In [11]:

```
import matplotlib.pyplot as plt
import matplotlib.lines as mlines
import seaborn as sns
import numpy as np
import pandas as pd
import os

import warnings

import pickle
import time

import re
from bs4 import BeautifulSoup
import nltk
from nltk.tokenize import ToktokTokenizer
from nltk.stem.wordnet import WordNetLemmatizer
from nltk.corpus import stopwords
from string import punctuation

from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.decomposition import LatentDirichletAllocation
from sklearn.preprocessing import MultiLabelBinarizer
from sklearn.model_selection import train_test_split
from sklearn.model_selection import learning_curve
from sklearn.model_selection import ShuffleSplit
from sklearn.dummy import DummyClassifier
from sklearn.naive_bayes import MultinomialNB
from sklearn.linear_model import SGDClassifier
from sklearn.linear_model import LogisticRegression
from sklearn.multiclass import OneVsRestClassifier
from sklearn.svm import LinearSVC
from sklearn.linear_model import Perceptron
from sklearn.linear_model import PassiveAggressiveClassifier
from sklearn.neural_network import MLPClassifier
from sklearn.ensemble import RandomForestClassifier
from sklearn import model_selection
from sklearn.metrics import make_scorer
from sklearn.metrics import confusion_matrix
from sklearn.metrics import hamming_loss
from sklearn.cluster import KMeans


import logging

from scipy.sparse import hstack
warnings.filterwarnings("ignore")
plt.style.use('bmh')
%matplotlib inline
```

In [12]:

```
np.random.seed(seed=11)
```

In [13]:

```
#print(os.listdir("../Questions.csv"))
df = pd.read_csv("Questions.csv", encoding="ISO-8859-1")
```

In [14]:

```
df.head(5)
```

Out[14]:

	Id	OwnerUserId	CreationDate	ClosedDate	Score	Title	Body
0	80	26.0	2008-08-01T13:57:07Z	NaN	26	SQLStatement.execute() - multiple queries in o...	<p>I've written a database generation script i...
1	90	58.0	2008-08-01T14:41:24Z	2012-12-26T03:45:49Z	144	Good branching and merging tutorials for Torto...	<p>Are there any really good tutorials explain...
2	120	83.0	2008-08-01T15:50:08Z	NaN	21	ASP.NET Site Maps	<p>Has anyone got experience creating <strong>...
3	180	2089740.0	2008-08-01T18:42:19Z	NaN	53	Function for creating color wheels	<p>This is something I've pseudo-solved many t...
4	260	91.0	2008-08-01T23:22:08Z	NaN	49	Adding scripting functionality to .NET applica...	<p>I have a little game written in C#. It uses...

In [15]:

```
tags = pd.read_csv("Tags.csv", encoding="ISO-8859-1", dtype={'Tag': str})
tags.head(5)
```

Out[15]:

	Id	Tag
0	80	flex
1	80	actionscript-3
2	80	air
3	90	svn
4	90	tortoisesvn

In [16]:

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1264216 entries, 0 to 1264215
Data columns (total 7 columns):
 #   Column        Non-Null Count    Dtype  
---  ------        --------------    -----  
 0   Id            1264216 non-null  int64  
 1   OwnerUserId   1249762 non-null  float64
 2   CreationDate  1264216 non-null  object 
 3   ClosedDate    55959 non-null    object 
 4   Score         1264216 non-null  int64  
 5   Title         1264216 non-null  object 
 6   Body          1264216 non-null  object 
dtypes: float64(1), int64(2), object(4)
memory usage: 67.5+ MB
```

In [17]:

```
tags.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 3750994 entries, 0 to 3750993
Data columns (total 2 columns):
 #   Column  Dtype 
---  ------  ----- 
 0   Id      int64 
 1   Tag     object
dtypes: int64(1), object(1)
memory usage: 57.2+ MB
```

In [18]:

```
tags['Tag'] = tags['Tag'].astype(str)
```

In [19]:

```
grouped_tags = tags.groupby("Id")['Tag'].apply(lambda tags: ' '.join(tags))
```

In [23]:

```
grouped_tags.head(5)
```

Out[23]:

```
Id
80                             flex actionscript-3 air
90         svn tortoisesvn branch branching-and-merging
120                                sql asp.net sitemap
180    algorithm language-agnostic colors color-space
260           c# .net scripting compiler-construction
Name: Tag, dtype: object
```

In [24]:

```
grouped_tags.reset_index()
```

Out[24]:

	Id	Tag
0	80	flex actionscript-3 air
1	90	svn tortoisesvn branch branching-and-merging
2	120	sql asp.net sitemap
3	180	algorithm language-agnostic colors color-space
4	260	c# .net scripting compiler-construction
...	...	...
1264211	40143210	php .htaccess
1264212	40143300	google-bigquery
1264213	40143340	android android-studio
1264214	40143360	javascript vue.js
1264215	40143380	npm mocha babel

1264216 rows × 2 columns

In [25]:

```
grouped_tags_final = pd.DataFrame({'Id':grouped_tags.index, 'Tags':grouped_tags.values
})
```

In [26]:

