Final Project Document

Name: Jinhui Jia

Uniquename: jiajh

**Project code:**

Github repo: <https://github.com/qmkazo/507final>

Readme.md and requirements.txt are provided in github.

**Data sources：**

**Data origin**:

<https://www.usnews.com/best-colleges>

<https://www.mapquest.com/>

**Access techniques:**

In this program, Beautifulsoup4 is used to crawl data from websites. Once proper condition is provided, the program can automatically fetch data from usnews.com. Colleges data will be fetched, storage into database and processed into instances. The program will then use the instances to build figures and graphs. Caching is used in this program. A file “final\_cache.py” will storage the data fetched from websites. When the data are used for a second time, the program will use cache file rather than fetching again.

**Data summary:**

The raw data is HTML of the two websites. The program can store the raw data into json caches and sql database. Every time the user checks the list of a state, the data will be stored in database as well as in the json cache file.

**Database**

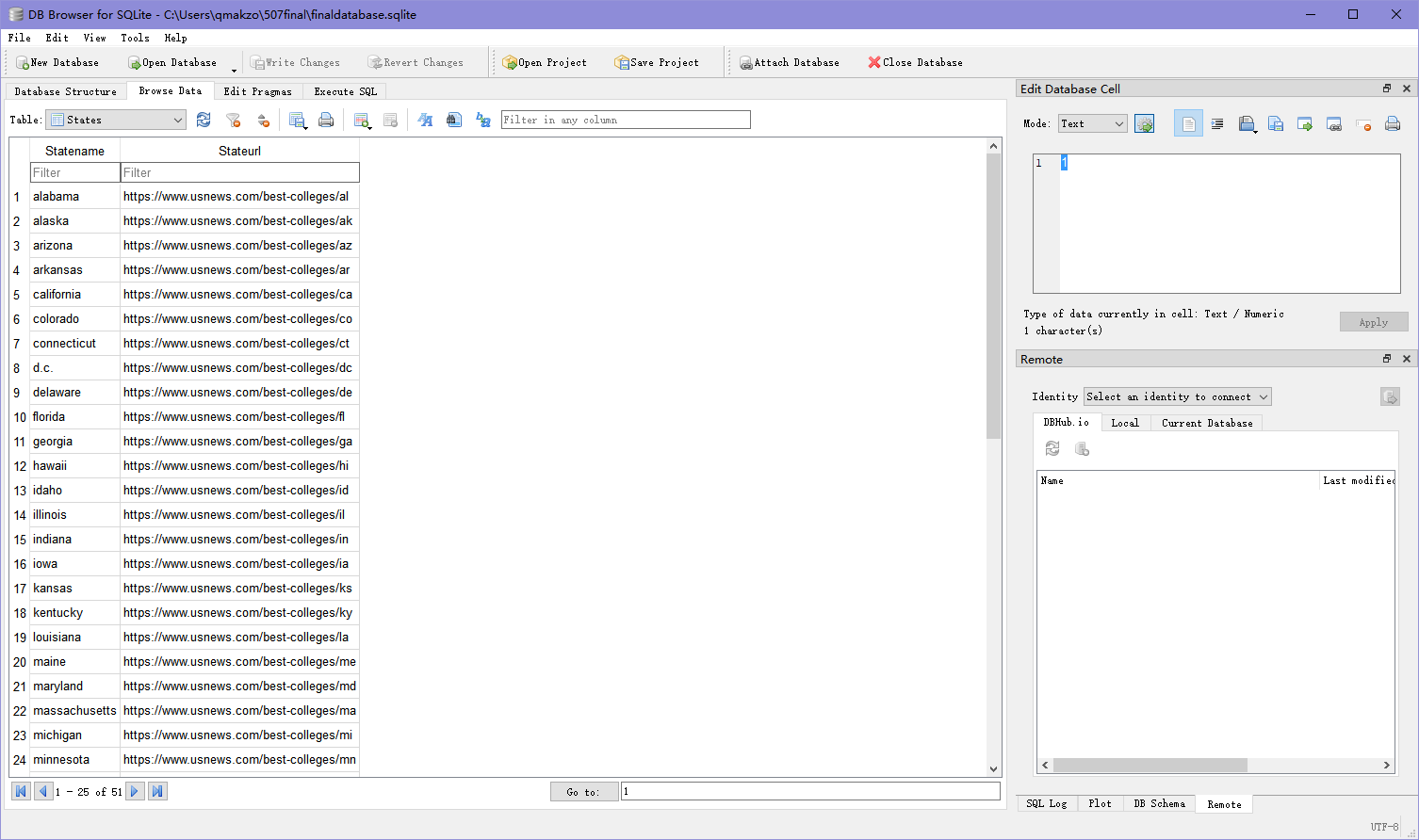
The database consists of 2 tables, which are built in “main.py” using SQL CREATE TABLE command while processing data.

