

# tf\_idf\_cal

October 27, 2021

Name : Kaustubh Agawane

Student ID : 21260460

Email : kaustubh.agawane2@mail.dcu.ie

Assignment 1: Data Analysis

Task 4 : Calculate TF-IDF with MapReduce/Pig/Hive

## 1) Package Imports

```
[1]: from pyhive import hive
import pandas as pd
from sklearn.feature_extraction.text import TfidfVectorizer
```

## 2) Hive Connection setup

```
[2]: host="127.0.0.1"
port=10000
username="kaustubh_agawane2"
password="5567170637024851622"
conn = hive.Connection(host, port, username, password)
```

## 3) Finding top 10 users to calculate the per-user TF-IDF

```
[3]: data = pd.read_sql("select OwnerUserId, max(DisplayName) as Displayname,
    ↳sum(score) as score from stackexpsts group by OwnerUserId order by score
    ↳desc limit 10", conn)
display(data)
```

	owneruserid	displayname	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

#### 4) Selecting Post Body and Titles for each of top 10 users

```
[4]: owner_user_id=list(data['owneruserid'])
query=f"select owneruserid,displayname,body,title,score from stackexposts where
      ↳owneruserid in {tuple(owner_user_id)}"
df = pd.read_sql(query, conn)
df['Body-Title']=df['body'] + ' ' + df['title']
```

#### 5) Function to calculate TF-IDF on post Body and Title fields and finding top 10 terms for each of the top 10 user

Steps performed in below function Step 1 : Passing Title dataframe

Step 2 & Step 3 : TfidfVectorizer.fit\_transform() function filters the stop words, cleans the data for unwanted characters and calculates TF-IDF on entire list of Title dataframe for each term.

Step 4 : Creating a new dataframe after calculating TF-IDF

Step 5 : To find the top 10 terms, we use dataframe sum() function to sum TF-IDF for every term and using nlargest() function we find top 10 terms.

```
[5]: def calc_tf_idf(pass_df):
      df_tfidf=pass_df
      vectorizer = TfidfVectorizer(stop_words='english', lowercase=True)
      stop_words=vectorizer.fit_transform(df_tfidf) #Passing Title Dataframe to
      ↳calculate TF-IDF
      df_calculated_tfidf = pd.DataFrame(stop_words.toarray(),columns=vectorizer.
      ↳get_feature_names())
      sum_of_tf_idf=df_calculated_tfidf.sum(axis = 0)
      calc_top_10=sum_of_tf_idf.nlargest(10)
      find_top10_words = list(calc_top_10.index)
      return df_calculated_tfidf[find_top10_words]
```

#### 6) Displaying Top 10 terms TF-IDF for each of top 10 users

```
[6]: for i in data.index:
      extract_namer=data['displayname'].values[i]
      df1=df.query(f"displayname == '{extract_namer}'")
```

```

df2=calc_tf_idf(df1['Body-Title'])
print("User ID:",extact_namer + "                                Top 10 Words\
↪TF-IDF")

↵
↪print("-----")
df2.insert(0,'displayname',extact_namer)
display(df2)

```

User ID: GManNickG

Top 10 Words TF-IDF

```

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

      quot      swap  arraysize      array
0  0.086364  0.000000   0.283895  0.194318
1  0.252073  0.000000   0.000000  0.000000
2  0.000000  0.314215   0.000000  0.079656

