Stack Exchange Question Classifier ☆

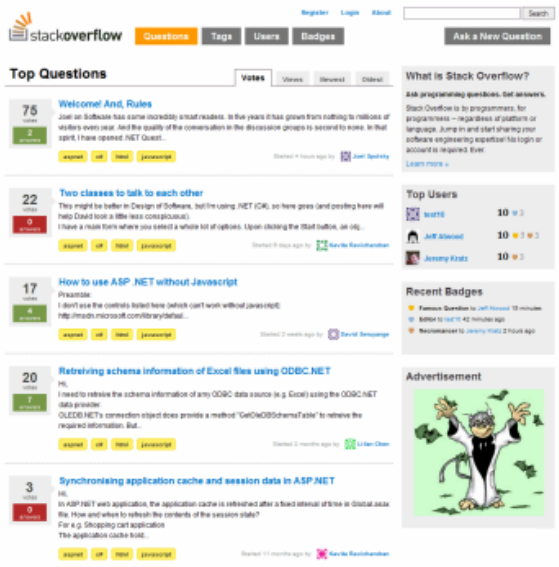
Points: 390.30000000000007 Rank: 1302

Problem Submissions Leaderboard Editorial

RATE THIS CHALLENGE

☆ ☆ ☆ ☆ ☆

Stack Exchange is an information powerhouse, built on the power of crowdsourcing. It has 105 different topics and each topic has a library of questions which have been asked and answered by knowledgeable members of the StackExchange community. The topics are as diverse as travel, cooking, programming, engineering and photography.

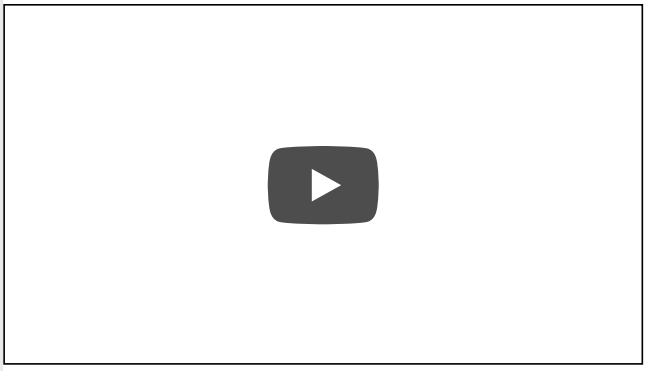


We have hand-picked ten different topics (such as Electronics, Mathematics, Photography etc.) from Stack Exchange, and we provide you with a set of questions from these topics.

Given a question and an excerpt, your task is to identify which among the 10 topics it belongs to.

Getting started with text classification

For those getting started with this fascinating domain of text classification, here's a wonderful Youtube video of Professor Dan Jurafsky from Stanford, explaining the Naive Bayes classification algorithm, which you could consider using as a starting point.



Input Format

The first line will be an integer N. N lines follow each line being a valid JSON object. The following fields of raw data are given in json

- question (string) : The text in the title of the question.
- excerpt (string) : Excerpt of the question body.
- topic (string) : The topic under which the question was posted.

The input for the program has all the fields but topic which you have to predict as the answer.

Constraints

1 <= N <= 22000

topic is of ascii format

question is of UTF-8 format

excerpt is of UTF-8 format

Output Format

For each question that is given as a JSON object, output the topic of the question as predicted by your model separated by newlines.

The training file is available here. It is also present in the current directory in which your code is executed.

Sample Input

```
12345
json_object
json_object
json_object
.
.
.
json_object
```

Sample Output

```
electronics
security
photo
.
.
.
mathematica
```

Sample testcases can be downloaded here for offline training. When you submit your solution to us, you can assume that the training

file can be accessed by reading "training.json" which will be placed in the same folder as the one in which your program is being executed.

Scoring

While the contest is going on, the score shown to you will be on the basis of the Sample Test file. The final score will be based on the Hidden Testcase only and there will be no weightage for your score on the Sample Test.

Score = MaxScore for the test case * (C/T)

Where C = Number of topics identified correctly and

T = total number of test JSONs in the input file.

Change Theme C++

Find No Results

```
1  #include <cmath>
2  #include <cstdio>
3  #include <vector>
4  #include <iostream>
5  #include <algorithm>
6  using namespace std;
7
8
9  int main() {
10     /* Enter your code here. Read input from STDIN. Print output to STDOUT */
11     return 0;
12 }
13
```

Line: 1 Col: 1

Upload Code as File Test against custom input Run Code Submit Code