

crawling

Framework

- [puppeteer](https://github.com/puppeteer/puppeteer) : <https://github.com/puppeteer/puppeteer>
- [Crawlab](https://github.com/crawlab-team/crawlab) : <https://github.com/crawlab-team/crawlab>
- [ScrapydWeb](https://github.com/my8100/scrapydw) : <https://github.com/my8100/scrapydw>
- [Gerapy](https://github.com/Gerapy/Gerapy) : <https://github.com/Gerapy/Gerapy>

Ref

- <https://tiktikeuro.tistory.com/174>
- <https://tiktikeuro.tistory.com/171>

Framework	Technology	Pros	Cons	Github Stats
Crawlab	Golang + Vue	Not limited to Scrapy, available for all programming languages and frameworks. Beautiful UI interface. Naturally support distributed spiders. Support spider management, task management, cron job, result export, analytics, notifications, configurable spiders, online code editor, etc.	Not yet support spider versioning	
ScrapydWeb	Python Flask + Vue	Beautiful UI interface, built-in Scrapy log parser, stats and graphs for task execution, support node management, cron job, mail notification, mobile. Full-feature spider management platform.	Not support spiders other than Scrapy. Limited performance because of Python Flask backend.	
Gerapy	Python Django + Vue	Gerapy is built by web crawler guru Germey Cui. Simple installation and deployment. Beautiful UI interface. Support node management, code edit, configurable crawl rules, etc.	Again not support spiders other than Scrapy. A lot of bugs based on user feedback in v1.0. Look forward to improvement in v2.0	
SpiderKeeper	Python Flask	Open-source Scrapyhub. Concise and simple UI interface. Support cron job.	Perhaps too simplified, not support pagination, not support node management, not support spiders other than Scrapy.	

Plugin Backlinks:

From:
<https://jace.link/> - **Various Ways**

Permanent link:
<https://jace.link/open/crawling>

Last update: **2021/06/15 01:30**