```

User ID: readonly

Top 10 Words TF-IDF

```

-----
displayname      python      use      ruby      list      table      branch \
0  readonly  0.000000  0.000000  0.000000  0.000000  0.000000  0.000000
1  readonly  0.000000  0.000000  0.000000  0.000000  0.000000  0.000000
2  readonly  0.000000  0.000000  0.000000  0.000000  0.000000  0.000000
3  readonly  0.000000  0.000000  0.413988  0.000000  0.000000  0.000000
4  readonly  0.000000  0.000000  0.000000  0.000000  0.000000  0.636117
5  readonly  0.152171  0.174938  0.000000  0.000000  0.000000  0.000000
6  readonly  0.000000  0.191672  0.000000  0.000000  0.000000  0.000000
7  readonly  0.000000  0.000000  0.000000  0.000000  0.000000  0.000000
8  readonly  0.000000  0.000000  0.204627  0.000000  0.000000  0.000000
9  readonly  0.251336  0.000000  0.000000  0.000000  0.000000  0.000000
10  readonly  0.195328  0.112276  0.000000  0.000000  0.000000  0.000000
11  readonly  0.000000  0.000000  0.000000  0.135482  0.000000  0.000000
12  readonly  0.000000  0.000000  0.000000  0.000000  0.000000  0.000000
13  readonly  0.000000  0.000000  0.404077  0.000000  0.000000  0.000000
14  readonly  0.101808  0.234079  0.000000  0.418562  0.000000  0.000000
15  readonly  0.000000  0.120988  0.000000  0.000000  0.310520  0.000000
16  readonly  0.000000  0.000000  0.000000  0.000000  0.000000  0.000000
17  readonly  0.000000  0.213146  0.000000  0.000000  0.000000  0.000000
18  readonly  0.000000  0.000000  0.000000  0.000000  0.000000  0.000000
19  readonly  0.088347  0.000000  0.000000  0.000000  0.000000  0.000000
20  readonly  0.149597  0.000000  0.000000  0.000000  0.000000  0.000000
21  readonly  0.194472  0.000000  0.000000  0.000000  0.000000  0.000000
22  readonly  0.000000  0.000000  0.000000  0.000000  0.000000  0.000000

```

23	readonly	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
24	readonly	0.000000	0.000000	0.000000	0.000000	0.000000	0.402695
25	readonly	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
26	readonly	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
27	readonly	0.070218	0.000000	0.000000	0.000000	0.000000	0.000000
28	readonly	0.000000	0.202521	0.000000	0.000000	0.259890	0.000000
29	readonly	0.000000	0.000000	0.000000	0.000000	0.521156	0.000000
30	readonly	0.000000	0.000000	0.000000	0.202044	0.000000	0.000000
31	readonly	0.163253	0.000000	0.223728	0.447455	0.000000	0.000000
32	readonly	0.121700	0.000000	0.000000	0.000000	0.000000	0.000000

	way	difference	git	process
0	0.000000	0.000000	0.000000	0.350231
1	0.000000	0.000000	0.203342	0.000000
2	0.000000	0.000000	0.149441	0.000000
3	0.000000	0.000000	0.000000	0.000000
4	0.000000	0.000000	0.202994	0.000000
5	0.000000	0.000000	0.000000	0.000000
6	0.000000	0.491935	0.000000	0.000000
7	0.000000	0.000000	0.000000	0.000000
8	0.000000	0.000000	0.000000	0.000000
9	0.144470	0.000000	0.000000	0.000000
10	0.000000	0.288162	0.000000	0.000000
11	0.000000	0.000000	0.000000	0.000000
12	0.000000	0.000000	0.000000	0.000000
13	0.000000	0.000000	0.000000	0.000000
14	0.000000	0.000000	0.000000	0.000000
15	0.120988	0.000000	0.000000	0.000000
16	0.000000	0.000000	0.000000	0.187558
17	0.000000	0.000000	0.238205	0.000000
18	0.000000	0.000000	0.000000	0.000000
19	0.000000	0.000000	0.000000	0.000000
20	0.000000	0.220697	0.000000	0.000000
21	0.000000	0.000000	0.000000	0.000000
22	0.179925	0.000000	0.000000	0.000000
23	0.000000	0.000000	0.000000	0.000000
24	0.114986	0.000000	0.192758	0.000000
25	0.000000	0.000000	0.000000	0.000000
26	0.000000	0.000000	0.000000	0.000000
27	0.000000	0.000000	0.000000	0.000000
28	0.101261	0.000000	0.000000	0.389835
29	0.000000	0.000000	0.000000	0.000000
30	0.169489	0.000000	0.000000	0.000000
31	0.187678	0.000000	0.000000	0.000000
32	0.000000	0.000000	0.000000	0.000000

