

DSCI 560: Group Laboratory 6

Table of Contents

- [Team Members](#)
- [Installation](#)
- [Execution](#)

Team Members

1. Hanlu Ma (USC ID: 1392-9443-71)
2. Zhenyu Chen (USC ID: 2242-3773-15)
3. Fariha Sheikh (USC ID: 9625-7343-53)

Installation

Install the require libraries:

1. Install the **sqlalchemy** Library:

```
1 pip install sqlalchemy
```

2. Install the **OpenCV (cv2)** Library:

```
1 pip install opencv-python
```

3. Install the **numpy** Library:

```
1 pip install numpy
```

4. Install the **requests** Library:

```
1 pip install requests
```

4. Install the **pytesseract** Library:

```
1 pip install requests
```

5. Install the **pytesseract** Library:

```
1 pip install pytesseract
```

6. Install the **tqdm** Library:

```
1 pip install tqdm
```

7. Install the **pdf2image** Library:

```
1 pip install pdf2image
```

8. Install the **joblib** Library:

```
1 pip install joblib
```

9. Install the **BeautifulSoup (bs4)** Library:

```
1 pip install bs4
```

10. Install the **lxml** Library:

```
1 pip install lxml
```

Execution

First, we execute **database.py** file to create the table in the SQL database

```
1 python database.py
2 "Please enter the username for the database:" [username] #Enter
  MySQL username.
3 "Please enter the password for the database:" [password] #Enter
  MySQL passowrd.
4 "Please enter the database name:" [database] #Enter the database
  name where you want to store the data.
```

Then, we execute **pdf_extraction.py** file to parse the information from pdf.

```
1 python pdf_extraction.py
2 "Please enter the username for the database:" [username] #Enter
  MySQL username.
3 "Please enter the password for the database:" [password] #Enter
  MySQL passowrd.
4 "Please enter the database name:" [database] #Enter the database
  name where you want to store the data.
```

Lastly, we execute **web_scraper.py** file to parse the information from the web.

```
1 python web_scraper.py
2 "Please enter the username for the database:" [username] #Enter
  MySQL username.
3 "Please enter the password for the database:" [password] #Enter
  MySQL passowrd.
4 "Please enter the database name:" [database] #Enter the database
  name where you want to store the data.
```

Notes

We include many other programs in the script folder, but they are not necessarily all useful as we mainly use it as a function call reference if we need. You can freely check any other scripts for more information that you are interested in. However, our main scripts for this lab will be `database.py`, `pdf_extraction.py`, and `web_scraper.py`.