

## OPEX-462 Handshake Automation

Google Drive Link to store the document created using this template-

JIRA : <https://jira.jerseystem.org/browse/OPEX-462>

Google Collab Link :

<https://colab.research.google.com/drive/1BcwVm0hbeJZGmVdwvuvM7MFHycAw8FLk?usp=sharing>

### Description :

Handshake Automation is a Python tool for extracting student candidate data from Handshake, an online recruiting platform for higher education students and alumni. Mainly use Seleium and Beautiifulsoup library to see the web interactions and gspread for manipulation of Google Sheets.

Here are some key functions included in the library:

#### 1. Initial Setup and Imports

- **Selenium and tqdm installation:** These are library installations for web scraping (selenium) and for displaying progress bars (tqdm).
- **Google Sheets Setup:** The script uses the gspread library to interact with Google Sheets. Users are prompted to enter a Google Sheet URL, which is then accessed and manipulated through the script.

#### 2. Key Method

- **login(url, email, password) :**
  - **Purpose:** This method is responsible for logging into a website.
  - **Parameters:**
    - URL: The URL of the login page.
    - Email: The user's email address used for login.
    - Password: The user's password used for login.
  - **Description:** The function navigates to the specified URL, enters the provided email and password into the corresponding fields, and submits the login form. It typically returns a logged-in session or driver object that can be used for subsequent interactions with the website.
- **filter\_school(driver, school\_names) :**
  - **Purpose:** This method is responsible for filtering schools on a web page.

- **Parameters:**
    - **driver:** The web driver or session object used for interacting with the website.
    - **school\_names:** A list of school names to be used as filters.
  - **Description:** The function interacts with the web page, selects the specified schools from a list or dropdown, and applies the filters. It could be part of a broader data extraction process where the user wants to focus on specific schools.
- **extract\_data(driver) :**
    - **Purpose:** This method is responsible for extracting data from a web page.
    - **Parameters:**
      - **driver:** The web driver or session object used for interacting with the website.
    - **Description:** The function uses the provided driver to access the current page's HTML, extracts relevant data (such as names, active statuses, degrees, education, and skills), and returns the data in a structured format, often as a DataFrame.

These methods, when combined, can form the core functionality of a web scraping or automation script for a specific website. The `login` function establishes a session, `filter_school` refines the data to be extracted, and `extract_data` retrieves the relevant information from the web page.

## Important Notes

### How to Use the Handshake Student Data Scraping

1. Open Colab using the [link](#)
2. Go to runtime, and click Run All
3. Enter the proper email and password for the Handshake recruiting account.
4. Answer whether you need a filtering option.

### Demo video :

[https://drive.google.com/file/d/1wGA-Jxe5EoB7He2cg1hzatBDFe-9mvVH/view?usp=drive\\_link](https://drive.google.com/file/d/1wGA-Jxe5EoB7He2cg1hzatBDFe-9mvVH/view?usp=drive_link)