ТЗ на разработку парсера постов в Инстаграмм

Источник: Instagram

Результат работ: парсер в виде скрипта на python, запакованный в docker контейнер, и протестированный на работтоспособность на машине EC2 AWS. Код должен предусматривать обход блокировок Instagramm.

Технологии: : Желательно использовать Selenium, (но возможны и другие варианты, по предложения исполнителя)

Требования к парсеру:

1. Парсер должен принимать на вход список аккаунтов в Инстаграмм которые необходимо скачать.

2. Есть ли возможность выгружать с картинкой строку с описанием того, что на картинке (пока картинка прогружается такую строку можно видеть в браузере)?

3. Парсер должен быть выполнен в докер контейнере, который будет размещен на машине в облаке AWS. Исполнителю необходимо будет проверить чтобы все работало непосредственно в облаке. То есть нужно учесть то что Инстграмм может блокировать конкретные IP от AWS. Парсер должен обеспечивать возможность обхода блокировки.

Расписание и логика работы парсера: Пользователь запускает самостоятельно парсер, т.е. не требуется его работа в автоматическом режиме по расписанию.

Что должен качать парсер:

1. Скачивать нужно в обязательном порядке Фотографию в оригинальном разрешении и timestamp поста, количество лайков, а также все сопутствующую информацию которую выдает Инстаграмм. (Пример того что выкачивали ранее предоставим)
2. Текст поста, Геолокацию, Количество лайков, Комментарии (текст, эмодзи), Хэштеги, Теги, Ники фолоуверов аккаунта и т.п.

В каком виде должен записывать парсер данные (с примером):

1. Текстовые данные в формате json
2. Картинки в jpeg
3. Можно ли качать видео? (уточнить у исполнителя)

Как именовать картинки и текст:

№	Что скачать	Нейминг спарсенных данных
1	Картинка	Название Аккаунта_timestamp, *Пример (привести пример): iri_che_ 1607678575.jpg*
2	Текст	Название Аккаунта_timestamp *Пример (привести пример): iri_che_ 1607678575.json*

Язык нейминга: английский

Куда должен парсер записывать данные: Данные каждого аккаунта должны сохраняться в одну папку(бакет на AWS S3)