

# Student Information Project

---

## Abstract

---

This program was created for the Carleton College registrar's office, in purpose of splitting student documents and parsing them into one pdf file. This README is written for users assuming that the users already have an IDE(Integrated Development Environment) like [Visual Studio Code](#) or [PyCharm](#) and python3 installed in their computers.

---

## Table of Contents

---

- [Description](#)
  - [Files](#)
  - [Instruction](#)
  - [Credits](#)
- 

## Description

---

This project was made to do the following:

- Split a text file with multiple academic review sheets into text files that contain academic review sheets.
  - Split a text file with multiple transcripts into text files that contain individual transcripts.
  - Split a html file with multiple degree audits into html files that contain individual degree audits.
  - Convert the individual files into pdf format.
  - Merge multiple pdf files for individual students into one pdf file.
  - Name the merged pdf file as desired file name.
- 

## Files

---

- **modify\_txt.py** : add line at and remove line from the end of text file
  - **split\_review\_files.py** : split file into individual review sheets and saves them as "student\_id\_1\_review.txt" in review\_sheets folder
  - **split\_transcript\_files.py** : split file into individual transcripts and saves them as "student\_id\_2\_transcript.txt" in transcripts folder
  - **split\_degree\_audit.py** : split file into individual degree audits and saves them as "student\_id\_3\_degree\_audit.html" in degree\_audit folder
  - **convert\_pdf.py** : Convert html and txt files into pdf format
  - **extract\_file\_name.py** : Extract file names from spreadsheet
  - **student.py** : Find, move, convert, merge, and name files for each student
  - **split\_program.py** : Main program
- 

## Instruction

---

If this is the first time you are downloading or using this program, you may need to download several packages that allow you to convert files into pdf and merge pdfs. The following are the package name, information link, and the command to download the packages.

### pyhtml2pdf

Further information about the package can be found [here](#).

```
python3 -m pip install pyhtml2pdf
```

### FPDF

Further information about the package can be found [here](#).

```
python3 -m pip install fpdf
```

### PyMuPDF

Further information about the package can be found [here](#).

```
python3 -m pip install --upgrade pip
python3 -m pip install --upgrade pymupdf
```

After downloading the packages above, you can now use the program by going through the following steps.

1. Add five files in the *student\_info\_project* folder and rename them like the following:
  - o a. **TEXT** file named *review\_sheet.txt* that contains the **Review Sheets**.
  - o b. **TEXT** file named *transcript.txt* that contains the **Transcripts**.
  - o c. **HTML** file named *degree\_audit.html* that contains the **Degree Audits**.

- d. **Folder** named *d\_f\_forms* that contains all D-F Forms.
  - e. **CSV** file named *file\_names.csv* that contains the desired file names for all students.
2. Open the terminal and run the following command:

```
python3 split_program.py review_sheet.txt transcript.txt degree_audit.html file_names.csv
```

After running the program, you will now find four sub-folders newly made in the *student\_info\_project* folder:

1. **review\_sheets** folder that contains individual review sheets.  
Files are named in format *student\_id\_1\_review.txt*.
2. **transcripts** folder that contains individual transcripts.  
Files are named in format *student\_id\_2\_transcript.txt*.
3. **degree\_audits** folder that contains individual degree audits.  
Files are named in format *student\_id\_3\_degree\_audit.html*.
4. **combined\_folder** folder that contains following folders.
  - a. **all** folder that contains combined pdf files for all students.  
Files are named in format *student\_id.pdf*.
  - b. Folders named by student ID(ones existing in the inputted files) each containing the following:
    - i. **raw** folder that contains all individual documents for each student in their original format.
    - ii. **pdf** folder that contains all individual documents for each student in a pdf format.
    - iii. **all** folder that contains a single pdf file with all documents combined.

---

## Credits

---

The splitting program for degree audits was provided by misplonj, and was modified during the process of putting this program together.