Project: Publication Scrapper

Introduction

The Publication Scrapper project is designed to extract and visualize data from scholarly publications. It consists of three program files: ***scholar.py, arxiv.py,*** and ***plot.py***. This document provides an overview of each file and instructions for running the project.

Files

scholar.py

This file is responsible for extracting and storing data from the Scholar platform. ***However, it's important to note that the Scholar platform may block the IP address if the number of requests exceeds a certain limit.*** To overcome this limitation, the code in scholar.py writes the extracted data to a CSV file. This allows the data to be plotted using plot.py without making additional requests to Scholar.

arxiv.py

The arxiv.py file is specifically designed to extract data from the ArXiv platform. It does not have the same IP blocking limitations as Scholar, so it directly extracts and stores the data in a CSV file and plot it aswell.

plot.py

If you only want to visualize the data, you can use plot.py. This file is a GUI application that reads the CSV file (generated by scholar.py or arxiv.py) and displays it as a multi-line graph.

Instructions

Before running the project, make sure you have installed all the required dependencies mentioned in the requirements.txt file. You can install the dependencies by running the following command:

Copy code:

```
pip install -r requirements.txt
```

To visualize the data, run plot.py. The GUI application will read the CSV file (generated by either scholar.py or arxiv.py) and display it as a multi-line graph.

Note: If you encounter any issues or errors during the execution of the project, please refer to the error messages or consult the project documentation for troubleshooting guidance or just contact me.