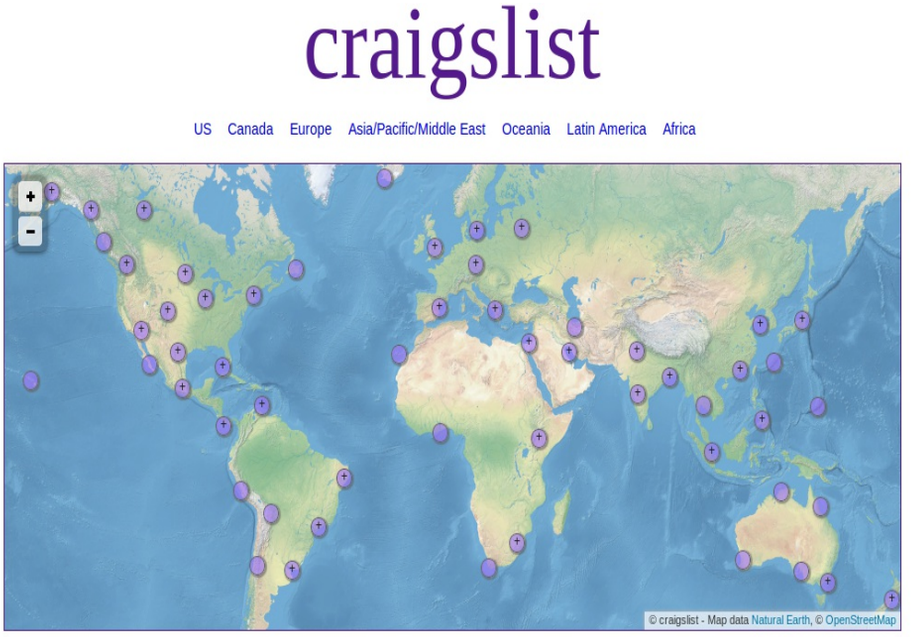
Craigslist Post Classifier: Identify the Category

Craigslist is a powerful platform and forum for local classified advertisements. It has 9 prominent **sections**: jobs, resumes, gigs, personals, housing, community, services, for-sale and discussion forums.

Each of these sections is divided into subsections called *categories*. For example, the services section has the following categories under it:
beauty, automotive, computer, household, etc.



For a set of sixteen different cities (such as newyork, Mumbai, etc.), we provide to you data from **four sections**

- for-sale
- housing
- community
- services

and we have selected a total of 16 **categories** from the above sections.

- activities
- appliances
- artists
- automotive
- cell-phones
- childcare
- general
- household-services

- housing
- photography
- real-estate
- shared
- temporary
- therapeutic
- video-games
- wanted-housing

Each category belongs to only 1 section. Given the city, section and heading of a Craigslist post, can you predict the category under which it was posted?

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

- city (string) : The city for which this Craigslist post was made.
- section (string) : for-sale/housing/etc.
- heading (string) : The heading of the post.

each of the fields have no more than 1000 characters. The input for the program has all the fields but *category* which you have to predict as the answer.

Constraints
1 <= N <= 22000
city is of ascii format
section is of ascii format
heading is of UTF-8 format

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

Training File
A total of approximately 20,000 records have been provided to you, proportionally represented across these sections, categories and cities. The format of training data is the same as input format but with an additional field "category", the category in which the post was made.

The training file (2.2 mb) 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

```
shared
automotive
cell-phones
...
...
...
```

Available Files for Training and (Sample) Tests

The training file as well as the sample tests are available here, to help you build your classification model.
Training File
Sample Test Input
Sample Test Output

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 categories identified correctly and T = total number of test JSONs in the input file.