DS115- Data Collection & Web Scrapping

أكاديمية طويق
Tuwaiq Academy
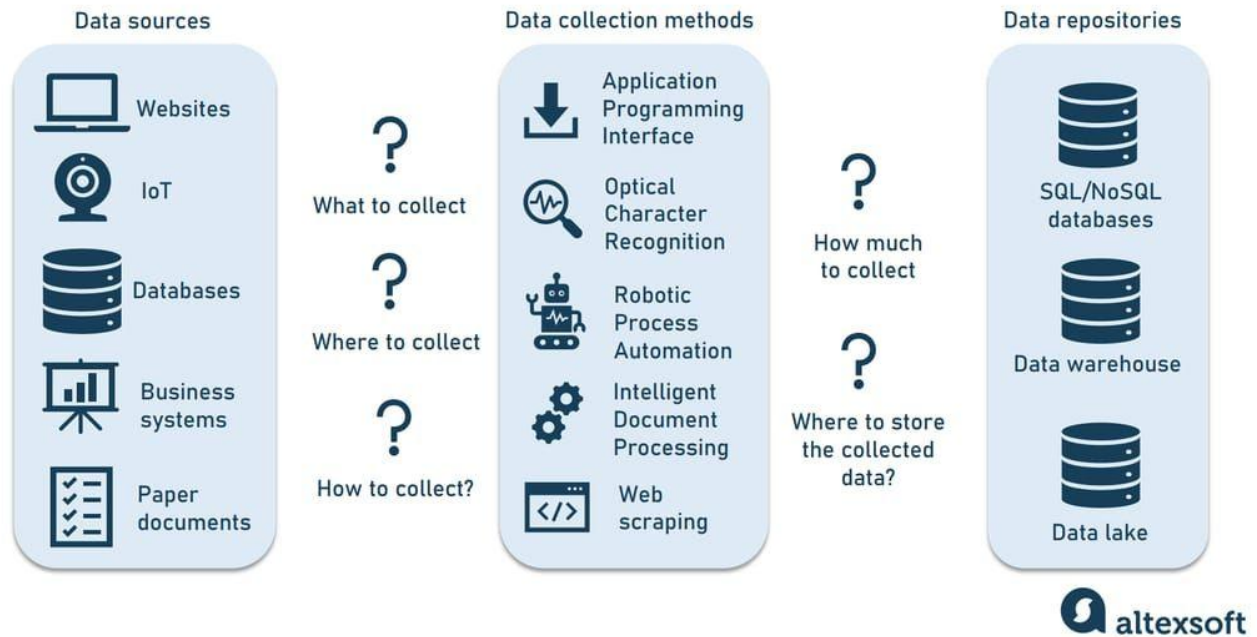
By: eng. Esraa Madhi

What is Data Collection?

Data collection involves gathering and analyzing information from various sources to address research issues, respond to queries, assess results, and predict trends and probabilities

https://youtu.be/Lb6Gi6lR-Kc

PILLARS OF DATA COLLECTION

Web scrapping:

https://youtu.be/Ct8Gxo8StBU

Why is Web Scraping Used?

Web scraping is used to collect large information from websites.

Webpages → Web Scraping → Structured Data

What are the different Python web scraping libraries?

1. <mark>Requests</mark>: Simple library for HTTP requests, handling cookies, and authentication with ease.

2. <mark>BeautifulSoup</mark>: Parses HTML and XML, allowing easy navigation and extraction of elements with robust error handling.

> https://www.tutorialspoint.com/beautiful_soup/index.htm

Beautiful Soup Tutorial •

3. <mark>Selenium</mark>: Automates web browsers for complex user interactions and web testing.

> https://www.geeksforgeeks.org/selenium-python-tutorial/

Selenium Python Tutorial - GeeksforGeeks • www.geeksforgeeks.org

> https://youtu.be/tcMdWM8wmqs

To extract data using web scraping with python, you need to follow these basic steps:

1. Find the URL that you want to scrape
2. Inspecting the Page to find the data you want to extract
3. Write the code to extract the data
4. **Store the data in the required format (cleaning)**

Resources:

- https://www.altexsoft.com/blog/data-collection-machine-learning/
- https://www.questionpro.com/blog/data-collection-methods/
- https://nanonets.com/blog/web-scraping-with-python-tutorial/