Homework #1: Firebase and JSON

Due: January 31, Wednesday (end of day)
100 points

In this homework, we provide you with a CSV file: "menu-200.csv". The file contains a list of 200 old restaurant menus. Each menu has an ID, name, sponsor, event, date, location, etc. Note that not every menu has values for all attributes. Here are your tasks:

- [60 points] Write a Python script (with REST requests embedded) called "load.py". The script will do two things:
  - Convert the given csv file into JSON format and store it in Firebase. You may need Python "requests" package as described in the lecture.
  - Create an inverted index for the event attribute of the menus. For this, you need to parse the content of event attribute and extracts unique tokens (which are defined as a sequence of alphabetic characters) among the values of all attributes.
  For example, "[DINNER]" has only one token "DINNER" (or "dinner" if you decide to lower case it), while "FRUHSTUCK/BREAKFAST;" has two tokens.
  You then create an index in Firebase to record, for each token, the list of ids of menus which contain the token.
  For example:

    ```
    {"index": {
       "breakfast": [12463, 12470, ...]
       ...
    }
    ```

    which says "breakfast" appears in menu 12463, 12470, …

  Execution format:

  - python load.py menu-200.csv

- [40 points] Write a Python script called "search.py". The script takes a list of keywords (which may contain multiple tokens) and return a list of ids of menus whose **event** field contains **any** keywords in the list. The search needs to be executed using the data stored in your Firebase database. Note that the search is NOT case-sensitive. For example,
  - python search.py "dinner breakfast"

  should return the ids of menus that contain dinner (and DINNER), breakfast, 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 2.7 (installed by default on EC2) for the coursework.