

Crawling healthy



MIKHAIL KAVALIOU

January, 2020



Mikhail Kavalious

Senior Java Developer

6+ years in Java
3+ years in EPAM

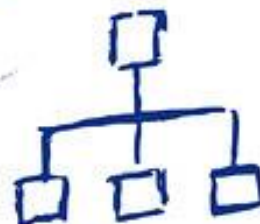


AGENDA

- **SEO**
- **Crawling**
- **Scrapy**
- **Cloud Crawling**



SITE
ARCHITECTURE



✓ ONPAGE
OFFPAGE

SEO

CONTENT



Search Engine Optimization

TRAFFIC

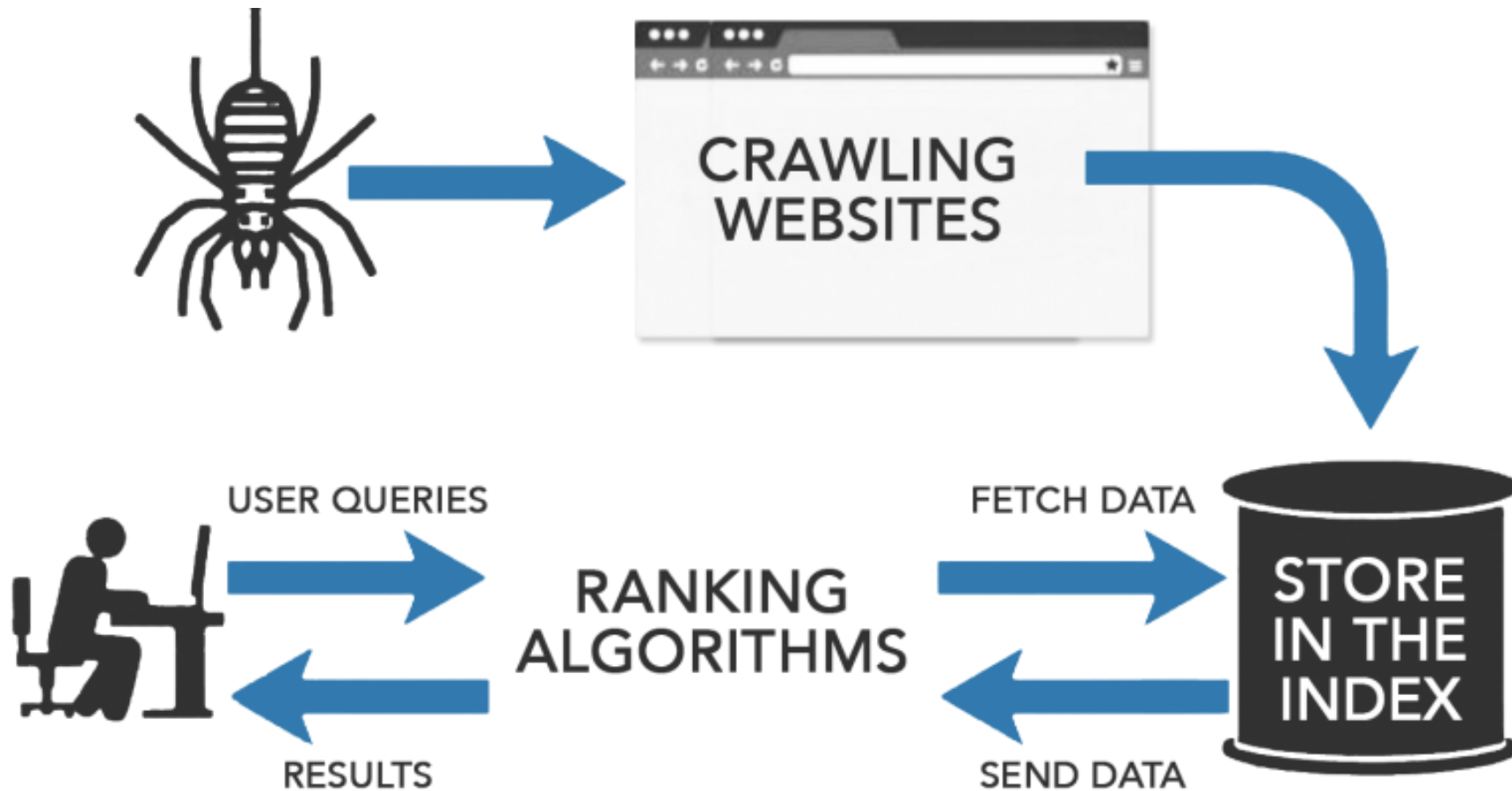
★★★★
RANKING



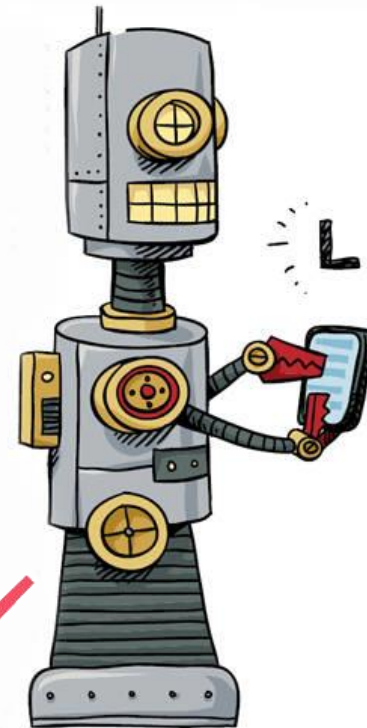
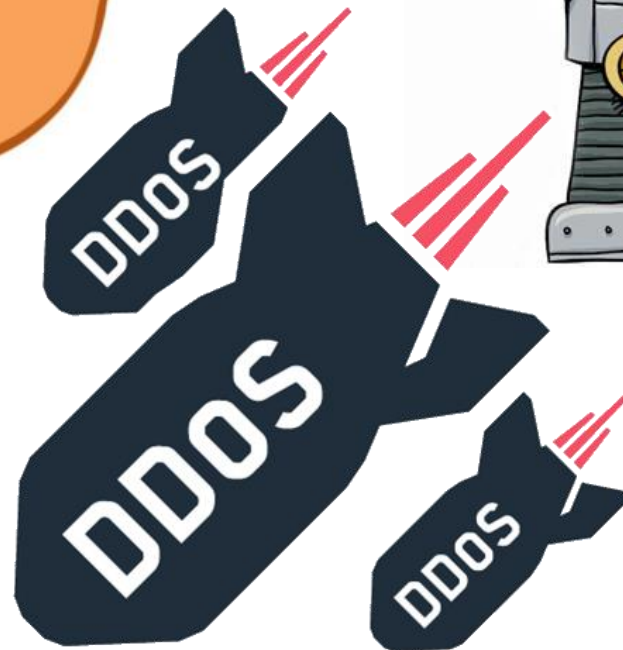
INDEXING
KEYWORDS

◀◀◀◀...
BACKLINKS

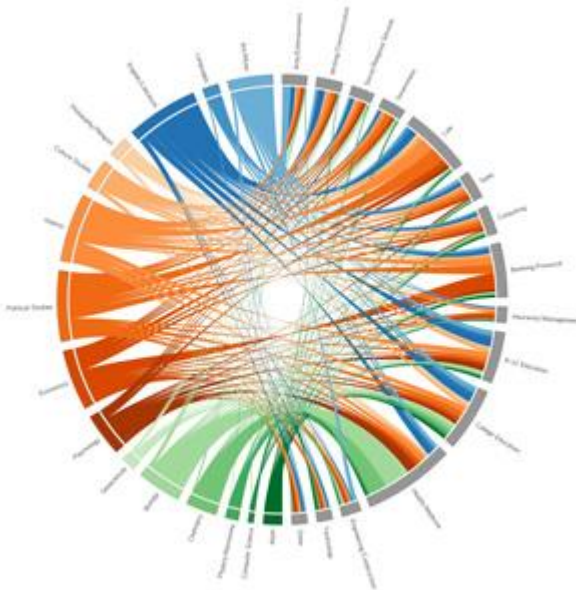
Что такое crawling?



Для чего не стоит использовать crawling?



