# Google Summer of Code 2019

Project Proposal

# Python Software Foundation

## Scrapy | Integrate Cerberus

*Integrating the powerful data validation package, Cerberus into Spidermon*

## About You

- **Name** - Vipul Gupta
- **Email address** - vipulgupta2048@gmail.com
- **Mobile** - ************
- **University -** Amity University, Bachelor of Technology,
  Computer Science and Engineering,
  3rd year (Junior), 2016-2020
- **First language** - English (Fluent)
- **Time Zone** - New Delhi, India (+5.30 GMT)
- **Blog Url -** Mixster
- **IRC nickname** - vipulgupta2048
- **Linkedin -** vipulgupta2048
- **Github -** vipulgupta2048
- **Twitter -** vipulgupta2048

For any contact, I would recommend reaching me at my email address, would be happy to help you in any way possible.

# Code Contribution

## Pull Request Opened:

[#142] Improve Slack Action Documentation

## Reviewed:

[#138] Provide API documentation for render_text_template and render_template

[#133] Add Mixin Docs

## Issues Created/Resolved/Participated:

[#140] `schematics` required but not installed by default - **Solved**

[#135] Document use of custom templates - In progress,

[#128] Action to restart spider in Scrapy Cloud

[#125] Add default template to email actions

[#107] Follow python end of life https://devguide.python.org/

[#95] Enhance Slack notifications - In progress

[#47] Option to stop spiders

And many more.

# Project Information

**Sub-org name** - Scrapy

## Why Cerberus?

There are several great reasons for including Cerberus as an option for users of Spidermon. I feel some are as follows.

- As Raph once mentioned, one of the primary uses for Spidermon is data validation. Having another great tool such as Cerberus would only be beneficial to the entire package.
- Cerberus has a wide range of support both from the community and its sponsors, leading to believe its usefulness, people’s trust and power in allowing for custom validation. Which is what we need.
- The official documentation makes note of another great advantage that isn’t present with jsonschema and schematics. Which is having no dependencies, effectively reducing dependencies issues that we have experienced in the community quite recently through [1] [2]
- Cerberus is thoroughly tested from Python 2.6 up to 3.5, PyPy and PyPy3 ensuring life long support and maintenance.

**I think all these features make Cerberus a vital inclusion that we should consider having in Spidermon as one of the ways data validation**

## Objective

To integrate include Cerberus as a new option for item validation available for the user in Spidermon. and PyPy

## Technologies and Skills Used

This list is just a small glimpse of technologies that I am thinking to use for this project. I am proficient in each of them, through my past projects. (Refer)

1. **Programming Language:** Python 3.6
   - Using packages such as Scrapy, Cerberus along with OOPS concepts
2. **Markup Languages** - Markdown, reStructuredText,and Sphinx for documentation generation.
3. **Python Packaging** - Setuptools, Virtualenv
4. **Testing and code styling** -Pytest, Unitest, Black
5. **Editors -** VScode, Geany, Nano.
6. **Version Control** - Git, GitHub

## Deliverables

1. The code will be off the **highest quality standards** having detailed documentation, black styling and will be well tested.
2. A detailed, well documented tutorial will be developed during the course of the summers implementing almost every feature of Spidermon to help developers as a reference and blogs will be written.
3. One blog **each week** (Total of **15 and more**) blogs regarding Scrapy, Spidermon and my experience, learning through the project on Mixster.
4. For the community to track progress, I will maintain a tracker with my latest developments containing week-to-week updates, and MoM of mentor meetings. This helps to maintain **accountability**, **transparency** and well, **keeping track**.

**Thus, by the deliverables mentioned above, 3 well-defined goals that determine project status which will help in the evaluation are as follows**

- **Evaluation 1**: Cerberus Integration 40% complete. Prepwork done, documentation and command needed finalised.

- **Evaluation 2**: Cerberus integration 80% complete. Alpha testing starts.

- **Final**: Cerberus ready for deployment, available for all platforms, documentation and beta testing complete. Tutorial and blogs ready for reference.

### Other Deliverables

1. **Contribute more** - After and during the project, I will help in **solving bigger challenges** and **bugs** faced in other Scrapy projects, wherever my help is needed by the community.
2. **Complete Misc. Tasks -** There are several miscellaneous non-code tasks that I would like to take up in the favor of giving back to the community such as mentoring and representing the same in global conferences.
3. **Remain** as an active contributor to Scrapy and Python Software Foundation, taking part in discussions for the future projects, implementations, and bugs. Maybe if the community agrees, also mentor someone else in the foreseeable future.

## Timeline

In accordance with the official timeline by Google, my college exams end somewhere in the first week of May or so. My college reopens early in July. During this time, I will remain in contact with my mentor.

This gives me ample time to code. I plan to begin coding before the official period coding period of GSoC starts from the starting of May. Which gives me a head start to finish the majority (70%) of the project before the second evaluation.

**Providing a total of 10 weeks to write code for integrating Cerberus (& other deliverables)**

I will also write a weekly or Bi-weekly blog post on my progress and updates on my project and post it diligently on the website to get the community involved.

The weekly timeline is as follows.

| **Period** | **Task** |
|---|---|
| May 6 – May 26<br>**Community Bonding Period** | ● Community Bonding, discussion of feature list that would be created, button down on evaluation goals with mentors.<br>● Research **Part 1**<br>○ Find ways to implement Cerberus the best way into Spidermon.<br>○ Study installation procedures and source code of Cerberus<br>○ Blog post on Community Bonding @ Scrapy<br>○ Setup development environment<br>○ Discuss results with the mentor, shortlist alternative methods that have a higher success rate.<br>○ Go through other validator's documentations (jsonschema & schematics |
| [#1] May 27 – June 17<br>**Student Coding + First evaluations (Till June 23)** | ● Official Coding period starts<br>● Write code for integration of Cerberus, and figuring out a way how would users add their own validation schema and pass it through validator classes.<br>● Extensive implementation testing<br>● Improving documentation of other validators side by side.<br>● Get code reviewed by mentors for the first evaluation. |
| [#2] June 24 - July 22<br>**Student Coding + 2nd evaluations (Till June 23)** | ● Work continues; Tweaking; suggestions implemented from community<br>● Fix problems with the integration and development<br>● Write relevant documentation and test to full proof integration<br>● Write more code for completing integration<br>● Alpha testing period end, Beta begins. |
| [#3] July 23 - August 24<br>**Final evaluations** | ● Cerberus ready for deployment, available for all platforms,<br>● Documentation complete.<br>● Beta testing complete.<br>● Tutorial and blogs ready for reference. |

## Other commitments

- List of any things that might affect your ability to work this summer.
  **Answer: None**

  - **List any exams** - My exams end on May 15, which works out just fine.
  - **Classes, Weddings, other jobs** - One, 24th June
  - **Holidays** - Maybe, but with my past experience with Google Summer of Code. I think I can handle it well. I usually inform mentors 2-3 weeks in advance, add changes to my timeline accordingly and see to that the work is done before the said period.

- **If you're applying to more than one organization, you can let us know which one you prefer in case of a tie.**
  - **None**, Scrapy is my first choice.
  - I have another proposal in the same org. For Developing the Spidermon Command line interface. I have put in a lot of thought and work into that. I would like to have that as my **first choice.**