.000000	0.000000
6	0.000000	0.263508	0.000000

7	0.000000	0.000000	0.280386
8	0.000000	0.000000	0.064802
9	0.000000	0.000000	0.145836
10	0.000000	0.119441	0.000000
11	0.000000	0.114542	0.000000
12	0.000000	0.141236	0.000000
13	0.654243	0.000000	0.100786
14	0.000000	0.000000	0.000000
15	0.000000	0.000000	0.174929
16	0.000000	0.000000	0.000000
17	0.000000	0.201112	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	0.124213	0.117155
22	0.000000	0.000000	0.000000
23	0.000000	0.000000	0.000000
24	0.000000	0.000000	0.059434
25	0.000000	0.000000	0.000000

'TF-IDF of the top 10 terms usre : Ali'

	gt	array	javascript	php	using	lt	jquery	id	key	\
0	0.0	0.0	0.000000	0.0	0.000000	0.0	0.0	0.000000	0.0	
1	0.0	0.0	0.104572	0.0	0.000000	0.0	0.0	0.085724	0.0	
2	0.0	0.0	0.000000	0.0	0.063737	0.0	0.0	0.000000	0.0	
3	0.0	0.0	0.096714	0.0	0.082236	0.0	0.0	0.000000	0.0	
4	0.0	0.0	0.000000	0.0	0.000000	0.0	0.0	0.000000	0.0	
..	...	...	...	...	...	...	...	...	...	
74	0.0	0.0	0.000000	0.0	0.000000	0.0	0.0	0.200529	0.0	
75	0.0	0.0	0.000000	0.0	0.000000	0.0	0.0	0.000000	0.0	
76	0.0	0.0	0.000000	0.0	0.000000	0.0	0.0	0.000000	0.0	
77	0.0	0.0	0.000000	0.0	0.000000	0.0	0.0	0.167844	0.0	
78	0.0	0.0	0.000000	0.0	0.070152	0.0	0.0	0.000000	0.0	

	file
0	0.000000
1	0.000000
2	0.000000
3	0.000000
4	0.000000
..	...
74	0.000000
75	0.000000
76	0.177553
77	0.000000
78	0.000000



[79 rows x 10 columns]

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

	file	timer	files	java	new	table	application \
0	0.000000	0.000000	0.097851	0.000000	0.000000	0.000000	0.251599
1	0.000000	0.000000	0.000000	0.146846	0.000000	0.000000	0.000000
2	0.000000	0.000000	0.084239	0.000000	0.000000	0.000000	0.000000
3	0.000000	0.000000	0.000000	0.026395	0.000000	0.533399	0.000000
4	0.351468	0.000000	0.312480	0.000000	0.000000	0.000000	0.000000
5	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
6	0.463315	0.000000	0.102980	0.000000	0.000000	0.000000	0.264787
7	0.000000	0.000000	0.000000	0.284629	0.000000	0.000000	0.000000
8	0.000000	0.782791	0.000000	0.066405	0.190127	0.000000	0.000000
9	0.064939	0.000000	0.000000	0.051849	0.371129	0.000000	0.000000

	jar	array	bits
0	0.251599	0.000000	0.000000
1	0.000000	0.49458	0.49458
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	0.264787	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

'TF-IDF of the top 10 terms usre : Click Upvote'

	array	session	gt	way	alternative	php	extract \
0	0.000000	0.000000	0.000000	0.000000	0.51278	0.101404	0.000000
1	0.396229	0.000000	0.000000	0.396229	0.000000	0.286662	0.000000
2	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
3	0.000000	0.598859	0.000000	0.000000	0.000000	0.088820	0.000000
4	0.521282	0.000000	0.529749	0.130320	0.000000	0.031428	0.000000
5	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.489813

	image	sprite	alphabetic
0	0.000000	0.000000	0.000000
1	0.000000	0.000000	0.483198
2	0.000000	0.000000	0.000000
3	0.000000	0.000000	0.000000
4	0.000000	0.000000	0.000000
5	0.489813	0.489813	0.000000

'TF-IDF of the top 10 terms usre : J. Pablo Fern&#225;ndez'

	java	clojure	gitignore	use	way	time	emacs \
0	0.000000	0.000000	0.000000	0.386479	0.000000	0.000000	0.000000
1	0.302215	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
2	0.000000	0.000000	0.000000	0.000000	0.144990	0.000000	0.383273
3	0.322419	0.319611	0.000000	0.000000	0.000000	0.000000	0.106537
4	0.000000	0.000000	0.000000	0.000000	0.108882	0.143911	0.000000
5	0.000000	0.000000	0.505586	0.000000	0.106861	0.000000	0.000000
6	0.131316	0.000000	0.000000	0.000000	0.131316	0.347126	0.000000
7	0.331672	0.219189	0.000000	0.109595	0.000000	0.000000	0.000000

	file	difference	double
0	0.000000	0.00000	0.00000
1	0.000000	0.47662	0.47662
2	0.383273	0.00000	0.00000
3	0.106537	0.00000	0.00000
4	0.000000	0.00000	0.00000
5	0.000000	0.00000	0.00000
6	0.000000	0.00000	0.00000
7	0.000000	0.00000	0.00000

'TF-IDF of the top 10 terms usre : Java PHP'

	used	case	letter	java	lower	start	byte \
0	0.39896	0.000000	0.000000	0.189242	0.000000	0.000000	0.265973
1	0.00000	0.398917	0.398917	0.141916	0.299188	0.299188	0.000000

  

	char	double	long
0	0.265973	0.265973	0.265973
1	0.000000	0.000000	0.000000

'TF-IDF of the top 10 terms usre : GManNickG'

	lt	copy	data	int	idiom	gt	quot \
0	0.431818	0.000000	0.51101	0.172727	0.000000	0.129546	0.086364
1	0.126037	0.000000	0.00000	0.252073	0.000000	0.252073	0.252073
2	0.000000	0.523691	0.00000	0.000000	0.418953	0.000000	0.000000

  

	swap	arraysize	array
0	0.000000	0.283895	0.194318
1	0.000000	0.000000	0.000000
2	0.314215	0.000000	0.079656

[ ]: