Rough ideas and resources for fyp

No-SQL vs SQL databases

Default with the Django framework is SQLite 3 which is a no-sql database and is easy to use but unfortunately might not be the best for large-scale web scraping

My next database I was thinking about would be MongoDB which would work well if the data scraped is in JSON format

Another database I thought about using is MySQL database which I have the most experience with it is a relational database and works well with structured data

PostgreSQL is a relational database that is great for structured data and complex queries. It also supports JSON data types if you need flexibility with semi-structured data. It also handles high-concurrency well, making it a good choice for web scraping applications that collect a lot of data.

From a web scraping perspective, responses or data that's available through APIs are

preferred over data that's retrieved using scraping scripts. This is due to the following

reasons:

An API's returned data is completely specific to the requests being performed,

along with the filters or parameters that have been applied to it.

Tasks such as parsing HTML or XML using Python libraries, such

as BeautifulSoup, pyquery, and lxml, isn't always required.

The format of the data is structured and easy to handle.

Data cleaning and processing for final listings will be more easy or might not be

required.

There will be significant reductions in processing time (compared to coding,

analyzing the web, and applying XPath and CSS selectors to retrieve data).

They are easy to process.

MLA 9th Edition (Modern Language Assoc.)

Anish Chapagain. Hands-On Web Scraping with Python : Perform Advanced Scraping Operations Using Various Python Libraries and Tools Such As Selenium, Regex, and Others. Packt Publishing, 2019.

APA 7th Edition (American Psychological Assoc.)

Anish Chapagain. (2019). Hands-On Web Scraping with Python : Perform Advanced Scraping Operations Using Various Python Libraries and Tools Such As Selenium, Regex, and Others. Packt Publishing.