News Aggregator

PyCK Project (Abstract)

About Project

I am planning to make a News Aggregator (web scraper) as my project. The aim will be to collect the latest news from some specific news websites that contain some specific keywords (according to personal preference). Then that collection will be displayed to the user in a well organised web page.

Libraries Used

requests
Beautiful soup
Django framework

Timeline

- Before end-sems (21 June): Start trying out Django and basics of making a web-app.
- After endsems(25 June):
  - Week1: Learn more about Django and implement a basic Django app that can extract information from HTML pages.
  - Week2: Complete the web app to filter the news according to keywords.
    Add a good look to the webpage so that it actually becomes usable. Add more features to it like extracting articles from websites that need user credentials (twitter/insta), organising the data according to number of likes/shares/trending, etc.

Extra things to learn

Django framework
Web scraping
HTML/CSS

Motivation

I learnt about the concept of web scraping quite long ago; it is interesting to know that our programs can interact with the web too. But I neither knew about web nor about python that much then, so I dropped the idea for then.
Recently I have been following HTML/CSS/bootstrap tutorials too from online resources. And now that I know enough about Python too, I think I can try it out.

Why a new scraper ? Simply because otherwise I hardly read news. Also, I haven't made any practically usable program till date. So it will be really cool if I can make something that I actually use in my daily life.

References

I was going through one of the links provided by the PyCK team when I remembered about web scraping.
https://data-flair.training/blogs/python-project-ideas/
https://data-flair.training/blogs/django-project-ideas/
https://data-flair.training/blogs/django-project-news-aggregator-app/
The blog is itself quite thorough in explaining the project and code but as I can see in the comments, there are quite a few errors and also it shows all the latest news from the web page rather than filtering them according to keywords (as I have planned).