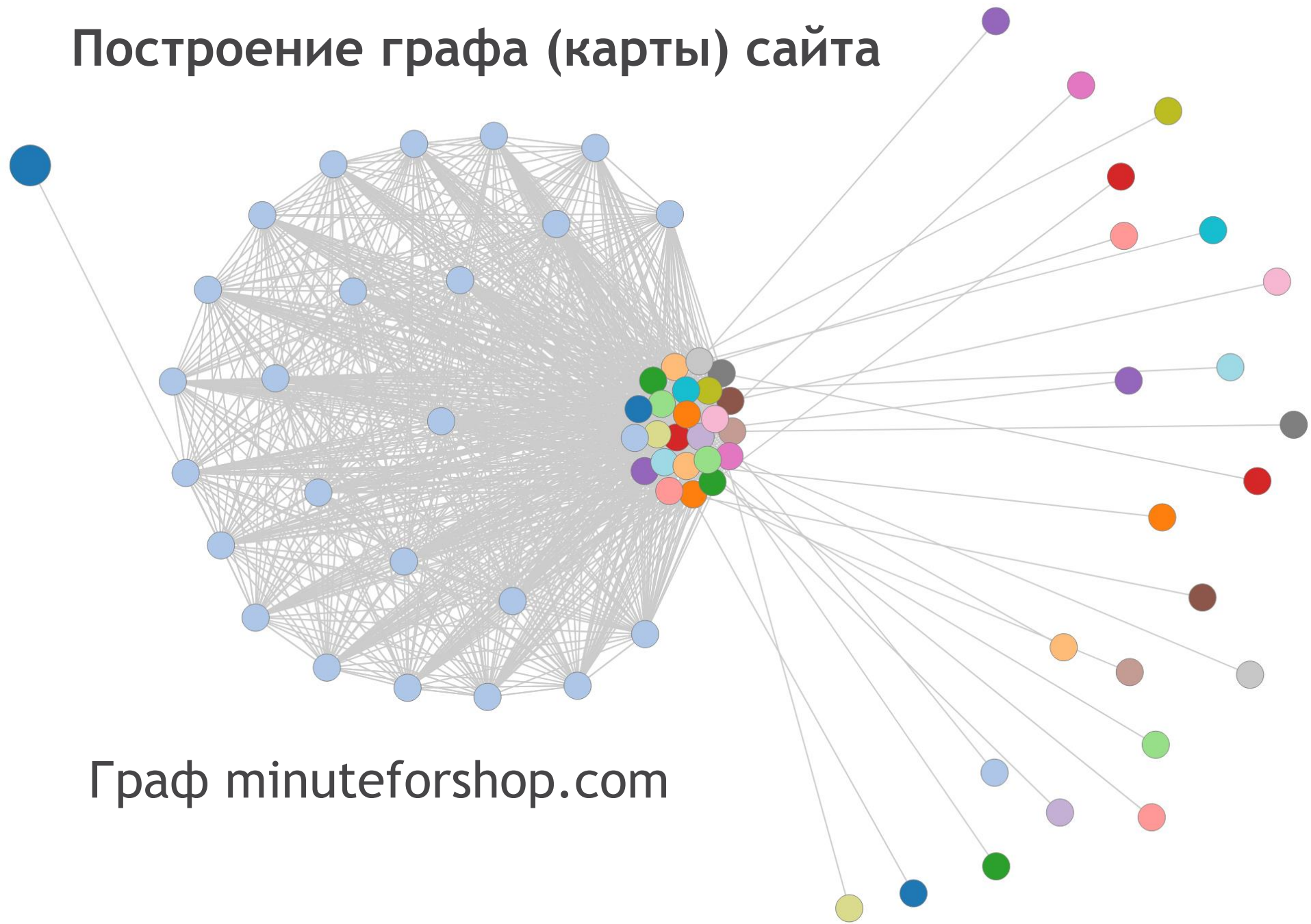
Для чего нужен crawling?