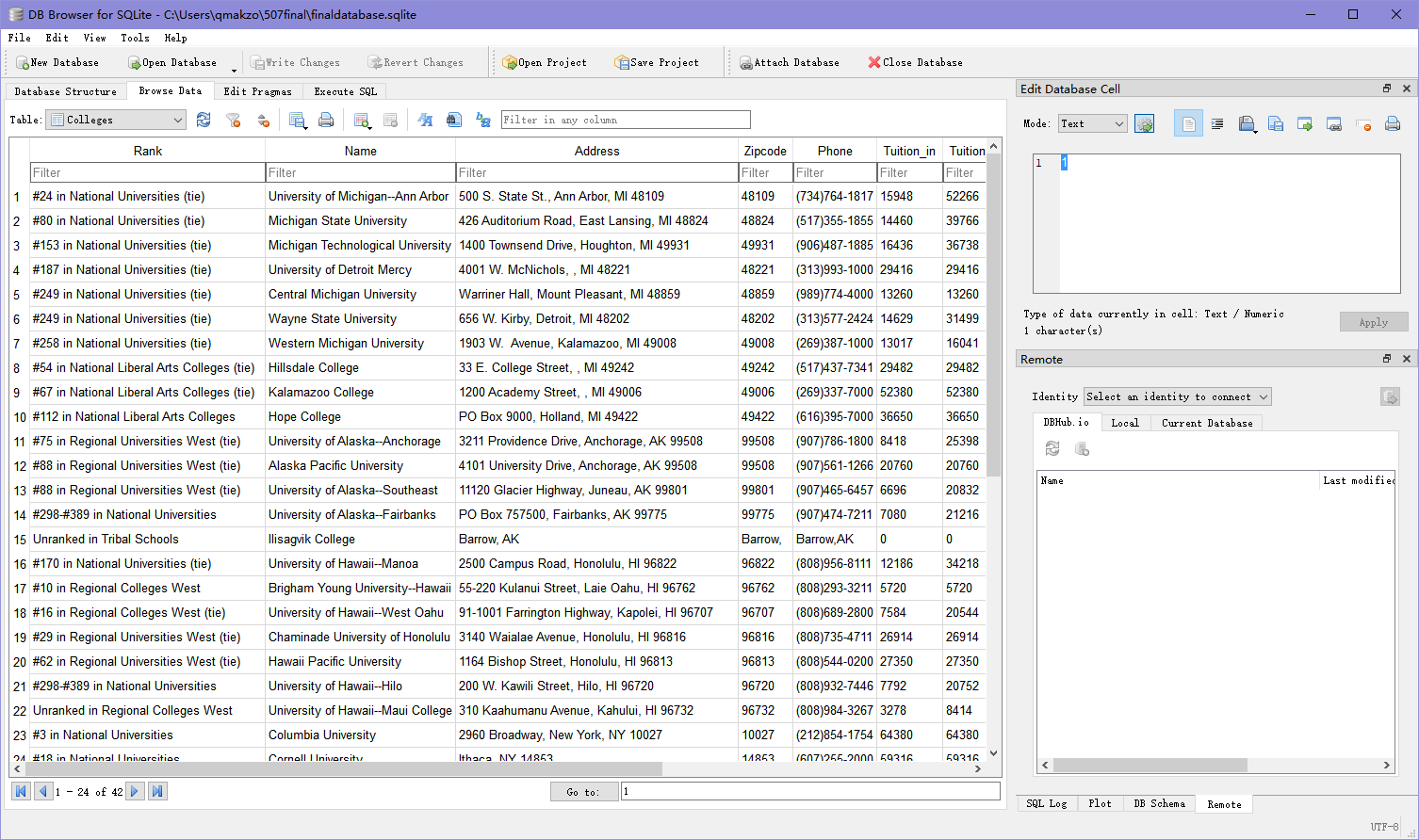
The two tables are:

1.States. Attributes include names and urls of the states.

2.College. Attributes include rank, name, address, zipcode, in-state tuition, out-of-state tuition, url, enrollment and state of the colleges.

The key-primary key linking the two tables is “state” and “statename”





**User interaction and options:**

First, users type a state name in command line prompt. Then a list of top colleges in this state will be shown. Meanwhile, the information will be stored in file “finaldatabase.sqlite”.

In this page, user have 4 options：

1 Type a college number to check the detail of this college.

2.Type “tuition” to check a graph comparing tuition among the colleges

3.Type “enrollment” to check a graph comparing enrollment among the colleges

4.Type “details” to check the detail table of the colleges

If the user chooses to check the detail of a college, the next page will ask the user to check the website of the college. User can type “y” or “n” to choose. If user chooses “y”, the website of the college will open in a browser.

Then, the page will ask the user to check the map and nearest places of the college. User can type “y” or “n” to choose. If user chooses “y”, a table of nearest places and a website of map will be opened.

Users can type “back” or “exit” at any time to get back to former page or exit.

**Data presentation**

This program uses Plotly to present data. File “my\_plotly.py” is the graph module of this program. Main file will import this file and use the functions to make graphs and figures. Bar graphs and tables can be presented by this program.

**Demo video link**: https://youtu.be/zGKCZsPZsHg