```
grouped_tags_final.head(5)
```

Out[26]:

	Id	Tags
0	80	flex actionscript-3 air
1	90	svn tortoisesvn branch branching-and-merging
2	120	sql asp.net sitemap
3	180	algorithm language-agnostic colors color-space
4	260	c# .net scripting compiler-construction

In [27]:

```
df.drop(columns=['OwnerUserId', 'CreationDate', 'ClosedDate'], inplace=True)
```

In [28]:

```
df = df.merge(grouped_tags_final, on='Id')
```

In [ ]:

```
df.head(5)
```

In [ ]:

```
new_df = df[df['Score']>5]
```

In [ ]:

```
plt.figure(figsize=(5, 5))
new_df.isnull().mean(axis=0).plot.barh()
plt.title("Ratio of missing values per columns")
```

In [ ]:

```
print('Dupplicate entries: {}'.format(new_df.duplicated().sum()))
new_df.drop_duplicates(inplace = True)
```

In [ ]:

```
new_df.drop(columns=['Id', 'Score'], inplace=True)
```

In [29]:

```
new_df.head(5)
```

Out[29]:

	Title	Body	Tags
0	SQLStatement.execute() - multiple queries in o...	<p>I've written a database generation script i...	flex actionscript-3 air
1	Good branching and merging tutorials for Torto...	<p>Are there any really good tutorials explain...	svn tortoisesvn branch branching-and-merging
2	ASP.NET Site Maps	<p>Has anyone got experience creating <strong>...	sql asp.net sitemap
3	Function for creating color wheels	<p>This is something I've pseudo-solved many t...	algorithm language-agnostic colors color-space
4	Adding scripting functionality to .NET applica...	<p>I have a little game written in C#. It uses...	c# .net scripting compiler-construction

In [ ]:

```
new_df['Tags'] = new_df['Tags'].apply(lambda x: x.split())
```

In [ ]:

```
all_tags = [item for sublist in new_df['Tags'].values for item in sublist]
```

In [ ]:

```
len(all_tags)
```

In [ ]:

```
my_set = set(all_tags)
unique_tags = list(my_set)
len(unique_tags)
```

In [ ]:

```
flat_list = [item for sublist in new_df['Tags'].values for item in sublist]

keywords = nltk.FreqDist(flat_list)

keywords = nltk.FreqDist(keywords)

frequencies_words = keywords.most_common(100)
tags_features = [word[0] for word in frequencies_words]
```

In [ ]:

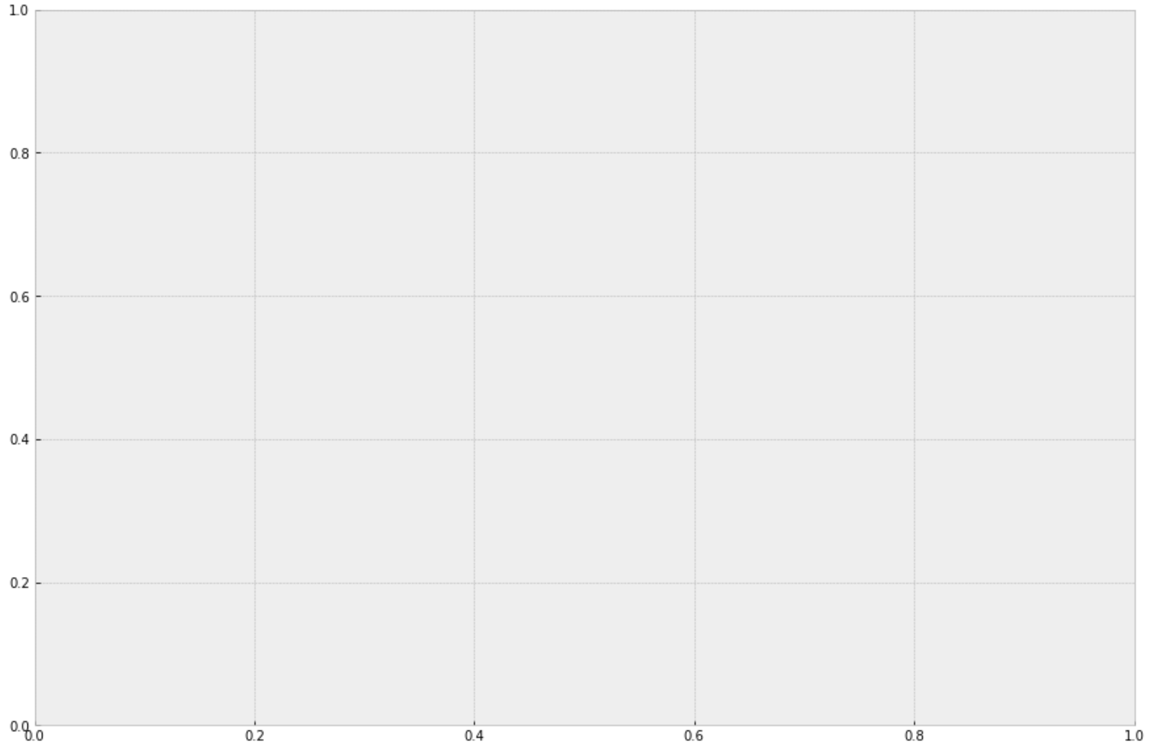
```
tags_features
```

In [5]:

