Setting up the Project

1. Download and install pycharm Community edition

2. Download and install python 3.6

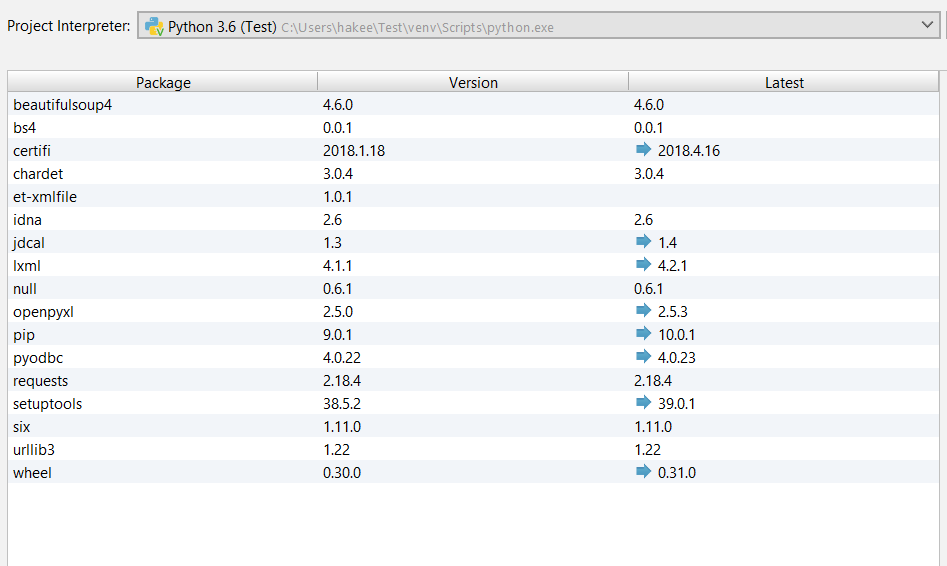
3. Open pycharm community edition and create a new project.

4. Choose an interpreter for the project: (File -> Settings -> Project Interpreter)

5. Install the necessary packages (File -> Settings -> Project Interpreter -> Click the green plus button -> Type the name of the following packages and install one at a time) -

pyodbc, lxml, bs4, openpyxl, requests

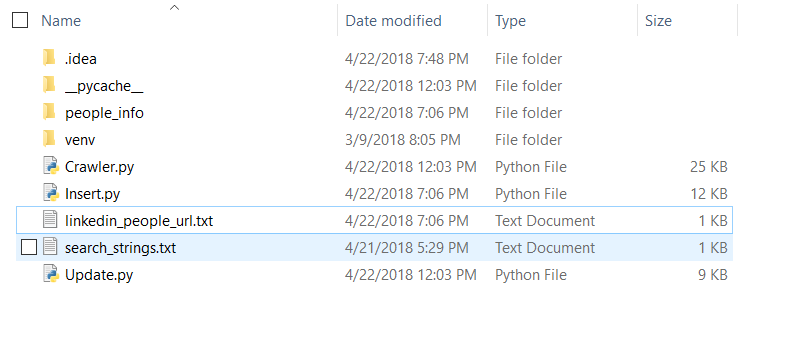
Your project interpreter window should resemble this afterwards:



6. Add the scripts to your folder. You may need to create the script files (using the same names) first within pycharm and then copy the contents into the newly created files.

7. Add the people\_info folder, search\_strings.txt file, and the linkedin\_people\_url.txt file to your project folder.

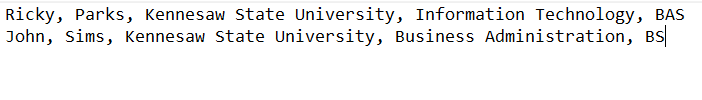
Your project folder should resemble this:



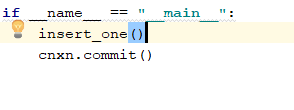
Running the Scripts

* Only the Insert and Update scripts will insert/update data in the database
* There are two methods in the Insert script that will run the crawler and insert data in the database:
  + insert\_one(): This method will read from the search\_strings.txt file and use those lines as search parameters. If you want to add another alumni to lookup, and another line that’s in the following format: FirstName, LastName, University Name, Field of study, Degree type (All separated by commas)

Example:



* + insert\_many(): This method will prompt you for user input. The method will use your input as a search string to crawl many students and insert their information in the database.
  + If you want to change which method to run, change the method in the main method of the insert script.



* Run the update script to update data in the database. It will retrieve all the linkedin urls from the database and

Changing the Username and password

If you want to use a different username and password for the linkedin account, edit the username and password variables in the Insert.py script.

Viewing the data in the database:

* Install SQL Server
* Connect to the server: itcapstone.database.windows.net (username: capstone, password: Alumnidatabase!)
* Also, the DDL is included in this folder if you want to create a local database.

