PyCK Project Outline

About the Project

The project will be based in the Data Science field. It is a web-based project app/website, responsive in nature catering to statistics regarding the Covid-19 situation in India/World. The figures in mind to be projected include daily figures, archive, vaccination availability, etc.

Libraries Used

The project is a web-based application, and hence Django would be used much of the time for coding the backend and frontend portion. Along with that web scraping mechanism would be used to fetch data from Covid-rerlated websites, hence, Python libraries like Pandas, Selenium, BeautifulSoup, Matplotlib and frameworks like scrapy could be used to achieve this.

Walkthrough

I am well-versed with HTML, CSS and JavaScript for creating a good responsive frontend design for the website. Django would be used in the frontend and backend portion. Web scraping would be carried out using scrapy and Pandas framework. Afaik it would deal in writing rules for spiders that crawl the web page, and items for the data you want to scrape. Selenium is a testing library that would be used to automate browser activities. BeautifulSoup is a Python package for parsing HTML and XML documents. It creates parse trees that are helpful to extract the data easily. I am planning on using the scraped data, which could be stored in multi-dimensional arrays, made easy using Numpy, further utilised for plotting graphs using Matplotlib (using Mplot3d for 3D graphs). I could try to

use the CoWin API for vaccine availability data, however, it seems unlikely due to its tedious procedure.

Timeline for 2-weeks

Starting with the frontend development, I would code the basic skeleton for the webapp. Thereafter, learning the use of Django in frontend, it would be incorporated accordingly. Next, basics of web scraping using Selenium, and Beautiful Soup, and incorporating the data as maybe a JSON file, or any other method.

Other Things to Learn

The above project requires skills in web development which would include HTML, CSS (or SASS), and JavaScript. For backend programming, Django could be used for interacting with the database or Excel. Libraries such as BeautifulSoup and Selenium would also come handy and need to be learnt.

Motivation

I have already done projects in the web development sector like a blogging website along with user login and authentication, working with databases (mySQL) and websites using CRUD commands of SQL. This project takes this up a notch. It would act as an introduction to data science, as well as enhance the scope of data portrayal on the website, which earlier dealt mostly in hard-coded entries. I have little experience in working with APIs and hence I could use the scraped/fetched data, and fit it dynamically on the website. This project will be an improvement in my web development skills and would allow me to learn about data science in detail. The scope of Python could be understood in web development which for now had been limited in my experience.

References

1. https://docs.scrapy.org/en/latest/
2. https://github.com/checkcheckzz/python-github-projects/blob/master/WebFrameworkandRelatedTool.md
3. https://github.com/checkcheckzz/python-github-projects/blob/master/WebMining.md
4. https://www.edureka.co/blog/web-scraping-with-python/
5. https://www.edureka.co/blog/python-pandas-tutorial/
6. https://www.youtube.com/watch?v=uufDGjTuq34 (Code with Harry)
7. https://www.youtube.com/watch?v=XVv6mJpFOb0 (FreeCodeCamp crash course in BeautifulSoup)
8. I am following a Udemy course on Python and Django for web development.