```
fig, ax = plt.subplots(figsize=(15, 10))
keywords.plot(100, cumulative=False)
```

```
--------------------------------------------------------------------------
-
NameError                                 Traceback (most recent call las
t)
<ipython-input-5-a594a37c0360> in <module>
      1 fig, ax = plt.subplots(figsize=(15, 10))
----> 2 keywords.plot(100, cumulative=False)

NameError: name 'keywords' is not defined
```



In [ ]:

```
def most_common(tags):
    tags_filtered = []
    for i in range(0, len(tags)):
        if tags[i] in tags_features:
            tags_filtered.append(tags[i])
    return tags_filtered
```

In [ ]:

```
new_df['Tags'] = new_df['Tags'].apply(lambda x: most_common(x))
new_df['Tags'] = new_df['Tags'].apply(lambda x: x if len(x)>0 else None)
```

In [ ]:

```
new_df.shape
```

In [ ]:

```
new_df.dropna(subset=['Tags'], inplace=True)
```

In [ ]:

```
new_df.shape
```

In [ ]:

```
new_df['Body'] = new_df['Body'].apply(lambda x: BeautifulSoup(x).get_text())
```

In [12]:

```
def clean_text(text):
    text = text.lower()
    text = re.sub(r"what's", "what is ", text)
    text = re.sub(r"\'s", " ", text)
    text = re.sub(r"\'ve", " have ", text)
    text = re.sub(r"can't", "can not ", text)
    text = re.sub(r"n't", " not ", text)
    text = re.sub(r"i'm", "i am ", text)
    text = re.sub(r"\'re", " are ", text)
    text = re.sub(r"\'d", " would ", text)
    text = re.sub(r"\'ll", " will ", text)
    text = re.sub(r"\'scuse", " excuse ", text)
    text = re.sub(r"\'\n", " ", text)
    text = re.sub(r"\'\xa0", " ", text)
    text = re.sub('\s+', ' ', text)
    text = text.strip(' ')
    return text
```

In [ ]:

```
new_df['Body'] = new_df['Body'].apply(lambda x: clean_text(x))
```

In [45]:

```
token=ToktokTokenizer()
```

In [46]:

```
punct = '!"#$%&\'()*+,./:;<=>?@[\\]^_`{|}~'
```

In [47]:

```
def strip_list_noempty(mylist):
    newlist = (item.strip() if hasattr(item, 'strip') else item for item in mylist)
    return [item for item in newlist if item != '']
```

In [48]:

```
def clean_punct(text):
    words=token.tokenize(text)
    punctuation_filtered = []
    regex = re.compile('[%s]' % re.escape(punct))
    remove_punctuation = str.maketrans(' ', ' ', punct)
    for w in words:
        if w in tags_features:
            punctuation_filtered.append(w)
        else:
            punctuation_filtered.append(regex.sub('', w))

    filtered_list = strip_list_noempty(punctuation_filtered)

    return ' '.join(map(str, filtered_list))
```

In [49]:

```
new_df['Body'] = new_df['Body'].apply(lambda x: clean_punct(x))
```

In [ ]:

```
new_df['Body'][2]
```

In [77]:

```
lemma=WordNetLemmatizer()
nltk.download('stopwords')
stop_words = set(stopwords.words("english"))
```

```
[nltk_data] Downloading package stopwords to
[nltk_data]     C:\Users\ramyalavanya\AppData\Roaming\nltk_data...
[nltk_data]   Unzipping corpora\stopwords.zip.
```

In [78]:

```
def lemitizeWords(text):
    words=token.tokenize(text)
    listLemma=[]
    for w in words:
        x=lemma.lemmatize(w, pos="v")
        listLemma.append(x)
    return ' '.join(map(str, listLemma))

def stopWordsRemove(text):

    stop_words = set(stopwords.words("english"))

    words=token.tokenize(text)

    filtered = [w for w in words if not w in stop_words]

    return ' '.join(map(str, filtered))
```

In [80]:

```
nltk.download('wordnet')
new_df['Body'] = new_df['Body'].apply(lambda x: lemitizeWords(x))
new_df['Body'] = new_df['Body'].apply(lambda x: stopWordsRemove(x))
```

```
[nltk_data] Downloading package wordnet to
[nltk_data]     C:\Users\ramyalavanya\AppData\Roaming\nltk_data...
[nltk_data]   Unzipping corpora\wordnet.zip.
```

In [81]:

```
new_df['Title'] = new_df['Title'].apply(lambda x: str(x))
new_df['Title'] = new_df['Title'].apply(lambda x: clean_text(x))
new_df['Title'] = new_df['Title'].apply(lambda x: clean_punct(x))
new_df['Title'] = new_df['Title'].apply(lambda x: lemitizeWords(x))
new_df['Title'] = new_df['Title'].apply(lambda x: stopWordsRemove(x))
```

In [82]:

```
no_topics = 20
```

In [83]:

```
text = new_df['Body']
```

In [84]:

```
vectorizer_train = TfidfVectorizer(analyzer = 'word',
                                       min_df=0.0,
                                       max_df = 1.0,
                                       strip_accents = None,
                                       encoding = 'utf-8',
                                       preprocessor=None,
                                       token_pattern=r"(?u)\S\S+", # Need to repeat tok
en pattern
                                       max_features=1000)
```

In [85]:

```
TF_IDF_matrix = vectorizer_train.fit_transform(text)
```

In [87]:

```
lda = LatentDirichletAllocation(n_components=no_topics, max_iter=5, learning_method='online', learning_offset=50,random_state=11).fit(TF_IDF_matrix)
```

In [88]:

```
def display_topics(model, feature_names, no_top_words):
    for topic_idx, topic in enumerate(model.components_):
        print("--------------------------------------------")
        print("Topic %d:" % (topic_idx))
        print(" ".join([feature_names[i]
                        for i in topic.argsort()[:-no_top_words - 1:-1]]))
        print("--------------------------------------------")


no_top_words = 10
display_topics(lda, vectorizer_train.get_feature_names(), no_top_words)
```

```
--------------------------------------------
Topic 0:
div function var html script class page jquery id input
--------------------------------------------
--------------------------------------------
Topic 1:
use app ios would find xcode like know android application
--------------------------------------------
--------------------------------------------
Topic 2:
00 scala font 09 01 02 tab utf-8 socket constant
--------------------------------------------
--------------------------------------------
Topic 3:
array list number data value would use like way function
--------------------------------------------
--------------------------------------------
Topic 4:
android width color height div background scroll style image top
--------------------------------------------
--------------------------------------------
Topic 5:
import python module modules def random filter statement docs 22
--------------------------------------------
--------------------------------------------
Topic 6:
button view android activity app event click menu item screen
--------------------------------------------
--------------------------------------------
Topic 7:
date datetime format 17 echo 00 time 15 convert year
--------------------------------------------
--------------------------------------------
Topic 8:
int public class return string void std type new foo
--------------------------------------------
--------------------------------------------
Topic 9:
file project error run use build command try install version
--------------------------------------------
--------------------------------------------
Topic 10:
table select query row column sql database id key value
--------------------------------------------
--------------------------------------------
Topic 11:
image json td tr cell label selector picture table img
--------------------------------------------
--------------------------------------------
Topic 12:
request server use error user file get url send connection
--------------------------------------------
--------------------------------------------
Topic 13:
rail gem ruby 12 11 14 pdf 13 10 18
--------------------------------------------
--------------------------------------------
Topic 14:
124 string character line match node text file regex print
--------------------------------------------
--------------------------------------------
```

```
Topic 15:
git branch commit push repository merge github master remote pull
--------------------------------------------
--------------------------------------------
Topic 16:
controller class model view use name public property get object
--------------------------------------------
--------------------------------------------
Topic 17:
thread exception catch queue throw lock bitmap child wait exit
--------------------------------------------
--------------------------------------------
Topic 18:
nsstring nil animate animation self dictionary cursor yes hide iphone
--------------------------------------------
--------------------------------------------
Topic 19:
use would function memory code like question test time one
--------------------------------------------
```

In [89]:

```
X1 = new_df['Body']
X2 = new_df['Title']
y = new_df['Tags']
```

In [90]:

```
multilabel_binarizer = MultiLabelBinarizer()
y_bin = multilabel_binarizer.fit_transform(y)
```

In [91]:

```
vectorizer_X1 = TfidfVectorizer(analyzer = 'word',
                                       min_df=0.0,
                                       max_df = 1.0,
                                       strip_accents = None,
                                       encoding = 'utf-8',
                                       preprocessor=None,
                                       token_pattern=r"(?u)\S\S+",
                                       max_features=1000)

vectorizer_X2 = TfidfVectorizer(analyzer = 'word',
                                       min_df=0.0,
                                       max_df = 1.0,
                                       strip_accents = None,
                                       encoding = 'utf-8',  preprocessor=None,
                                       token_pattern=r"(?u)\S\S+",
                                       max_features=1000)
```

In [92]:

```
X1_tfidf = vectorizer_X1.fit_transform(X1)
X2_tfidf = vectorizer_X2.fit_transform(X2)
```

In [93]:

```
X_tfidf = hstack([X1_tfidf,X2_tfidf])
```

In [94]:

```
X_train, X_test, y_train, y_test = train_test_split(X_tfidf, y_bin, test_size = 0.2, random_state = 0) # Do 80/20 split
```

In [95]:

```
param_grid = {'estimator__C':[1,10,100,1000]
             }
```

In [103]:

```
svc = OneVsRestClassifier(LinearSVC())
CV_svc = model_selection.GridSearchCV(estimator=svc, param_grid=param_grid, cv= 5, verb
ose=10)
CV_svc.fit(X_train, y_train)
```

```
Fitting 5 folds for each of 4 candidates, totalling 20 fits
[CV] estimator__C=1 ..................................................

[Parallel(n_jobs=1)]: Using backend SequentialBackend with 1 concurrent wo
rkers.

[CV] ...................... estimator__C=1, score=0.367, total=  15.4s
[CV] estimator__C=1 ..................................................

[Parallel(n_jobs=1)]: Done   1 out of   1 | elapsed:   15.4s remaining:
0.0s

[CV] ...................... estimator__C=1, score=0.360, total=  14.8s
[CV] estimator__C=1 ..................................................

[Parallel(n_jobs=1)]: Done   2 out of   2 | elapsed:   30.2s remaining:
0.0s

[CV] ...................... estimator__C=1, score=0.360, total=  15.7s
[CV] estimator__C=1 ..................................................

