Университет ИТМО, факультет программной инженерии и компьютерной техники

Двухнедельная отчётная работа по «Информатике»: аннотация к статье

Дата лекции: \_\_\_\_\_\_\_\_\_\_ Дата сдачи: \_\_\_\_\_\_\_\_\_\_\_

Выполнил(а) Кулаков Н.В , № группы *P3130* , оценка

Фамилия И.О. студента не заполнять

|  |  |  |
| --- | --- | --- |
| **Название статьи/главы книги/видеолекции**  Web scraping вашего сайта: непрошеные гости и как их встречают | | |
| **ФИО автора статьи (или e-mail)**  *Shapelez* | **Дата публикации**  **(не старше 2018 года)**  "29" июля 2020 г. | **Размер статьи**  **(от 400 слов)**  4951 слов |
| **Прямая полная ссылка на источник и сокращённая ссылка (bit.ly, goo.gl, tr.im и т.п.)**  *Прямая ссылка:*  [*https://habr.com/ru/company/qrator/blog/513014/*](https://habr.com/ru/company/qrator/blog/513014/)  *Сокращенная ссылка:* [*https://bit.ly/31ODOQe*](https://bit.ly/31ODOQe) | | |
| **Теги, ключевые слова или словосочетания**  Web scraping, скрейпинг, парсинг, защита, ddos, l7 | | |
| **Перечень фактов, упомянутых в статье**   1. Web scraping нацелен на массовый сбор публичных данных с веб-ресурсов с использованием ботов. 2. Парсить HTML регулярными выражениями – не очень хорошая идея. 3. В 2019 году суд Linkedln к организации, собиравшей публичные данные пользователей, постановил действия скрейпинговой компании легальными; в результате этого прецедента скрейпинг стал считаться более законным, и это положительно повлияло на интерес к данной области. 4. Однажды бот скупил не глядя много пар кроссовок, цена которых была равна 100 тыс. долларов. 5. С помошью скрейпинга можно осуществлять DDOS-атаки. 6. Скрейпер в большинстве своем реализовывается на скриптах. 7. Для предотвращения скрейпинга можно использовать разнообразный арсенал ботов-детекторов, функции которых - анализ, сбор цифровых отпечатков, CAPTCHA | | |
| **Позитивные следствия и/или достоинства описанной в статье технологии (минимум три пункта)**   1. Scraping значительно экономит время поиска необходимой информации в интернете. 2. На скрейпинге можно хорошо заработать, т.к вовремя собранная информация стоит больших денег. 3. С правовой точки зрения Scraping является легальным. 4. Благодаря использованию JS при скрейпинге, мы можем интегрировать сам скрипт в нашу страницу и таким образом обойти проверку сайта. 5. Для решения captcha, которые время от времени появляются на сайтах, можно использовать человеческий труд, а само решение посылать на ввод программе-скрейперу. | | |
| **Негативные следствия и/или недостатки описанной в статье технологии (минимум три пункта)**   1. Злоумышленник может заскрейпить чужой сайт, и таким образом лишить как прибыли и части аудитории самого владельца, так и покупателей денег и времени. 2. Из-за того, что данный метод является относительно быстрым, его можно использовать для скупки лимитированных товаров, а это может помешать приобретению данного продукта реальным покупателям. 3. Ботами-скрейперами можно опустошить инвентарь онлайн-магазинов путем добавления его [инвентаря] в корзину. 4. Программа, использующая scrapping для поиска информации, может сильно нагружать сайт, таким образом мешать его нормальной работе. 5. На основе скрейпинга можно осуществлять взлом путем перебора логинов и паролей. 6. Во время парсинга сайта вас могут вычислить: вам могут ограничить доступ к данному информационному источнику. 7. При скрейпинге мы может столкнуться с проблемами в виде разнообразных проверок и капч. | | |
| **Ваши замечания, пожелания преподавателю *или* анекдот о программистах[[1]](#footnote-1)** | | |

1. Наличие этой графы не влияет на оценку [↑](#footnote-ref-1)