Предсказание результата



Построение графа (карты) сайта



Граф minuteforshop.com

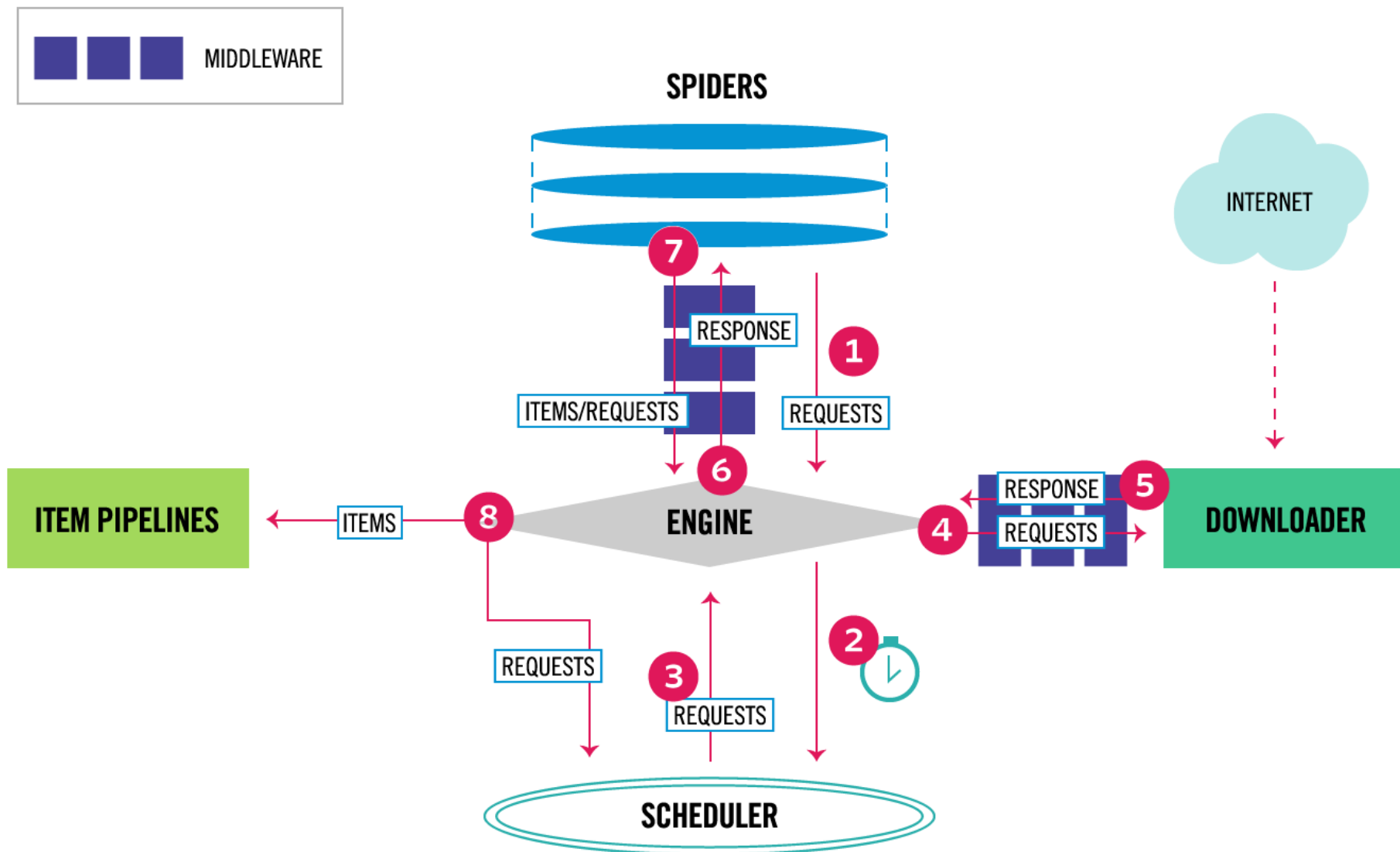
Решение технических проблем



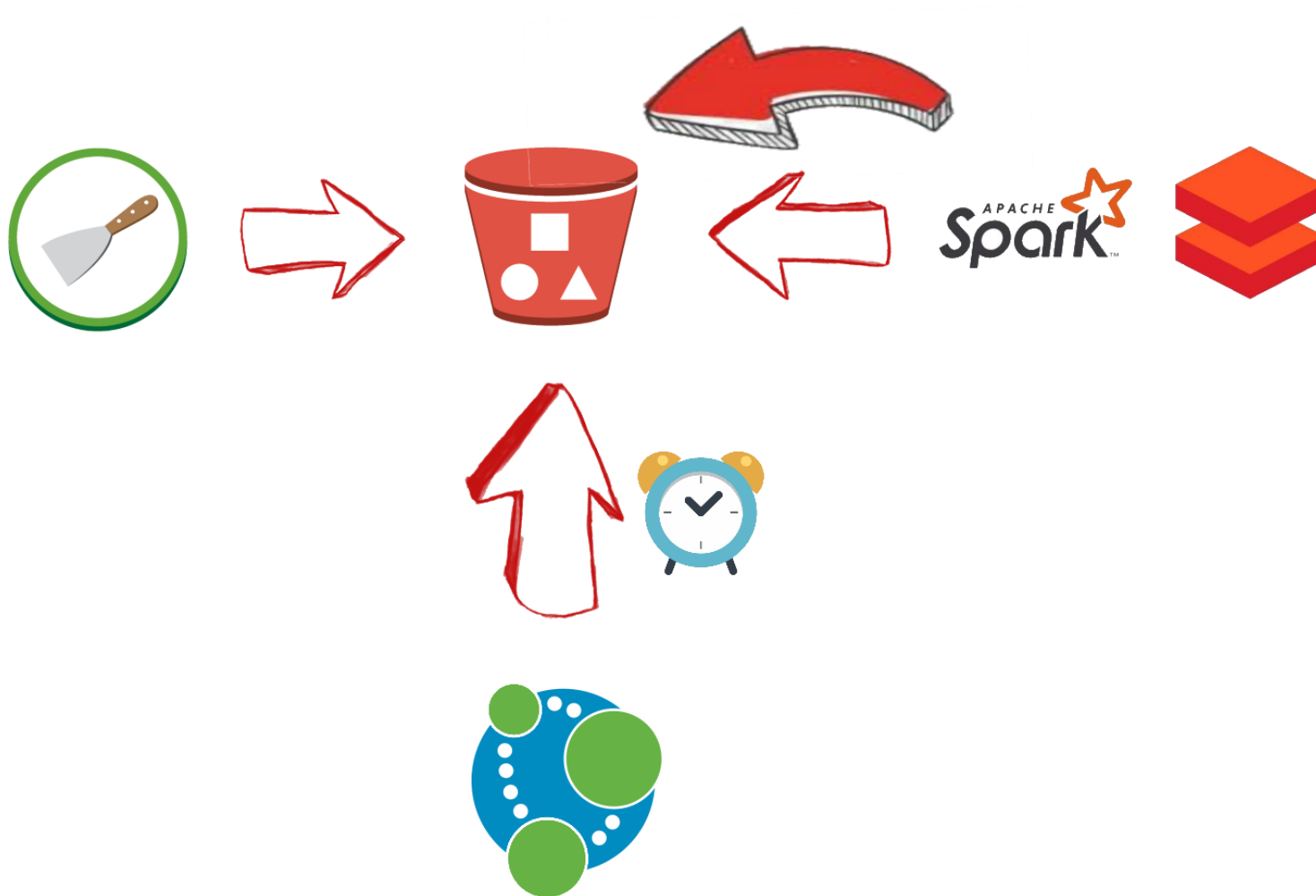
