INF 551 – Fall 2019

Homework #1: Firebase and JSON

Due: September 22, Sunday (end of day)
100 points

In this homework, we consider the LA restaurants health inspection data set: restaurants.csv, in the CSV format. We are only concerned with two columns in the data set: serial_number, facility_name and score.

1. [60 points] Write a Python script (with REST requests embedded) called "load.py". The script will do two things:
   - Convert the data (only the needed three columns) into the JSON format and load the dataset into Firebase. You may need Python "requests" package as shown in class.
   - Create an inverted index for the facility_name column. The index stores, for each unique word in the name (you can assume that words are delimited by white spaces and punctuation characters), the serial_number of restaurant whose name contains the word.

     For example, the name of the first restaurant (serial number = DAJ00E07B) has 3 unique words: habitat, coffee, and shop. You should lower case the words in the index. The index looks like the following:

```
{"index": {
      "habitat":  [DAJ00E07B, …],
      “coffee”:  [DAJ00E07B, …],
      “shop”:   [DAJ00E07B, …],
      …
}
```

   Execution format:

   - python load.py restaurants.csv

- [40 points] Write a Python script called "search.py". The script takes a list of keywords and returns names and scores of restaurants whose **name** contains one or more keywords in the list. The search needs to be executed using the data stored in your Firebase database and use the above index. Note that the search is NOT case-sensitive. For example,
  - python search.py "coffee shop"

  should return the restaurants whose name contains "coffee" or "shop" or both.

Submissions: Name your 2 scripts as below and submit to Blackboard by the due time. **DO NOT** place them in a folder or zip file.

- <FirstName>_<LastName>_load.py
- <FirstName>_<LastName>_search.py

Note: Please use Python 3.6 for your homework. To install Python 3.6 on EC2, execute this:

```
sudo yum install python36 python36-pip
```

However, please do not remove Python2 from EC2, which may be needed for Spark.

To execute the new Python, type: python3, instead of python.

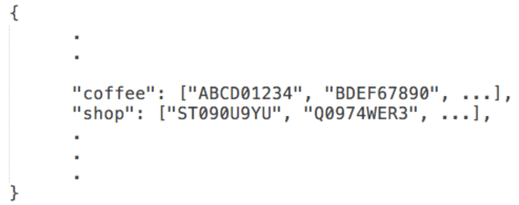
Note that the new usage of pip in Python 3:

```
sudo python3 -m pip install <package-name>
```

The Python packages allowed in this homework are: requests, json, pandas, and numpy.

Output Format:

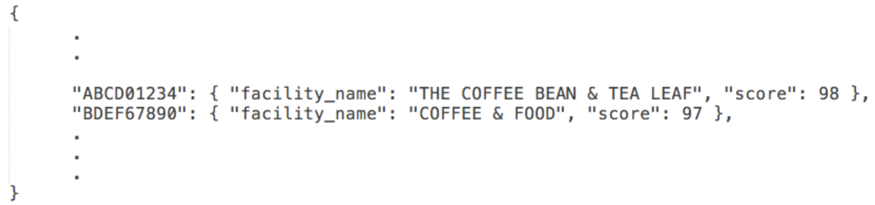
For <FirstName>_<LastName>_load.py, print the response in JSON format after using requests.put to upload your inverted index to the database.



```
{
      .
      .

      "coffee": ["ABCD01234", "BDEF67890", ...],
      "shop": ["ST090U9YU", "Q0974WER3", ...],
      .
      .
      .
}
```

For <FirstName>_<LastName>_search.py, print the result in the following format.



```
{
      .
      .

      "ABCD01234": { "facility_name": "THE COFFEE BEAN & TEA LEAF", "score": 98 },
      "BDEF67890": { "facility_name": "COFFEE & FOOD", "score": 97 },
      .
      .
      .
}
```