

Tech Talk Ideas

10/01/2023, COMP 523 Team-C

Tech Talk: Web Scraping and API Interaction using Python

In the digital era nowadays, data is one of the most important things in people's daily life. It is significantly crucial for our developers to manage the ability to extract, manipulate, and analyze data efficiently. Therefore, our tech talk idea will delve into the knowledge of web scraping and interacting with possible APIs using Python as a higher-level programming language to provide us convenience, combining with possible Python packages like `requests`, `BeautifulSoup`, `json`, `aiohttp`, `lxml`, ...etc. With these convenient tools, we will explore scraping data from websites, alongside interacting with APIs to access data seamlessly.

In the process of scraping data from the web, we need to utilize each of our tools. To be more specific, with `requests`, we could simply make HTTP requests to the target website, and parsing each HTML content using `BeautifulSoup` or `lxml`. Also, we could use `Selenium` with `Javascript` to technically bypass the possible human verification on specific websites. If the website is dynamic, we could also use alternative tools like `Puppeteer` to scrap it. Therefore, as we can see, it is important to know and utilize different convenient packages for us to better scrap the data that we need.

Also, there are lots of well-built websites that provide public/private APIs to the developer. In this scenario, we could possibly utilize the interactions with APIs rather than scraping it by our own. By the interaction, we could usually get standardized data format responses like `JSON` or `XML`, making us easier to work with these data. Also, working with APIs could possibly increase the efficiency since it is usually faster than the web-scraping processes. Other than that, APIs from websites also show better reliability since they are intended for data access and usually have stable endpoints with versioning.