

How to import top200 records into database

Description:

The top 200 records are the result of the analysis performed by Jon Tryggiv from department of Chemistry and Biochemistry at University of Arizona. The result of his analysis is available on his university blog and in PDF format only and have not been updated since 2018. For more information please refer to his blog: <https://njardarson.lab.arizona.edu/>

How to run the code:

Pre-installation requirement:

The following libraries are required to run the import script “top_pharmaceuticals_poster_load.py”

Please run the following commands one by one to ensure the libraries are installed, if planning to run this from parent project folder, requirement.txt includes these libraries.

```
1. pip install numbers
2. pip install numpy
3. pip install pandas
4. pip install SQLAlchemy
5. pip install pylint-flask-sqlalchemy
```

Database credentials:

open top_pharmaceuticals_poster_load.py file and provide database credential as following:

```
10 DATABASE_URL = "postgresql://YOUR_USERNAME:YOUR_PASSWORD@DB_URL:5432/drugdata"
11
```

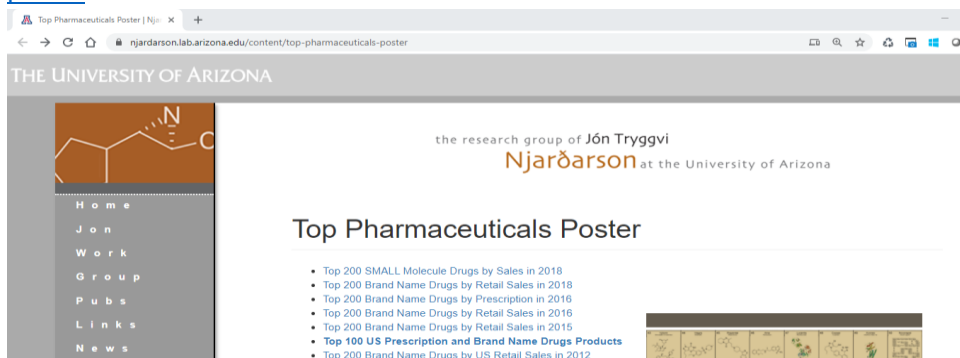
Run the script:

- Navigate the folder that top_pharmaceuticals_poster_load.py exist and ensure the source file “top_pharmaceuticals_poster_data.xlsx” exist.
- Run the script by running:

```
C:\Development\School\CSCI-E 599\M3\code_4_22_2020\import_script_doc\import_scripts\top200>python top_pharmaceuticals_poster_load.py
top 200 records are successfully inserted
```

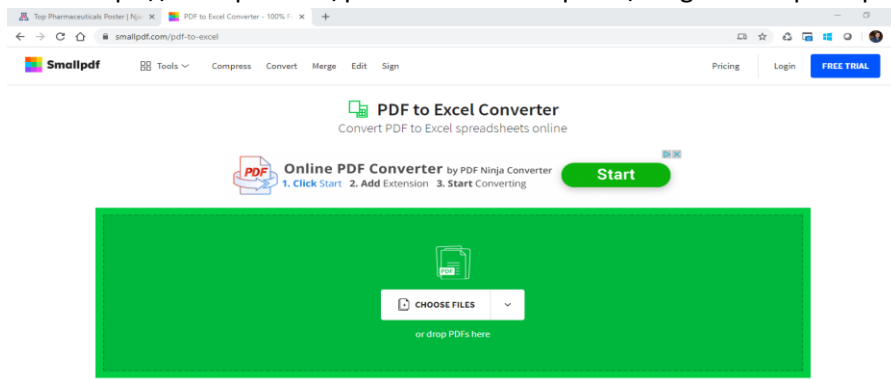
Steps to convert PDF to excel for future use cases:

1. Download the PDF files from <https://njardarson.lab.arizona.edu/content/top-pharmaceuticals-poster>



Please note to convert the PDF pictures to excel files, please note many libraries have been attempted to properly parse the PDF pictures to excel file, however the best optimal option is to leverage a free tool <https://smallpdf.com/pdf-to-excel>

2. Go to <https://smallpdf.com/pdf-to-excel> and upload/drag and drop the pdf file to



3. Click Download

