

Web Scrapping

Spoken Tutorial Project

<https://spoken-tutorial.org>

National Mission on Education through ICT

Sai Sathwik Matury

Domain: Dr. T. Subbulakshmi

VIT Chennai

3 October 2024



Learning Objectives

In this tutorial, we will learn to



Learning Objectives

In this tutorial, we will learn to

- Scrape data from any website



Learning Objectives

In this tutorial, we will learn to

- Scrape data from any website
- Extract it to a CSV file



Learning Objectives

In this tutorial, we will learn to

- Scrape data from any website
- Extract it to a CSV file
- **Perform basic data analysis**



Learning Objectives

In this tutorial, we will learn to

- Scrape data from any website
- Extract it to a CSV file
- Perform basic data analysis
- **Generate visualizations**



System Requirements

To record this tutorial, I am using



System Requirements

To record this tutorial, I am using

- **Ubuntu Linux OS 22.04**



System Requirements

To record this tutorial, I am using

- Ubuntu Linux OS 22.04
- Python 3.12.3



Pre-requisites

- You must have basic knowledge of using Linux Terminal and Python



Pre-requisites

- You must have basic knowledge of using Linux Terminal and Python
- For pre-requisite Linux and Python Tutorials, please visit <https://www.spoken-tutorial.org/>



Pre-requisites

- You must have basic knowledge of using Linux Terminal and Python
- For pre-requisite Linux and Python Tutorials, please visit <https://www.spoken-tutorial.org/>
- Python libraries required for automation must be installed



Code Files

- The files used in this tutorial are provided in the Code files link



Code Files

- The files used in this tutorial are provided in the Code files link
- Please download and extract the files



Code Files

- The files used in this tutorial are provided in the Code files link
- Please download and extract the files
- Make a copy and then use them while practicing



Web Scrapping

- **Web Scrapping is the automated process of extracting data from websites with software**



Web Scrapping

- Web Scrapping is the automated process of extracting data from websites with software
- We will automate extracting data and information from web pages and parsing HTML content



Web Scrapping - Libraries



Web Scrapping - Libraries

- Requests **library** to fetch **HTML content** from a **web page**



Web Scrapping - Libraries

- Requests **library** to fetch **HTML content** from a web page
- BeautifulSoup **library** to **parse and extract information** from the **HTML content**



Web Scrapping - Libraries

- **Matplotlib library to create static, animated, and interactive visualizations**



Web Scrapping - Libraries

- Matplotlib **library to create static, animated, and interactive visualizations**
- Pandas **library to provide data structures and data analysis tools**



Web Scrapping - Libraries

- **Matplotlib library to create static, animated, and interactive visualizations**
- **Pandas library to provide data structures and data analysis tools**
- **Xlsxwriter library is used for creating and formatting Excel files**



Web Scrapping - Libraries

- Datetime library handles date operations like parsing strings into date objects



Web Scrapping - Example

- For this tutorial, we will extract data from the spoken Tutorial statistics webpage

<https://spoken-tutorial.org/statistics/training/>



Web Scrapping - Example

- **Data analysis is done with workshops conducted between 2022 and 2023 on certain software**



Web Scrapping - Example

- Data analysis is done with workshops conducted between 2022 and 2023 on certain software
- Data such as State, City, Institution, Department, Organizer, Date and Participants are handled



Applications of Web Scraping



Applications of Web Scrapping

- **Price Monitoring:**



Applications of Web Scraping

- **Price Monitoring:**
 - **E-commerce websites scrape their competitor websites**



Applications of Web Scraping

- **Price Monitoring:**
 - E-commerce websites scrape their competitor websites
 - They monitor their prices and adjust theirs accordingly



Applications of Web Scrapping

- **Academic Research:**



Applications of Web Scraping

- **Academic Research:**
 - Researchers can scrape data from academic journals and websites



Applications of Web Scrapping

- **Academic Research:**
 - Researchers can scrape data from academic journals and websites
 - They collect information for studies and research



Applications of Web Scrapping

- **Financial Data Analysis:**



Applications of Web Scrapping

- **Financial Data Analysis:**
 - Analysts use web scrapping to collect data from financial websites



Applications of Web Scraping

- **Financial Data Analysis:**
 - Analysts use web scraping to collect data from financial websites
 - They analyze stock prices, market trends etc.



Summary

In this tutorial, we have learnt to

- **Extract data from websites**
- **Save data to a CSV file**
- **Perform basic data analysis**
- **Generate visualizations**



Assignment

As an assignment, please do the following:

- **Extract the workshop data using different foss filters, start and end date.**
- **Write the data to an Excel sheet**



About the Spoken Tutorial Project

- Watch the video available at http://spoken-tutorial.org/What_is_a_Spoken_Tutorial
- It summarises the Spoken Tutorial project
- If you do not have good bandwidth, you can download and watch it



Spoken Tutorial Workshops

The Spoken Tutorial Project Team

- Conducts workshops using spoken tutorials
- Gives certificates to those who pass an online test
- For more details, please write to contact@spoken-tutorial.org



Answers for THIS spoken tutorial

- Questions in THIS Spoken Tutorial
- Visit <https://forums.spoken-tutorial.org>
- Choose the minute and second where you have the question
- Explain your question briefly
- The Spoken Tutorial project will ensure an answer
- You will have to register to ask questions



FOSSEE Forum

- For any general or technical questions on [Python for Automation](#), visit the FOSSEE forum and post your question <https://forums.fossee.in/>



Acknowledgement

- The Spoken Tutorial project was established by Ministry of Education, Govt. of India



Thank you

- This is Sai Sathwik, a FOSSEE Semester Long Intern 2024, IIT Bombay