User ID: e-satis

Top 10 Words TF-IDF

---

	displayname	python	git	head	like	does	using \
0	e-satis	0.000000	0.070489	0.000000	0.059616	0.000000	0.000000
1	e-satis	0.000000	0.000000	0.000000	0.101237	0.119702	0.000000
2	e-satis	0.525068	0.000000	0.000000	0.000000	0.000000	0.000000
3	e-satis	0.444462	0.000000	0.000000	0.000000	0.000000	0.000000
4	e-satis	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
5	e-satis	0.000000	0.000000	0.000000	0.000000	0.384412	0.000000
6	e-satis	0.000000	0.000000	0.000000	0.000000	0.000000	0.118825
7	e-satis	0.000000	0.000000	0.000000	0.000000	0.227740	0.000000
8	e-satis	0.000000	0.421761	0.349226	0.000000	0.140587	0.000000
9	e-satis	0.106392	0.000000	0.000000	0.224950	0.000000	0.000000
10	e-satis	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
11	e-satis	0.000000	0.000000	0.000000	0.156069	0.000000	0.000000
12	e-satis	0.000000	0.075851	0.000000	0.000000	0.000000	0.151701
13	e-satis	0.000000	0.000000	0.000000	0.169226	0.000000	0.000000
14	e-satis	0.000000	0.000000	0.000000	0.000000	0.000000	0.236845
15	e-satis	0.000000	0.188162	0.701109	0.000000	0.000000	0.000000
16	e-satis	0.077636	0.000000	0.000000	0.131321	0.155273	0.000000
17	e-satis	0.000000	0.000000	0.000000	0.034854	0.000000	0.164845
18	e-satis	0.000000	0.338695	0.000000	0.000000	0.000000	0.000000
19	e-satis	0.109937	0.000000	0.000000	0.092978	0.000000	0.000000
20	e-satis	0.000000	0.188113	0.000000	0.000000	0.062704	0.000000
21	e-satis	0.000000	0.000000	0.134811	0.000000	0.000000	0.000000
22	e-satis	0.000000	0.000000	0.000000	0.000000	0.000000	0.308775
23	e-satis	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
24	e-satis	0.080333	0.000000	0.000000	0.203822	0.000000	0.080333

	use	gt	way	know
0	0.000000	0.000000	0.000000	0.062828
1	0.000000	0.000000	0.000000	0.000000
2	0.262534	0.000000	0.000000	0.000000
3	0.000000	0.000000	0.279196	0.000000
4	0.000000	0.000000	0.064273	0.000000
5	0.000000	0.000000	0.144885	0.205578
6	0.118825	0.000000	0.000000	0.000000
7	0.113870	0.000000	0.000000	0.000000
8	0.140587	0.000000	0.000000	0.000000
9	0.000000	0.654206	0.100248	0.047414
10	0.000000	0.000000	0.000000	0.000000
11	0.000000	0.000000	0.173878	0.082239
12	0.000000	0.000000	0.000000	0.135212
13	0.200092	0.000000	0.000000	0.000000
14	0.000000	0.000000	0.000000	0.000000
15	0.000000	0.000000	0.000000	0.000000
16	0.000000	0.000000	0.000000	0.069198
17	0.123634	0.000000	0.116494	0.146928

18	0.000000	0.000000	0.000000	0.000000
19	0.000000	0.000000	0.000000	0.000000
20	0.000000	0.000000	0.059083	0.000000
21	0.000000	0.296631	0.000000	0.000000
22	0.000000	0.000000	0.000000	0.000000
23	0.000000	0.000000	0.000000	0.000000
24	0.000000	0.000000	0.000000	0.143203

User ID: pupeno

Top 10 Words TF-IDF

	displayname	file	java	android	sql	git	way \
0	pupeno	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
1	pupeno	0.000000	0.000000	0.000000	0.000000	0.728914	0.000000
2	pupeno	0.000000	0.000000	0.130150	0.000000	0.000000	0.000000
3	pupeno	0.000000	0.000000	0.174881	0.000000	0.000000	0.000000
4	pupeno	0.000000	0.000000	0.000000	0.000000	0.000000	0.140540
5	pupeno	0.000000	0.362108	0.000000	0.000000	0.000000	0.000000
6	pupeno	0.000000	0.000000	0.275010	0.000000	0.000000	0.000000
7	pupeno	0.000000	0.000000	0.000000	0.219137	0.000000	0.000000
8	pupeno	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
9	pupeno	0.000000	0.000000	0.151689	0.000000	0.000000	0.000000
10	pupeno	0.000000	0.000000	0.337085	0.000000	0.000000	0.125140
11	pupeno	0.000000	0.000000	0.000000	0.000000	0.000000	0.079078
12	pupeno	0.000000	0.000000	0.000000	0.105631	0.000000	0.000000
13	pupeno	0.656317	0.000000	0.000000	0.000000	0.417075	0.085457
14	pupeno	0.108337	0.000000	0.000000	0.000000	0.000000	0.000000
15	pupeno	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
16	pupeno	0.371197	0.000000	0.000000	0.000000	0.000000	0.000000
17	pupeno	0.000000	0.000000	0.352152	0.000000	0.000000	0.000000
18	pupeno	0.000000	0.000000	0.000000	0.000000	0.000000	0.084689
19	pupeno	0.000000	0.364640	0.000000	0.000000	0.000000	0.000000
20	pupeno	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
21	pupeno	0.312392	0.000000	0.000000	0.000000	0.000000	0.122027
22	pupeno	0.082427	0.366784	0.000000	0.000000	0.000000	0.000000
23	pupeno	0.000000	0.000000	0.000000	0.000000	0.000000	0.083640
24	pupeno	0.000000	0.000000	0.000000	0.000000	0.000000	0.026801
25	pupeno	0.000000	0.000000	0.000000	0.000000	0.143193	0.088019
26	pupeno	0.000000	0.071220	0.000000	0.000000	0.000000	0.050015
27	pupeno	0.000000	0.000000	0.000000	0.000000	0.000000	0.081616
28	pupeno	0.000000	0.531605	0.000000	0.000000	0.000000	0.000000
29	pupeno	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
30	pupeno	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
31	pupeno	0.000000	0.000000	0.000000	0.000000	0.000000	0.081609
32	pupeno	0.116987	0.000000	0.000000	0.692403	0.000000	0.000000
33	pupeno	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
34	pupeno	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
35	pupeno	0.187244	0.000000	0.000000	0.277057	0.000000	0.109712