[Parallel(n_jobs=1)]: Done   3 out of   3 | elapsed:   46.0s remaining:
0.0s

[CV] ...................... estimator__C=1, score=0.354, total=  15.7s
[CV] estimator__C=1 ..................................................

[Parallel(n_jobs=1)]: Done   4 out of   4 | elapsed:  1.0min remaining:
0.0s

[CV] ...................... estimator__C=1, score=0.359, total=  16.1s
[CV] estimator__C=10 .................................................

[Parallel(n_jobs=1)]: Done   5 out of   5 | elapsed:  1.3min remaining:
0.0s

[CV] ..................... estimator__C=10, score=0.296, total=  45.5s
[CV] estimator__C=10 .................................................

[Parallel(n_jobs=1)]: Done   6 out of   6 | elapsed:  2.1min remaining:
0.0s

[CV] ..................... estimator__C=10, score=0.291, total=  43.7s
[CV] estimator__C=10 .................................................

[Parallel(n_jobs=1)]: Done   7 out of   7 | elapsed:  2.8min remaining:
0.0s


[CV] ..................... estimator__C=10, score=0.296, total=  43.3s
[CV] estimator__C=10 .................................................

[Parallel(n_jobs=1)]: Done   8 out of   8 | elapsed:  3.5min remaining:
0.0s

[CV] ..................... estimator__C=10, score=0.283, total=  45.1s
[CV] estimator__C=10 .................................................

[Parallel(n_jobs=1)]: Done   9 out of   9 | elapsed:  4.3min remaining:
0.0s
```

```
[CV] ..................... estimator__C=10, score=0.288, total=  46.0s
[CV] estimator__C=100 ................................................
[CV] .................... estimator__C=100, score=0.242, total= 1.1min
[CV] estimator__C=100 ................................................
[CV] .................... estimator__C=100, score=0.238, total= 1.1min
[CV] estimator__C=100 ................................................
[CV] .................... estimator__C=100, score=0.241, total= 1.1min
[CV] estimator__C=100 ................................................
[CV] .................... estimator__C=100, score=0.231, total= 1.1min
[CV] estimator__C=100 ................................................
[CV] .................... estimator__C=100, score=0.242, total= 1.1min
[CV] estimator__C=1000 ...............................................
[CV] ................... estimator__C=1000, score=0.196, total= 1.2min
[CV] estimator__C=1000 ...............................................
[CV] ................... estimator__C=1000, score=0.215, total= 1.3min
[CV] estimator__C=1000 ...............................................
[CV] ................... estimator__C=1000, score=0.209, total= 1.3min
[CV] estimator__C=1000 ...............................................
[CV] ................... estimator__C=1000, score=0.208, total= 1.3min
[CV] estimator__C=1000 ...............................................
[CV] ................... estimator__C=1000, score=0.213, total= 1.4min
```

```
[Parallel(n_jobs=1)]: Done  20 out of  20 | elapsed: 16.9min finished
```

Out[103]:

```
GridSearchCV(cv=5, estimator=OneVsRestClassifier(estimator=LinearSVC()),
             param_grid={'estimator__C': [1, 10, 100, 1000]}, verbose=10)
```

In [104]:

```
CV_svc.best_params_
```

Out[104]:

```
{'estimator__C': 1}
```

In [105]:

```
best_model = CV_svc.best_estimator_
```

In [107]:

```
y_pred = best_model.predict(X_test)

#print_score(y_pred, best_model)
```

In [108]:

```
for i in range(y_train.shape[1]):
    print(multilabel_binarizer.classes_[i])
    print(confusion_matrix(y_test[:,i], y_pred[:,i]))
    print("")
```

```
.net
[[12033    54]
 [  447   100]]

ajax
[[12526    15]
 [   55    38]]

algorithm
[[12445    21]
 [   87    81]]

android
[[11528    40]
 [  194   872]]

angularjs
[[12443     7]
 [   49   135]]

api
[[12567    14]
 [   45     8]]

arrays
[[12405    55]
 [  103    71]]

asp.net
[[12346    30]
 [  166    92]]

asp.net-mvc
[[12402    41]
 [   99    92]]

asp.net-mvc-3
[[12566    11]
 [   44    13]]

bash
[[12520     7]
 [   46    61]]

c
[[12220    58]
 [  203   153]]

c#
[[11060   175]
 [  618   781]]

c++
[[11674    91]
 [  260   609]]

c++11
[[12453    31]
 [  103    47]]

class
```

```
[[12574    12]
 [   45     3]]

cocoa
[[12553     4]
 [   69     8]]

cocoa-touch
[[12530    10]
 [   90     4]]

css
[[12186    74]
 [  123   251]]

css3
[[12561    13]
 [   45    15]]

database
[[12503    29]
 [   84    18]]

datetime
[[12571    17]
 [   36    10]]

debugging
[[12523    27]
 [   50    34]]

delphi
[[12542     2]
 [   39    51]]

design-patterns
[[12573    11]
 [   36    14]]

django
[[12465     7]
 [   31   131]]

eclipse
[[12472     9]
 [   55    98]]

emacs
[[12591     1]
 [   16    26]]