Выбор лучшей библиотеки



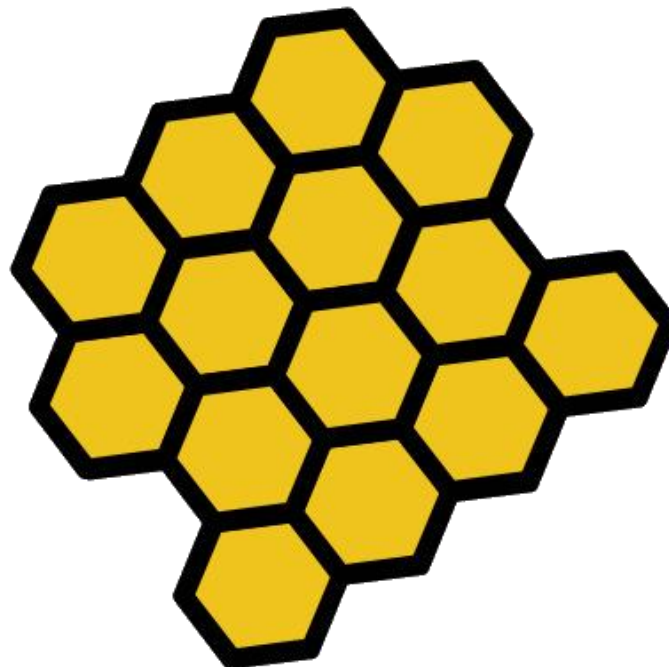
Как устроен Scrapy?



Архитектура кроулинга



Гибкость настройки и кастомизации



Запуск crawling процесса