36	pupeno	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
----	--------	----------	----------	----------	----------	----------	----------

	like	application	data	dump
0	0.000000	0.000000	0.000000	0.000000
1	0.000000	0.000000	0.000000	0.000000
2	0.000000	0.000000	0.000000	0.000000
3	0.000000	0.000000	0.000000	0.000000
4	0.000000	0.000000	0.000000	0.000000
5	0.000000	0.000000	0.000000	0.000000
6	0.000000	0.436139	0.261367	0.000000
7	0.000000	0.000000	0.000000	0.000000
8	0.032758	0.000000	0.040432	0.000000
9	0.000000	0.000000	0.000000	0.000000
10	0.000000	0.356390	0.000000	0.000000
11	0.164020	0.000000	0.000000	0.000000
12	0.072300	0.000000	0.267708	0.617252
13	0.000000	0.000000	0.000000	0.000000
14	0.000000	0.000000	0.123813	0.000000
15	0.000000	0.000000	0.000000	0.000000
16	0.000000	0.000000	0.000000	0.000000
17	0.000000	0.000000	0.000000	0.000000
18	0.263487	0.000000	0.000000	0.000000
19	0.000000	0.000000	0.000000	0.000000
20	0.000000	0.000000	0.095551	0.000000
21	0.126552	0.000000	0.000000	0.000000
22	0.066783	0.000000	0.000000	0.000000
23	0.000000	0.000000	0.000000	0.000000
24	0.055589	0.000000	0.000000	0.000000
25	0.091283	0.000000	0.000000	0.000000
26	0.000000	0.000000	0.000000	0.000000
27	0.000000	0.000000	0.000000	0.000000
28	0.064529	0.177202	0.000000	0.000000
29	0.089201	0.000000	0.000000	0.000000
30	0.000000	0.000000	0.000000	0.000000
31	0.000000	0.000000	0.000000	0.000000
32	0.094784	0.000000	0.116987	0.323682
33	0.000000	0.026204	0.000000	0.000000
34	0.000000	0.000000	0.000000	0.000000
35	0.000000	0.000000	0.093622	0.000000
36	0.021477	0.029489	0.000000	0.000000

User ID: Hamza Yerlikaya

Top 10 Words TF-IDF

	displayname	file	timer	files	java	new \
0	Hamza Yerlikaya	0.000000	0.000000	0.084239	0.000000	0.000000
1	Hamza Yerlikaya	0.000000	0.000000	0.000000	0.026395	0.000000
2	Hamza Yerlikaya	0.351468	0.000000	0.312480	0.000000	0.000000