entity-framework
[[12547    20]
 [   36    31]]

exception
[[12551    17]
 [   43    23]]

facebook
[[12571     9]
```

```
 [   17    37]]

function
[[12576    11]
 [   44     3]]

gcc
[[12566    12]
 [   35    21]]

generics
[[12558     9]
 [   31    36]]

git
[[12425     8]
 [   17   184]]

github
[[12575     5]
 [   23    31]]

google-chrome
[[12559     8]
 [   33    34]]

haskell
[[12497     5]
 [   70    62]]

hibernate
[[12570     5]
 [   16    43]]

html
[[12109   102]
 [  261   162]]

html5
[[12506     7]
 [   77    44]]

http
[[12563    14]
 [   41    16]]

image
[[12561    15]
 [   50     8]]

ios
[[12001    93]
 [  283   257]]

ipad
[[12573     2]
 [   42    17]]

iphone
[[12227    49]
 [  246   112]]
```

```
java
[[11130   144]
 [  565   795]]

javascript
[[11338   168]
 [  511   617]]

jquery
[[12011    65]
 [  199   359]]

json
[[12474    24]
 [   65    71]]

linq
[[12536    14]
 [   37    47]]

linux
[[12420    27]
 [  121    66]]

list
[[12540    19]
 [   61    14]]

math
[[12577    10]
 [   46     1]]

maven
[[12563    15]
 [   15    41]]

mongodb
[[12569     1]
 [   33    31]]

multithreading
[[12468    25]
 [   85    56]]

mysql
[[12363    32]
 [   74   165]]

node.js
[[12446    13]
 [   71   104]]

numpy
[[12579     4]
 [   26    25]]

objective-c
[[12199    81]
 [  249   105]]
```

```
oop
[[12573     5]
 [   56     0]]

optimization
[[12584     4]
 [   46     0]]

oracle
[[12576     3]
 [   25    30]]

osx
[[12493    15]
 [   88    38]]

performance
[[12437    29]
 [  133    35]]

perl
[[12563     0]
 [   27    44]]

php
[[12001    38]
 [  218   377]]

postgresql
[[12572     0]
 [   29    33]]

python
[[11715    51]
 [  240   628]]

qt
[[12585     2]
 [   24    23]]

r
[[12384    31]
 [   97   122]]

regex
[[12499    10]
 [   41    84]]

rest
[[12565     9]
 [   30    30]]

ruby
[[12311    36]
 [  158   129]]

ruby-on-rails
[[12304    41]
 [  105   184]]

ruby-on-rails-3
```

```
[[12526    21]
 [   75    12]]

scala
[[12488     7]
 [   36   103]]

security
[[12564     9]
 [   44    17]]

shell
[[12554    13]
 [   43    24]]

spring
[[12516    15]
 [   35    68]]

sql
[[12269    70]
 [  158   137]]

sql-server
[[12390    50]
 [   98    96]]

string
[[12432    54]
 [   93    55]]

svn
[[12586     6]
 [   12    30]]

swift
[[12521     5]
 [   35    73]]

swing
[[12582     2]
 [   28    22]]

templates
[[12542    13]
 [   44    35]]

tsql
[[12562    16]
 [   52     4]]

twitter-bootstrap
[[12565     2]
 [   21    46]]

unit-testing
[[12478    28]
 [   70    58]]

vim
[[12571     2]
```

```
 [   15    46]]

visual-studio
[[12497    29]
 [   72    36]]

visual-studio-2010
[[12542    17]
 [   52    23]]

wcf
[[12583     0]
 [   12    39]]

windows
[[12454    30]
 [  112    38]]

winforms
[[12556    14]
 [   42    22]]

wpf
[[12463     8]
 [   55   108]]

xcode
[[12426    26]
 [  102    80]]

xml
[[12507    17]
 [   58    52]]
```

In [109]:

```
def print_top10(feature_names, clf, class_labels):
    """Prints features with the highest coefficient values, per class"""
    for i, class_label in enumerate(class_labels):
        top10 = np.argsort(clf.coef_[i])[-10:]
        print("---------------------------------------------")
        print("%s: %s" % (class_label, " ".join(feature_names[j] for j in top10)))
        print("--------------------------------------------")
```

In [110]:

```
feature_names = vectorizer_X1.get_feature_names() + vectorizer_X2.get_feature_names()
```

In [111]:

```
print_top10(feature_names, best_model, multilabel_binarizer.classes_)
```

```
--------------------------------------------
.net: determine finally sender windows mouse assembly microsoft consolewri
teline .net .net
--------------------------------------------
--------------------------------------------
ajax: much 10 img request web items security load ajax ajax
--------------------------------------------
--------------------------------------------
algorithm: log input additional stick give although diff problem algorithm
algorithm
--------------------------------------------
--------------------------------------------
android: webview eandroidruntime linearlayout fragment edittext intent tex
tview activity android android
--------------------------------------------
--------------------------------------------
angularjs: resource factory integrate token config directive controller sc
ope angular angularjs
--------------------------------------------
--------------------------------------------
api: virtual border interface retrieve deal webview docs dependency api ap
i
--------------------------------------------
--------------------------------------------
arrays: particular chart autocomplete rest success username step come arra
y array
--------------------------------------------
--------------------------------------------
asp.net: external port entire trace website control server iis asp.net as
p.net
--------------------------------------------
--------------------------------------------
asp.net-mvc: try ie inherit provider action platform view razor mvc mvc
--------------------------------------------
--------------------------------------------
asp.net-mvc-3: us instead prefer description successfully side download vi
ew razor mvc3
--------------------------------------------
--------------------------------------------
bash: short local home 21 22 variables syntax script echo bash
--------------------------------------------
--------------------------------------------
c: relevant sizeof define complex char struct embed cc gcc printf
--------------------------------------------
--------------------------------------------
c#: datetime remain sender winforms linq typeof foreach c# consolewritelin
e c#
--------------------------------------------
--------------------------------------------
c++: c++11 std opencv boost cout cc stl std c++ c++
--------------------------------------------
--------------------------------------------
c++11: obviously initialize tuple standard via std template move auto c++1
1
--------------------------------------------
--------------------------------------------
class: thing tuple public whenever outside next compiler quite cout class
--------------------------------------------
--------------------------------------------
cocoa: window last notification launch super 21 easily nslog mac cocoa
--------------------------------------------
```

```
--------------------------------------------
cocoa-touch: cocoa amount upgrade animation management programatically sev
eral validate visible longer
--------------------------------------------
--------------------------------------------
css: trace stick override couple web area less fix css css
--------------------------------------------
--------------------------------------------
css3: initialize ways must transform 50 unknown title initialization selec
tor css3
--------------------------------------------
--------------------------------------------
database: cross ms configuration special iterate db resolve progress datab
ase database
--------------------------------------------
--------------------------------------------
datetime: reason 09 provider ex behavior wonder prefer original datetime d
atetime
--------------------------------------------
--------------------------------------------
debugging: finish value print suppose drive bug attach debug debug debugge
r
--------------------------------------------
--------------------------------------------
delphi: sample unit old 2010 finally component end begin procedure delphi
--------------------------------------------
--------------------------------------------
design-patterns: think developer specifically script useful compilation de
sign clean pattern pattern
--------------------------------------------
--------------------------------------------
django: instal iterate temporary play textjavascript place admin model dja
ngo django
--------------------------------------------
--------------------------------------------
eclipse: model properties shortcut others problem explorer pull software e
clipse eclipse
--------------------------------------------
--------------------------------------------
emacs: search enable recently behaviour configure 24 mode nil buffer emacs
--------------------------------------------
--------------------------------------------
entity-framework: context property apply framework db first entity framewo
rk entity ef
--------------------------------------------
--------------------------------------------
exception: instead wo validate exception insert kind exceptions sense exce
ption exceptions
--------------------------------------------
--------------------------------------------
facebook: class confuse post sdk callback ask effect customize facebook fa
cebook
--------------------------------------------
--------------------------------------------
function: twice bool 24 function length datetime enough sign function defi
nition
--------------------------------------------
--------------------------------------------
gcc: native enable target tutorial margin 32 information declare gcc gcc
--------------------------------------------
--------------------------------------------
```

```
generics: setup typeof progress others recently extend via generic generic
generics
--------------------------------------------
--------------------------------------------
git: diff repo local commit commit master history branch git git
--------------------------------------------
--------------------------------------------
github: live tag bundle whole unexpected access push upgrade github github
--------------------------------------------
--------------------------------------------
google-chrome: href absolute instead history problems garbage 404 bug chro
me chrome
--------------------------------------------
--------------------------------------------
haskell: bool perhaps derive type xs language append maybe instance haskel
l
--------------------------------------------
--------------------------------------------
hibernate: every retrieve generate nothing entity jpa fetch session transa
ction hibernate
--------------------------------------------
--------------------------------------------
html: overflow finally td along fix elements onclick div html html
--------------------------------------------
--------------------------------------------
html5: progress maximum proper overflow browser css3 main canvas audio htm
l5
--------------------------------------------
--------------------------------------------
http: enable client param js lose functionality request large http http
--------------------------------------------
--------------------------------------------
image: onclick around 2d confuse activity image yes img image picture
--------------------------------------------
--------------------------------------------
ios: nslog delegate ui due app alloc swift apple ios ios
--------------------------------------------
--------------------------------------------
ipad: finish visible orientation exit ways else detail less hard ipad
--------------------------------------------
--------------------------------------------
iphone: uitableview rather resolution push core 40 release delegate iphone
iphone
--------------------------------------------
--------------------------------------------
java: final jar extend tomcat servlet jvm swing systemoutprintln java java
--------------------------------------------
--------------------------------------------
javascript: chart js chrome onclick undefined consolelog var js javascript
javascript
--------------------------------------------
--------------------------------------------
jquery: complete global protect click clone plugin ajax ready jquery jquer
y
--------------------------------------------
--------------------------------------------
json: serialize description callback retrieve absolute anything field orig
inal json json
--------------------------------------------
--------------------------------------------
linq: inherit collection source expression member cod solve var well linq
```

