

Price Monitoring Platform

State-of-the-art

What the others did:

By using web scraping, information available on multiple websites can be merged into a more convenient format. The data can then be easily visualized, sorted and analysed using data mining.

Price monitoring with web scraping was used for:

- price comparison in online stores in real time/over time
- targeted sales strategy (tracking promotions of competitors)
- alerts on dynamic price changes on price modifications
- updates about changes in state of the market
- monitoring consumers reviews/sentiments about products

Some commonly used tools by price monitoring platform are:

Jaunt - Java library for web-scraping, web-automation and JSON querying

Selenium + Python

jArvest (Java web harvesting library) - web scraping framework

Scrapy - web scraping framework

Using web scraping we can gather and sort information about society

Techniques and methodologies used

1. Manually **copy and paste data from a web** page into a text file or spreadsheet:

- + the simplest solution
- + when the websites have barriers to prevent automated scraping
- time/information
- errors

2. Send a **HTTP** GET request to the site URL to get the **HTML** code and then use **regular expression** on the raw information from the web and keep the data that matches the pattern:

- + easy
- + time
- obfuscation
- if the content is displayed dynamically, the HTML code doesn't contain the needed informations (it is populated at runtime)

3. Build the **DOM** (Document Object Model) and execute the scripts located in the HTML source code:

- + resolves the problem with the dynamic content

4. **Computer vision** web-page analysis - there are attempts to identify and extract information from web pages by interpreting pages visually as a human being would.

5. Vertical aggregation:

- bots can be detected (excessive traffic / CAPTCHA)

6. Semantic annotation recognition

Evaluation methods

Many papers compare web scraping methods using metrics that evaluate quantity like:

- time (for the process to be completed)
- resources (CPU power and memory)
- cyclomatic complexity (number of independent paths through a piece of code)
- lines of code
- depth of inheritance
- number of error messages thrown

Results and research

Today, some sites have many solutions to stop scraping. A reliable web scraper should have CAPTCHA solutions, be able to scrap anonymously and be fast. Thus, research can be done in order to best imitate human behavior on web pages and to extract information from all of them.

There are many solutions for web scraping, that also help build marketing strategies:

- Mozenda
- Import.io
- Octoparse
- Data Crops
- Prsync
- Omnia Dynamic Pricing
- Price2Spy
- Repricer
- Minderest

Important names in the field

- **World Wide Web Wanderer (1993)** - first web robot, intended to measure the size of the web
- **Python Beautiful Soup (2004)** - library designed by Leonard Richardson for Python - helps pulling data out of HTML container
- **Visual web scraping software (2005-2006)** - Stefan Andresen and his Kapow Software launched Web Integration Platform version 6.0, something now understood as visual web scraping software, which allows users to simply highlight the content of a web page and structure that data into a usable excel file, or database

Relevant articles and links

1. Reasons to use price monitoring software:

<https://www.netrivals.com/resources/guides/price-monitoring-software-reasons-use/>

2. Benefits of using a price monitoring tool:

https://webdata-solutions.com/wp-content/uploads/2015/11/blackbee_whitepaper_Price-monitoring-on-the-Internet.pdf

3. Examples of price tracking apps:

<https://smallbiztrends.com/2018/03/price-tracking-apps.html>,

<https://www.octoparse.com/blog/top-10-price-monitoring-tool>,
<https://www.oberlo.com/blog/25-best-price-comparison-websites>

4. Common mistakes in price monitoring:

<https://priceintelligence.net/en/2020/07/24/7-most-common-mistakes-in-price-monitoring-and-how-to-avoid-them/>

5. Some important elements of a price monitoring tool with Python examples:

<https://blog.scrapinghub.com/2016/11/24/how-to-build-your-own-price-monitoring-tool>

6. Python web scraping examples: <https://www.edureka.co/blog/web-scraping-with-python/>,

<https://www.dataquest.io/blog/web-scraping-tutorial-python/>,

<https://realpython.com/python-web-scraping-practical-introduction/>,

[https://towardsdatascience.com/scraping-multiple-amazon-stores-with-python-5eab811453a](https://towardsdatascience.com/scraping-multiple-amazon-stores-with-python-5eab811453a8)

[8](https://towardsdatascience.com/scraping-multiple-amazon-stores-with-python-5eab811453a8)

Resources and tools available

Open-source python library for scraping::

<https://scrapy.org/>

<https://realpython.com/beautiful-soup-web-scraper-python/>

Base idea

<https://blog.scrapinghub.com/2016/11/24/how-to-build-your-own-price-monitoring-tool>

Data source:

EMAG

Amazon

PCGarage

Other price monitoring tool

<https://www.octoparse.com/blog/top-10-price-monitoring-tool>