

Analyzing the selected web scraping tools

Selected tools: **Selenium, Scrapy**

Scrapy

Advantages:

- Open-source collaborative framework.
- Very fast performance & one of a most powerful library .
- It has built on top of twisted.
- Asynchronous networking framework.
- Uses the non-blocking mechanism while sending the requests to the users.

Key features:

- Scrapy is a portable library and easily Extensible.
- It has built-in support for extracting data from HTML sources using XPath expression and CSS expression.
- Less memory usage.

Disadvantages:

- Documentation is not friendly for beginners.

When considering the Extensibility, the architecture has well designed to customize the middleware to add own custom functionality. So the projects will be more robust and flexible.

Selenium

Selenium is mainly designed to automate tests for Web applications. But it can be used for web scraping also.

Advantages:

- Beginner friendly and very easy to code.

Key Features:

- Selenium can easily work with core JavaScript concepts(DOM)
- It can easily handle AJAX and PJAX requests.

Disadvantages:

- Suitable for projects where the data size is small.

When considering the Extensibility when it is dealing with Core JavaScript featured website, selenium would be the best.