TECHNICAL SEMINAR
ON
WEB SCRAPING

26.03.2025

Under the Guidance

Mrs. M. Anusha

Assistant Professor,
Department of CSE

Presentation by

Ms .Srilekha Bayyapu (21C11A05G0)

Contents

- Abstract
- Introduction
- History
- Seminar Literature
- Applications
- Advantages
- Disadvantages
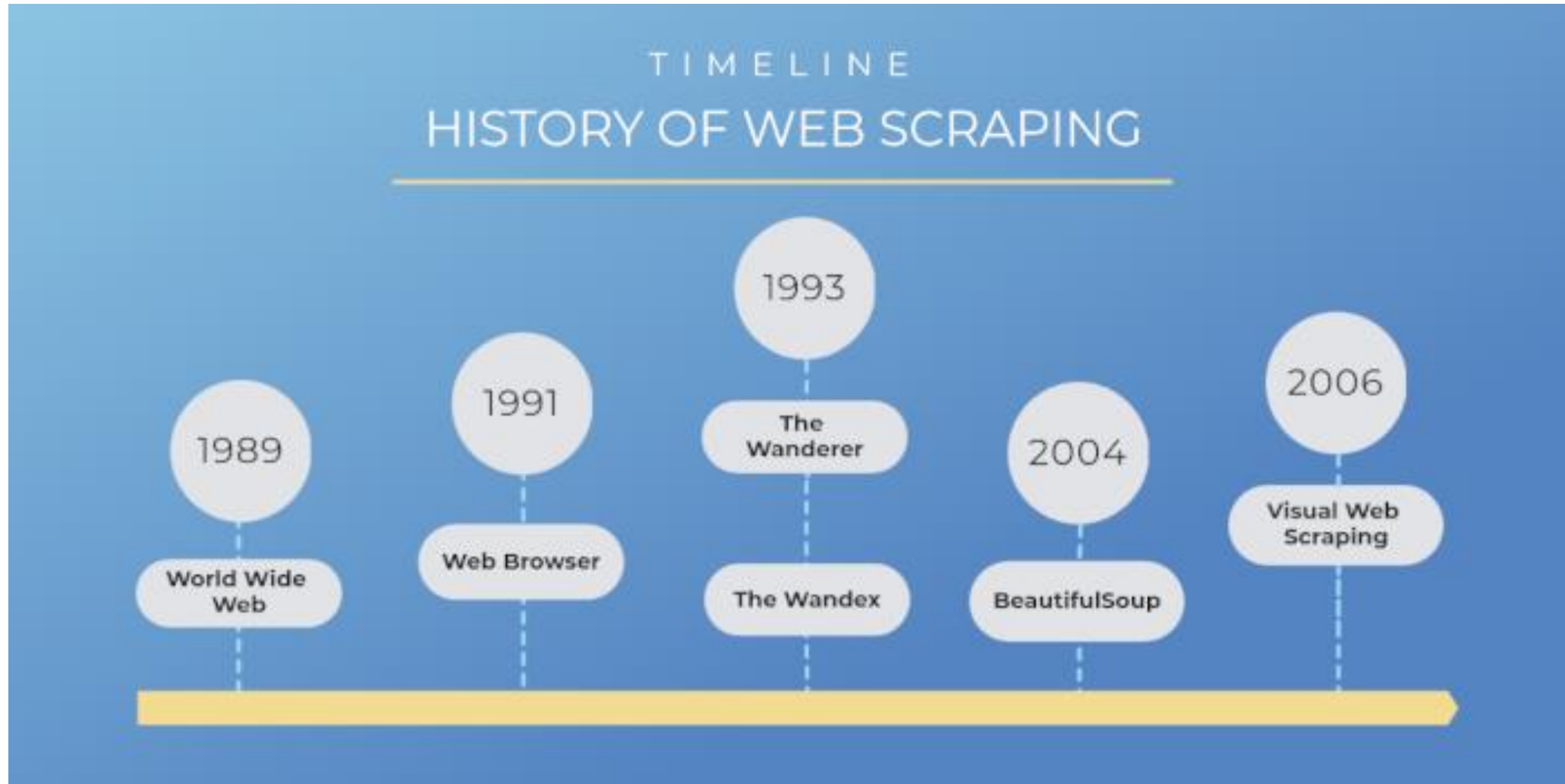- Conclusion
- References

Abstract

Web Scraping is the technique which allows user to fetch data from the World Wide Web. It gives a brief introduction to Web Scraping covering different technologies available and methods to prevent a website from getting scraped. Currently available software tools available for web scraping are also listed with a brief description. Web Scraping is explained with a practical example.

Introduction

- Web Scraping is a technique to fetch the data from websites .while surfing on the web, Many websites don't allow the user to save the data personally use.

- One way is to manually copy-paste the data, which both tedious and time-consuming.

- Web scraping is the automation of the data extraction process from websites .The event is done with the help of web scraping software known as web scrapers.

- They automatically load and extract data from the websites base on user requirements These can be custom built to work for one site or can be configured to work with any website.

History

Seminar Literature

Title: A framework for Web Scraping with Python

Author: Ryan Mitchell

➢ He has expensive experience in software developments, data analysis ,and web scraping

His focuses on making complex technical topics accessible to a broad audience

Title: Web scraping with python and Beautiful soup

Author: Kathryn Hodge

➢ He has worked with various organizations on web scraping and data analysis projects, helping them extract insights from large datasets

Title: web scraping A beginner's Guide and scraping the web

Author: Richard Lawson

➢He has written several books on web scraping, and developed many projects including web scraping , Data analysis and software development

Applications

❖Market Research and Competitive Analysis

❖Social Media Data Extraction

❖Job Listings

❖Inventory Management

❖Price comparison

❖Travel Trend Analysis

Advantages

1. Reduced manual labor

2. Increased efficiency

3. Cost-effective

4. Identifying business opportunities

5.Helps in market Research

Disadvantages

1. Complexity

2. Data quality issues

3. Scalability

4. Website changes

5. Anti-scraping measures

Conclusion

- Web scraping process, and most of the web scrapers are often quite similar and general in nature designed to carry out generic and easy jobs. By comparing the performance and features of different tools and frameworks, we found that Scrapy provides better results as it is fast, extensible, and powerful. Since Scrapy handles requests asynchronously, the results can be scraped rapidly. Further more Scrapy's architecture is based on a web crawler which enables easy data extraction.

Department of Computer Science and Engineering

Reference

- [1] Gunawan, R. et al. (2019). Comparison of web scraping techniques: regular expression, HTML DOM and Xpath. In International Conference on Industrial Enterprise and System Engineering, 2:283-287.

- [2] Sirisuriya, D. S. (2015). A comparative study on web scraping. In the Proc. 8th Int. Res. Conf. KDU, 135– 140.

- [3] Spangher, A. and May, J. (2021). A Web Application for Consuming and Annotating Legal Discourse Learning. arXiv preprint arXiv:2104.10263.

Department of Computer Science and Engineering