3	Hamza Yerlikaya	0.000000	0.000000	0.000000	0.000000	0.000000
4	Hamza Yerlikaya	0.463315	0.000000	0.102980	0.000000	0.000000
5	Hamza Yerlikaya	0.000000	0.000000	0.000000	0.284629	0.000000
6	Hamza Yerlikaya	0.000000	0.782791	0.000000	0.066405	0.190127
7	Hamza Yerlikaya	0.064939	0.000000	0.000000	0.051849	0.371129
8	Hamza Yerlikaya	0.000000	0.000000	0.097851	0.000000	0.000000
9	Hamza Yerlikaya	0.000000	0.000000	0.000000	0.146846	0.000000

	table	application	jar	array	bits
0	0.000000	0.000000	0.000000	0.000000	0.000000
1	0.533399	0.000000	0.000000	0.000000	0.000000
2	0.000000	0.000000	0.000000	0.000000	0.000000
3	0.000000	0.000000	0.000000	0.000000	0.000000
4	0.000000	0.264787	0.264787	0.000000	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	0.251599	0.251599	0.000000	0.000000
9	0.000000	0.000000	0.000000	0.49458	0.49458

User ID: Joan Venge

Top 10 Words TF-IDF

	displayname	python	string	like	list	gt	class	\
0	Joan Venge	0.145502	0.437078	0.071480	0.000000	0.0	0.000000	
1	Joan Venge	0.060947	0.000000	0.059882	0.366161	0.0	0.000000	
2	Joan Venge	0.000000	0.000000	0.051889	0.000000	0.0	0.000000	
3	Joan Venge	0.086448	0.649210	0.084937	0.000000	0.0	0.000000	
4	Joan Venge	0.194943	0.000000	0.000000	0.000000	0.0	0.000000	
..	...	...	...	...	...	...	...	
61	Joan Venge	0.101717	0.000000	0.000000	0.000000	0.0	0.000000	
62	Joan Venge	0.000000	0.000000	0.092332	0.000000	0.0	0.000000	
63	Joan Venge	0.000000	0.000000	0.000000	0.000000	0.0	0.699688	
64	Joan Venge	0.255396	0.000000	0.125466	0.000000	0.0	0.000000	
65	Joan Venge	0.000000	0.000000	0.000000	0.000000	0.0	0.064150	

	want	index	value	function
0	0.000000	0.000000	0.000000	0.000000
1	0.165147	0.223014	0.000000	0.000000
2	0.071552	0.289870	0.000000	0.000000
3	0.000000	0.000000	0.000000	0.000000
4	0.000000	0.000000	0.000000	0.000000
..	...	...	...	...
61	0.000000	0.000000	0.167135	0.000000
62	0.063660	0.000000	0.000000	0.000000
63	0.096154	0.000000	0.000000	0.000000
64	0.000000	0.000000	0.000000	0.224828
65	0.000000	0.000000	0.192451	0.000000



[66 rows x 11 columns]

User ID: Ali

Top 10 Words TF-IDF

```
-----
displayname      gt      array javascript      php      using      lt  \
0      Ali  0.000000  0.000000      0.09527  0.000000  0.000000  0.000000
1      Ali  0.000000  0.000000      0.00000  0.000000  0.000000  0.000000
2      Ali  0.226733  0.000000      0.00000  0.000000  0.169265  0.244775
3      Ali  0.000000  0.382036      0.14211  0.000000  0.000000  0.000000
4      Ali  0.000000  0.000000      0.00000  0.083204  0.000000  0.000000
..      ...      ...      ...      ...      ...      ...      ...
74     Ali  0.000000  0.602876      0.00000  0.112129  0.000000  0.000000
75     Ali  0.000000  0.000000      0.00000  0.000000  0.031352  0.000000
76     Ali  0.332715  0.000000      0.00000  0.000000  0.000000  0.359189
77     Ali  0.000000  0.000000      0.00000  0.000000  0.000000  0.000000
78     Ali  0.000000  0.000000      0.00000  0.000000  0.070152  0.000000
```

```
      jquery      id  key  file
0  0.000000  0.000000  0.0  0.0
1  0.000000  0.000000  0.0  0.0
2  0.244775  0.000000  0.0  0.0
3  0.000000  0.000000  0.0  0.0
4  0.000000  0.000000  0.0  0.0
..      ...      ...  ...  ...
74  0.000000  0.000000  0.0  0.0
75  0.000000  0.317373  0.0  0.0
76  0.000000  0.000000  0.0  0.0
77  0.000000  0.167844  0.0  0.0
78  0.000000  0.000000  0.0  0.0
```

[79 rows x 11 columns]

