

Homework #1: Firebase and JSON

Due: February 1st, Friday (end of day)

100 points

In this homework, we provide you with an JSON data set on bank accounts: “accounts.json”. The data set contains 995 accounts. Ignore the “... index ...” lines. For each account, it records information such as account number, the name, address, and gender of account holder.

1. [60 points] Write a Python script (with REST requests embedded) called “load.py”. The script will do two things:

- Load the dataset into Firebase. You may need Python “requests” package as shown in class.
- Create an inverted index for the address field of accounts. The index stores, for each unique word in the address (you can assume that words are delimited by white spaces), the id’s of account whose address contains the word.

For example, the address of the first account (id = 1) has 3 unique words: 880, homles, and lane. You should lower case the words in the index. The index looks like the following:

```
{ "index": {  
    "880": [1],  
    "homles": [1],  
    "lane": [1, 70, ...],  
    ...  
}
```

Execution format:

- python load.py accounts.json
- [40 points] Write a Python script called “search.py”. The script takes a list of keywords and return a list of ids of accounts whose **address attribute** contains some 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 “maple lane”

should return the ids of accounts whose address contains either “maple” or “lane” 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.

INF 551 – Spring 2019

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

Note: Please use Python 2.7 (installed by default on EC2) for the coursework. Note that print and other functions may be different in Python 3.