

README.md

Задание

Ycrawler

Задание: написать асинхронный краулер для новостного сайта news.ycombinator.com:

- краулер начинает обкачивать сайт с корня news.ycombinator.com/
- краулер должен обкачивать топ новостей, т.е. первые 30 на корневой станции, после чего ждать появления новых новостей в топе, которых он раньше не видел
- для того, чтобы "скачать" новость нужно скачать непосредственно страницу на которую она ведет и загрузить все страницы по ссылкам в комментариях к новости
- внутри скачанных страниц дальше новые ссылки искать не нужно
- скачанная новость со всеми скачанными страницами из комментариев должна лежать в отдельной папке на диске
- разрешается использовать стандартную библиотеку и aiohttp.

Цель задания: поближе познакомиться с асинхронным программированием, получить опыт написания и отладки асинхронных программ.

Критерии успеха: задание **обязательно**, критерием успеха является работающий согласно заданию код, для которого проверено соответствие pep8, написана документация с примерами запуска, в README, например. Далее успешность определяется code review.

Deadline

Задание нужно сдать через неделю. То есть ДЗ, выданное в понедельник, нужно сдать до следующего занятия в понедельник. Код, отправленный на ревью в это время, рассматривается в первом приоритете. Нарушение дедлайна (пока) не карается, но может повлиять на ранжирование при выборе топа студентов при окончании курса, пытаться сдать ДЗ можно до конца курса. Но код, отправленный с опозданием, когда по плану предполагается работа над более актуальным ДЗ, будет рассматриваться в более низком приоритете без гарантий по высокой скорости проверки

Обратная связь

Студент коммитит все необходимое в свой github/gitlab репозиторий. Далее необходимо зайти в ЛК, найти занятие, ДЗ по которому выполнялось, нажать "Чат с преподавателем" и отправить ссылку. После этого ревью и общение на тему ДЗ будет происходить в рамках этого чата.