

Webscraper-server

Рангел Плачков

1MI0800061

KN2 група 6

Документация

Повреме на разработка проекта беше разделен на множество по-малки автономни и поради това нямаше сериозни проблеми, върху които съм отделим голямо време, единствено голямо количество малки.

1. Клиент сървър приложение. Много подобно на чат. Остановява връзка между калсовете Client - Server чрез сокети. След това беше добавена възможността да се свързват повече от един клиенти наведнъж. Приложението не е направено да има никаква Client - Client комуникация, тъй че всеки е независим.
2. Обработка на стринговете подадени от клиента. Парсер в намиращ се в Command.java, който рабива стринга на команда и аргументи.
3. Executor е клас общо взето дълъг if-else if който разпознава команда (String) и повиква съответния метод.
4. Image И Html scraper оригинално направени ,като съвсем отделни java проекти и когато бяха готови им направих да се управляват main-а им чрез args[]. Двете приложения имат много подобен код. И са преминали през две версии -> една която теглеше файлове само от адреса който си им дал и втора ,която рекурсивно извиква себеси за списъка от линкове в сайта си.
5. Multithread използвайки "implements Runnable".Главният процес преминава през различни линкове и на всяка стъпка и дава на нов Thread да тегли снимките или HTML-а докато той продължава надолу.

Single vs Multithreading

Повреме на забелязах че има осезаема разлика използвайки multithread. Засякох времето за което <scrapeHTML <https://www.wikipedia.org/> 1>

С Multithread -> 2мин 14 сек.

Без Multithread -> 3мин 37 сек.