User ID: TIMEX

Top 10 Words TF-IDF

```
-----
displayname      python      want      user      string      gt  dictionary  \
0      TIMEX  0.285104  0.000000  0.000000  0.000000  0.0  0.00000
1      TIMEX  0.134952  0.000000  0.000000  0.000000  0.0  0.00000
2      TIMEX  0.000000  0.000000  0.000000  0.000000  0.0  0.00000
3      TIMEX  0.000000  0.000000  0.000000  0.177879  0.0  0.00000
4      TIMEX  0.000000  0.000000  0.000000  0.000000  0.0  0.00000
..      ...      ...      ...      ...      ...      ...      ...
110     TIMEX  0.071606  0.000000  0.000000  0.000000  0.0  0.00000
111     TIMEX  0.053461  0.000000  0.000000  0.000000  0.0  0.07638
112     TIMEX  0.000000  0.032024  0.000000  0.000000  0.0  0.00000
113     TIMEX  0.000000  0.000000  0.132608  0.000000  0.0  0.00000
```

114	TIMEX	0.000000	0.000000	0.000000	0.000000	0.0	0.000000
-----	-------	----------	----------	----------	----------	-----	----------

	lt	use	javascript	return
0	0.000000	0.000000	0.000000	0.000000
1	0.000000	0.000000	0.000000	0.000000
2	0.124939	0.056344	0.000000	0.000000
3	0.000000	0.000000	0.000000	0.000000
4	0.000000	0.000000	0.000000	0.000000
..	...	...	...	...
110	0.000000	0.000000	0.099799	0.210055
111	0.000000	0.000000	0.000000	0.000000
112	0.000000	0.000000	0.000000	0.000000
113	0.063177	0.000000	0.000000	0.068069
114	0.113867	0.000000	0.000000	0.000000

[115 rows x 11 columns]

User ID: Matthew Rankin

Top 10 Words TF-IDF

	displayname	install	pip	python	lt	gt	branch \
0	Matthew Rankin	0.000000	0.000000	0.000000	0.000000	0.043561	0.571753
1	Matthew Rankin	0.000000	0.000000	0.000000	0.650278	0.544984	0.000000
2	Matthew Rankin	0.349045	0.319958	0.319958	0.000000	0.000000	0.000000
3	Matthew Rankin	0.235219	0.211697	0.094087	0.000000	0.000000	0.000000
4	Matthew Rankin	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
5	Matthew Rankin	0.427251	0.284834	0.000000	0.000000	0.000000	0.000000
6	Matthew Rankin	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
7	Matthew Rankin	0.000000	0.000000	0.338392	0.000000	0.000000	0.000000

  

	installed	flask	version	django
0	0.000000	0.000000	0.000000	0.000000
1	0.000000	0.000000	0.000000	0.000000
2	0.000000	0.000000	0.000000	0.000000
3	0.109034	0.000000	0.000000	0.054517
4	0.447027	0.533396	0.533396	0.000000
5	0.000000	0.000000	0.000000	0.000000
6	0.000000	0.000000	0.000000	0.425746
7	0.000000	0.000000	0.000000	0.000000

User ID: flybywire

Top 10 Words TF-IDF

	displayname	file	python	want	java	vs	gt \
0	flybywire	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
1	flybywire	0.000000	0.000000	0.112601	0.000000	0.000000	0.000000
2	flybywire	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
3	flybywire	0.000000	0.120377	0.000000	0.000000	0.000000	0.000000

