

# Scrapy/Spidermon: Create a CLI to setup spidermon on projects

---

## About me

---

Leonardo Victor da Silva Rodrigues

UNICAMP / Computing Science / 2015 / 2020

Github: [leovictorsr](#)

BRT (Brasília Time) - UTC/GMT -3 hours

[Resume](#)

## Contributions

---

<https://github.com/scrapinghub/spidermon/pull/121>

## Project information

---

**Sub-org: Scrapy spidermon**

**Create a CLI to setup spidermon on projects**

Spidermon is a monitoring tool created to help Scrapy users.

It helps by creating monitors to be checked when Scrapy spiders run, like tests, and perform actions to notify Scrapy users that the spider could have any problem during run. Today an user can be notified via email and Slack, but anyone can create actions to be performed if you have an integration.

Today, Spidermon uses manual configuration to setup projects, which is good for experienced users, but it's not helpful for whom wants to have a quick setup. Also, Scrapy already have this feature, together with a good documentation.

In Spidermon we have 3 concepts: monitors, actions and item validation.

Monitors are the tests created that will validate that a spider run how it was expected to run. Simple monitors are item counts and item fields validation. Spidermon have

built-in monitors, and the CLI is going to allow to enable and configure those monitors..

Actions are performed after a monitor is run. With the result of the monitor, we can perform actions like send a message to an email or to a Slack channel. Spidermon also has built-in actions that the CLI is going to allow to enable and configure.

Validations are performed after a monitor is run. They check if an item is scrapped in conformation of what the user intends to be scrapped. The CLI will enable and create templates for validation schemas, and the user will be able to select from our current 2 (in the future, 3) options of validation method (jsonschema, schematics and cerberus)

My idea for this project is to create a robust CLI for Spidermon setup. A user would be able to enable and configure Spidermon monitors, actions and item validations, and be presented with a detailed help menu upon request.

My vision of usage of this feature can be read in this snippet:

<https://github.com/leovictorsr/gsoc-2019/blob/master/spidermon-cli-usage.md>

I will use the [Click](#) library for the implementation of the CLI, as Click is highly customizable and easy to use and extend, allowing me to have a good, well-documented interface using solid design principles quickly.

## **Timeline**

### **Community Bonding (May 7-26)**

During this time I'll be discussing with my mentors and interested contributors the design and implementation of the tool.

I will also continue to review and improve the documentation in areas I'll be contributing to.

### **Coding period**

I plan on incrementally adding features to the overall tool. I have scheduled the implementation, documentation, and testing for each individual piece in periods of 3 weeks. Every 3 weeks, I will have a deliverable of one subset of commands implemented. This does not mean I will have my work for review only every 3 weeks; I intend to have smaller deliveries every week for code, test, and documentation, and those 3 will compose the command deliverable while iterating with my mentors.

At the end of each period, I will have unit tests, documentation and code for the command to configure monitors. Every week will be used to interact and evaluate the code with the mentors in weekly online meetings that we can schedule, or by email.

1. Week 1 (May 27) to Week 3 (June 10) - 3 weeks
  - Implement monitors configuration
2. Week 4 (June 17) to Week 6 (July 1) - 3 weeks
  - Implement validation configuration
3. Week 7 (July 8) to Week 9 (July 22) - 3 weeks
  - Implement actions configuration
4. Week 10 (July 29) to Week 12 (August 12) - 3 weeks
  - I will use these weeks to have a good margin for possible delays and adjustments that weren't done during the previous weeks. I intend to have the CLI delivered before this period, where with the spare time I intend to continue to help with another tasks inside Spidermon.

I will commit to have at least one good feedback or report every week. This way, I can track any problems that I may have and ask help to my mentors for any questions I should have

I will commit to blog my experience in my personal blog hosted on Github Pages.

## Other commitments

---

- During the weeks 4 and 5 I'll be busy with the exams from university, so I'll lesser the workload these weeks by 3/4 of a regular. This shouldn't delay me though, because I can use my bigger 3-week periods to recover the delayed time during these weeks as after them I will be free from any university activities; also, I have a good 3 weeks period to have any delays caught up.