crawl my-spider

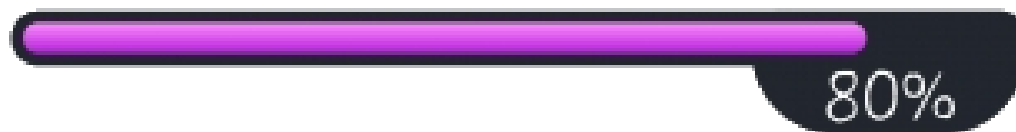
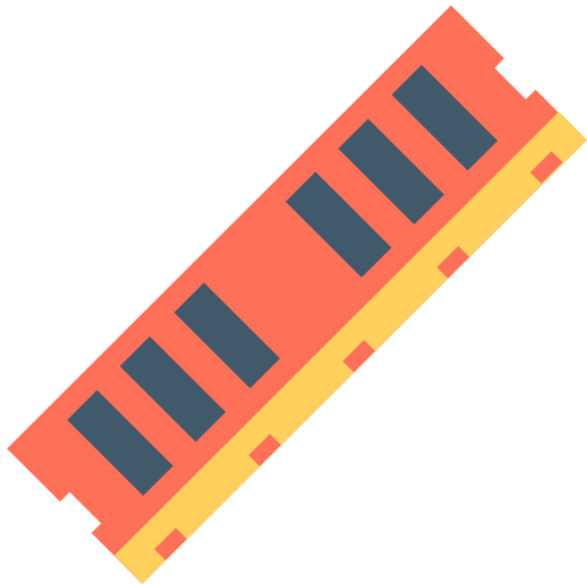
-a domains="[\"example.com\"]"

-a urls="[\"http://example.com/sitemap.xml\"]"