```
--------------------------------------------
--------------------------------------------
linux: setup fill edit process child cursor state ps linux linux
--------------------------------------------
--------------------------------------------
list: several horizontal faster -1 rather several clear big list list
--------------------------------------------
--------------------------------------------
math: computer hold unique hit log calculate logic fast assume evaluate
--------------------------------------------
--------------------------------------------
maven: info easy increase yes dependency modules advance execution maven m
aven
--------------------------------------------
--------------------------------------------
mongodb: group thus shell err driver db cursor document collection mongodb
--------------------------------------------
--------------------------------------------
multithreading: completely suggestions 100 solve prevent complete separate
thread lock thread
--------------------------------------------
--------------------------------------------
mysql: us users 24 limit database root per track mysql mysql
--------------------------------------------
--------------------------------------------
node.js: virtual schema â�� require node express node err nodejs node.js
--------------------------------------------
--------------------------------------------
numpy: amazon instead ignore import pure matrix array fix far numpy
--------------------------------------------
--------------------------------------------
objective-c: exceptions ios selector textjavascript super cocoa nsstring n
slog objective objective-c
--------------------------------------------
--------------------------------------------
oop: directly object fastest finish couple class obvious fit refer design
--------------------------------------------
--------------------------------------------
optimization: site 05 pad bitmap compiler faster couple root optimization
optimize
--------------------------------------------
--------------------------------------------
oracle: management schema number feature end developer max replace begin o
racle
--------------------------------------------
--------------------------------------------
osx: cocoa notification setup terminal os apple os mac mac osx
--------------------------------------------
--------------------------------------------
performance: efficient important faster entire ways speed faster performan
ce slow performance
--------------------------------------------
--------------------------------------------
perl: less upgrade trace print script warn job foreach fetch perl
--------------------------------------------
--------------------------------------------
php: modules construct foreach protect echo array codeigniter laravel php
php
--------------------------------------------
--------------------------------------------
postgresql: people much heroku several sqlalchemy docs database sure postg
```

```
res postgresql
--------------------------------------------
--------------------------------------------
python: import pip flask def sqlalchemy matplotlib pandas numpy python pyt
hon
--------------------------------------------
--------------------------------------------
qt: effect could 30 unsigned connect parent picture actually develop qt
--------------------------------------------
--------------------------------------------
r: variables frame datatable vector function library package na plot ggplo
t2
--------------------------------------------
--------------------------------------------
regex: either 17 operation replace several regular match regular regex reg
ex
--------------------------------------------
--------------------------------------------
rest: multiple anything entity finish service sense produce rest restful r
est
--------------------------------------------
--------------------------------------------
ruby: inject activerecord unfortunately longer put bundle gem end ruby rub
y
--------------------------------------------
--------------------------------------------
ruby-on-rails: config production heroku render devise model end activereco
rd rail rail
--------------------------------------------
--------------------------------------------
ruby-on-rails-3: blank rspec config original assignment 27 rail couple dep
loy rail
--------------------------------------------
--------------------------------------------
scala: repo transform case extend equal play def val scala scala
--------------------------------------------
--------------------------------------------
security: directly https certificate everything provider software injectio
n security secure security
--------------------------------------------
--------------------------------------------
shell: go sort correct thing 22 parent attach unix shell shell
--------------------------------------------
--------------------------------------------
spring: final response nest entities br primary info bean spring spring
--------------------------------------------
--------------------------------------------
sql: top system per varchar info whole select production sql sql
--------------------------------------------
--------------------------------------------
sql-server: select procedure sql email recommend transaction varchar serve
r studio server
--------------------------------------------
--------------------------------------------
string: remain sign large assign lose tab perform parent string string
--------------------------------------------
--------------------------------------------
svn: fine branch sit ask explorer commit repository svn subversion svn
--------------------------------------------
--------------------------------------------
swift: aware developer active optional protocol init var func let swift
```

```
--------------------------------------------
--------------------------------------------
swing: also practice confuse feel gui yet frame new component swing
--------------------------------------------
--------------------------------------------
templates: type codeigniter deal declare article drop simple template temp
late templates
--------------------------------------------
--------------------------------------------
tsql: statement 32 create interest transaction equal real already xml top
--------------------------------------------
--------------------------------------------
twitter-bootstrap: browser important render orientation launch center clas
s less twitter bootstrap
--------------------------------------------
--------------------------------------------
unit-testing: let several trouble things anybody test unit mock unit test
--------------------------------------------
--------------------------------------------
vim: visual let better modify mode stack plugin switch somewhere vim
--------------------------------------------
--------------------------------------------
visual-studio: pointer want whenever visual debug studio big studio visual
vs
--------------------------------------------
--------------------------------------------
visual-studio-2010: modules visual great anyway press hit 45 definition 20
10 2010
--------------------------------------------
--------------------------------------------
wcf: zero reason security iis layer software cancel bind service wcf
--------------------------------------------
--------------------------------------------
windows: checkbox whether kernel 40 drive dll apps batch windows windows
--------------------------------------------
--------------------------------------------
winforms: explorer operation success windows notification control form for
m close winforms
--------------------------------------------
--------------------------------------------
wpf: raise resource child control property visual xmlns ui bind wpf
--------------------------------------------
--------------------------------------------
xcode: expect bundle logic target step framework profile attach xcode xcod
e
--------------------------------------------
--------------------------------------------
xml: clean apple microsoft integration tutorial template xs layoutwidth xm
l xml
--------------------------------------------
```

In [ ]: