Draft Ideas for Google Summer of Code 2014

These are ideas that didn't make it to the list of final list for GSoC 2014, either because they need to be polished (or better specified) or because the scope wasn't suitable for 3 months of work.

Contents

Your own idea

Please submit your own idea following this template, keeping in mind these guidelines.

Brief explanation	
Expected results	
Required skills	
Difficulty level	Easy, Intermediate or Advanced
Mentor(s)	

Better generator support

TODO:
    This one should be better specified.

Brief explanation	Improve Scrapy API using generators
Expected results	Scrapy should provide an easy way to build a single item from several pages, ...
Required skills	Python, general understanding of async code, API design
Mentor(s)	Mikhail Korobov, Rolando Espinoza

There are areas where Scrapy usability and efficiency can be improved by using generators, for example:

- Integrate something like Rolando's https://github.com/darkrho/scrapy-inline-requests;
- ensure generators are not exhausted needlessly in various places;
- provide an easier alternative to spider_idle signal, something in line with https://github.com/scrapy/scrapy/issues/456
- ...

Reading list:

- http://www.python.org/dev/peps/pep-0342/
- http://www.tornadoweb.org/en/stable/gen.html
- http://twistedmatrix.com/documents/13.0.0/core/howto/defer.html
- https://twistedmatrix.com/trac/wiki/DeferredGenerator

Improve javascript integration

Brief explanation	Improve Javascript integration by using Splash to render and execute Javascript.
Expected results	A Scrapy middleware to integrate with Splash
Required skills	Scrapy
Mentor(s)	Mikhail Korobov, Daniel GraÃ±a

Multi-platform Scrapy GUI for running spiders

Brief explanation	Develop a multi-platform GUI interface for running Scrapy spiders.
Expected results	This interface is a companion to Scrapyd and must use (and possibly extend) the Scrapyd API. Basic features: schedule spider jobs (start, stop, pause), view/search items, view/filter/search logs, export items/logs.
Required skills	Multi-platform Python GUI development.

Difficulty level	Intermediate
Mentor(s)	

Building Social Spiders to better Crawl Social Networks

Brief explanation	Build Scrapy Spiders able to better extract content from Social Networks (target networks: Facebook, Google+, Twitter, Youtube and others if time suits us)
Expected results	Scrapy Spider classes easily put in use of the developer to crawl social networks
Required skills	Python Programming - Regular Expressions Background
Difficulty level	Intermediate
Mentor(s)	