4	flybywire	0.399837	0.000000	0.000000	0.000000	0.000000	0.300322
5	flybywire	0.166264	0.000000	0.155217	0.000000	0.000000	0.000000
6	flybywire	0.000000	0.000000	0.000000	0.657184	0.000000	0.000000
7	flybywire	0.099270	0.111844	0.000000	0.000000	0.000000	0.000000
8	flybywire	0.000000	0.461401	0.000000	0.000000	0.000000	0.000000
9	flybywire	0.000000	0.000000	0.123087	0.000000	0.000000	0.000000
10	flybywire	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
11	flybywire	0.227342	0.000000	0.106118	0.000000	0.000000	0.000000
12	flybywire	0.000000	0.146583	0.000000	0.000000	0.153493	0.000000
13	flybywire	0.076246	0.000000	0.284718	0.000000	0.000000	0.000000
14	flybywire	0.000000	0.112790	0.000000	0.000000	0.000000	0.225579
15	flybywire	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
16	flybywire	0.000000	0.000000	0.000000	0.544821	0.229172	0.000000
17	flybywire	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
18	flybywire	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
19	flybywire	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
20	flybywire	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
21	flybywire	0.000000	0.000000	0.086403	0.000000	0.000000	0.052138
22	flybywire	0.000000	0.000000	0.000000	0.000000	0.228055	0.000000
23	flybywire	0.071389	0.000000	0.066645	0.000000	0.000000	0.000000
24	flybywire	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
25	flybywire	0.000000	0.000000	0.116217	0.000000	0.000000	0.280514
26	flybywire	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
27	flybywire	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
28	flybywire	0.315679	0.000000	0.098234	0.000000	0.000000	0.000000
29	flybywire	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
30	flybywire	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
31	flybywire	0.000000	0.316739	0.000000	0.000000	0.000000	0.000000
32	flybywire	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
33	flybywire	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
34	flybywire	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
35	flybywire	0.000000	0.000000	0.068348	0.000000	0.000000	0.082486
36	flybywire	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
37	flybywire	0.175813	0.000000	0.082066	0.000000	0.000000	0.198083
38	flybywire	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
39	flybywire	0.000000	0.000000	0.000000	0.138204	0.000000	0.111033
40	flybywire	0.237010	0.000000	0.000000	0.000000	0.000000	0.000000
41	flybywire	0.000000	0.000000	0.000000	0.000000	0.194207	0.000000
42	flybywire	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
43	flybywire	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
44	flybywire	0.297907	0.000000	0.139056	0.000000	0.000000	0.000000
45	flybywire	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
46	flybywire	0.000000	0.321788	0.000000	0.000000	0.000000	0.000000
47	flybywire	0.000000	0.000000	0.000000	0.000000	0.220426	0.000000
48	flybywire	0.000000	0.000000	0.000000	0.000000	0.255143	0.000000

	use	bash	instance	standard
0	0.000000	0.000000	0.000000	0.000000

1	0.000000	0.000000	0.000000	0.000000
2	0.000000	0.000000	0.000000	0.000000
3	0.000000	0.000000	0.000000	0.000000
4	0.157240	0.165412	0.000000	0.498416
5	0.000000	0.412698	0.000000	0.466327
6	0.000000	0.000000	0.000000	0.000000
7	0.000000	0.000000	0.000000	0.000000
8	0.000000	0.000000	0.000000	0.000000
9	0.000000	0.000000	0.000000	0.000000
10	0.000000	0.000000	0.000000	0.000000
11	0.000000	0.000000	0.000000	0.000000
12	0.306986	0.000000	0.000000	0.000000
13	0.000000	0.000000	0.000000	0.000000
14	0.000000	0.000000	0.000000	0.000000
15	0.000000	0.000000	0.000000	0.000000
16	0.114586	0.000000	0.000000	0.000000
17	0.000000	0.000000	0.000000	0.000000
18	0.000000	0.000000	0.000000	0.000000
19	0.000000	0.000000	0.864911	0.000000
20	0.000000	0.000000	0.000000	0.000000
21	0.054596	0.172300	0.000000	0.000000
22	0.000000	0.000000	0.000000	0.000000
23	0.000000	0.177199	0.000000	0.000000
24	0.000000	0.000000	0.000000	0.000000
25	0.000000	0.000000	0.000000	0.000000
26	0.000000	0.000000	0.000000	0.000000
27	0.000000	0.000000	0.000000	0.000000
28	0.000000	0.261191	0.000000	0.000000
29	0.000000	0.000000	0.000000	0.000000
30	0.000000	0.000000	0.000000	0.000000
31	0.000000	0.000000	0.000000	0.000000
32	0.000000	0.000000	0.000000	0.000000
33	0.000000	0.000000	0.000000	0.000000
34	0.000000	0.000000	0.000000	0.000000
35	0.000000	0.000000	0.000000	0.000000
36	0.000000	0.000000	0.000000	0.000000
37	0.000000	0.000000	0.000000	0.000000
38	0.000000	0.000000	0.000000	0.000000
39	0.000000	0.000000	0.000000	0.000000
40	0.000000	0.000000	0.000000	0.000000
41	0.388413	0.000000	0.000000	0.000000
42	0.000000	0.000000	0.000000	0.207214
43	0.000000	0.000000	0.000000	0.000000
44	0.175732	0.000000	0.000000	0.000000
45	0.000000	0.000000	0.317777	0.000000
46	0.000000	0.000000	0.000000	0.000000
47	0.000000	0.000000	0.000000	0.000000
48	0.000000	0.000000	0.000000	0.000000