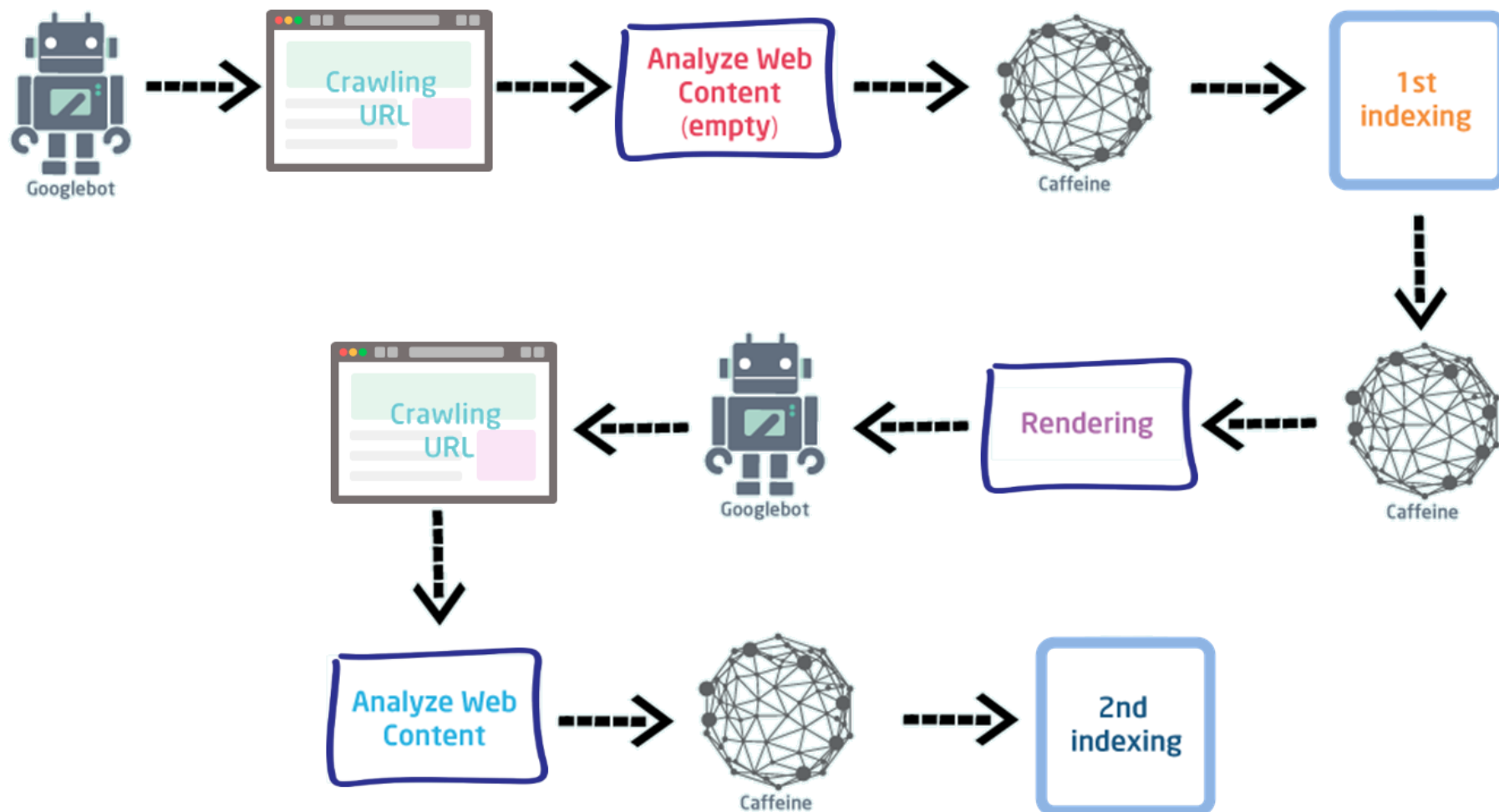
-s ROBOTSTXT_OBEY=True



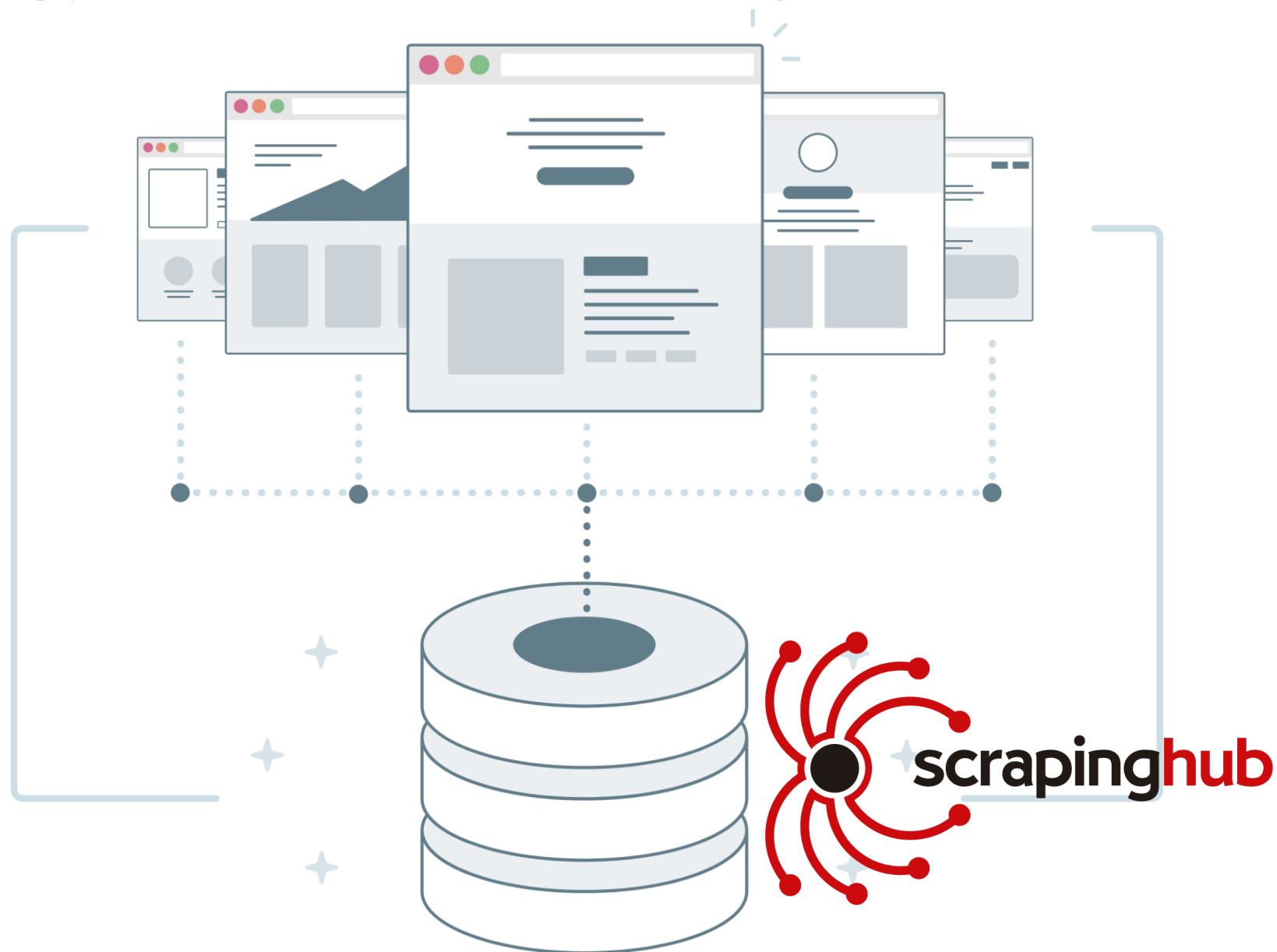
Очереди и состояние



Рендеринг JavaScript

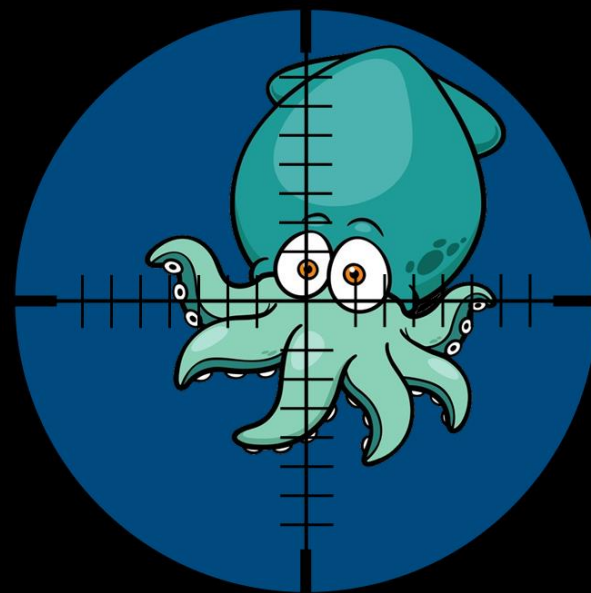


Scrapyhub и облачный crawling



ИТОГ

**Crawler - инструмент для бизнеса
и не только, имеющий множество
применений.**



USED RESOURCES

- https://ru.wikipedia.org/wiki/Поисковая_оптимизация
- https://ru.wikipedia.org/wiki/Поисковый_робот
- <https://docs.scrapy.org/en/latest/>
- <https://scrapinghub.com>

- <https://github.com/MikhailKavaliou/CrawlingHealthy>



Any
1.7
Questions