

Список использованных источников

1. О ветеринарии : Закон Российской Федерации от 14 мая 1993 г. №4979-1 : в ред. от 14 июля 2022 г.
2. Кодекс Российской Федерации об административных правонарушениях : 30 декабря 2001 г., N 195-ФЗ : в ред. от 18.03.2023 г., статья 10.8. Нарушение ветеринарно-санитарных правил перевозки, перегона или убоя животных либо правил заготовки, переработки, хранения или реализации продуктов животноводства.
3. Ветеринария и жизнь. Информационный портал и газета ВИЖ [Электронный ресурс] – Режим доступа: <https://vetandlife.ru/sobytiya/iskusstvennyj-intellekt-sistemy-merkurij-vysokocenili-mezhdunarodnye-ehksperty>. – Дата доступа: 21.03.2023.
4. OPSI Observatory of Public Sector Innovation [Электронный ресурс] – Режим доступа: <https://oecd-opsi.org/innovations/artificial-intelligence-reveals-counterfeit-and-falsified-products>. – Дата доступа: 21.03.2023.
5. Россельхознадзор. По материалам Россельхознадзора возбуждено 134 уголовных дела по фактам нарушений при обороте животноводческой продукции [Электронный ресурс] – Режим доступа: <https://fsvps.gov.ru/ru/fsvps/news/214161.html>. – Дата доступа: 22.03.2023.

УДК 004.738.52

РАЗРАБОТКА ВЕБ-ПРИЛОЖЕНИЯ «АГРЕГАТОР ОБЪЯВЛЕНИЙ О ПРОДАЖЕ МАШИН»

*Полушкин Д.В., студ., Лобанков А.А. к.т.н., доц.
Самарский университет, г. Самара, Российская Федерация*

Реферат. В статье рассмотрены методы работы парсеров, возможные проблемы, с которыми можно встретиться во время выполнения парсинга, а также пути их решения.

Ключевые слова: агрегатор объявлений, парсер, веб-скрейпинг, веб-приложение.

Интернет-продажи стали настолько популярны, что многие компании и частные продавцы публикуют свои объявления в интернете. Человеку достаточно иметь доступ к интернету, чтобы без проблем и совершенно бесплатно разместить или найти объявления по любой теме. Аренда квартир или домов, предложения различного вида услуг, вакансии по трудоустройству, объявления по купле и продаже – все это размещается и можно с легкостью найти в объявлениях, которые ежедневно и в огромном количестве размещаются в интернете.

Сайты-агрегаторы играют значительную роль для людей, в них собирается огромное количество информации сразу из многих источников.

Агрегаторы объявлений – удобная система, позволяющая за короткое время отыскать необходимые предложения. Использование сайта-агрегатора является гораздо более эффективным, методом для поиска объявлений, так как в наше время интернетом пользуются почти все люди [1].

Парсинг – это метод, при котором строка или текст анализируются и разбиваются на синтаксические компоненты. Затем полученные данные преобразуются в пригодный формат для дальнейшей обработки и использования в прикладных исследованиях. Получается, что один формат данных превращается в другой, более читаемый. Если получаемые данные представляют собой необработанный код HTML, то парсер принимает его и преобразует в формат, который можно легко проанализировать и понять [2].

Парсинг представляет инструмент, который помогает найти необходимые значения из различных форм данных. Результирующие данные хранятся в базе данных или в облачных ресурсах, также их можно сохранить на локальный компьютер.

Работа парсера начинается с отправки запроса типа GET на интересующий сайт, который в ответ на запрос отдает некие данные. Чаще всего результатом запроса является HTML-разметка, которая в дальнейшем анализируется программой. Завершающим этапом парсер проводит поиск заданных данных и выполняет преобразование в нужный формат.

Сбор и анализ информации посредством парсинга не является запрещенным. Но есть

условия, которые нельзя делать: нарушать авторские права, собирать охраняемую информацию (персональные данные пользователей), нарушать работу сайтов-доноров.

Имеется два ключевых способа парсинга: нисходящий и восходящий. Их различие заключается в порядке, в котором формируются узлы синтаксического дерева.

Сверху-вниз (нисходящий): при таком методе поиск идет сверху – с первого символа в разметке и ищет заданную ему синтаксические связи. Подобным образом, древо синтаксического анализа разрастается сверху вниз, в направлении более детальной разбивки.

Снизу-вверх (восходящий): поиск начинается снизу, с самого последнего символа строки, после этого устанавливаются все более крупные синтаксические связи. Это продолжается, пока программа не встретит первый символ разметки.

Методы работы парсера не являются важным моментом. Хорошо сделанный парсер, независимо от метода будет отлично различать, необходимую, в разметке информацию.

Проблемы парсеров

Хранилище данных: масштабное извлечение данных приводит к созданию большого объема информации для хранения. Если структура хранилища данных построена неправильно, поиск, хранение и экспорт этих данных станут тяжелой задачей. Следовательно, для крупномасштабного извлечения данных должна быть идеальная система хранения данных без каких-либо недостатков и ошибок.

Структура веб-сайта меняется: каждый веб-сайт периодически обновляет свой пользовательский интерфейс, чтобы улучшить его удобство и скорость работы. Это требует различных структурных изменений разметки. Поскольку парсеры настроены в соответствии с элементами кода веб-сайта в определенный момент времени, они также требуют изменений, чтобы настроить правильную добычу данных.

Технологии защиты от парсинга. Некоторые веб-сайты используют технологии защиты от скрейпинга, которые препятствуют любой попытке доступа к ресурсу.

Обход блокировки парсера

Если отправлять повторяющиеся запросы с одного и того же IP, сайты могут обнаружить подозрительную активность и заблокировать доступ по IP. Для того чтобы избежать этого, можно использовать динамические прокси. Вращающийся/динамический прокси – это прокси-сервер, который выделяет новый IP-адрес из набора прокси, хранящихся в пуле прокси. Использование такого прокси меняет IP адрес откуда поступает запрос. Это позволяет создать впечатление, что запросы посылает не бот, а человек из разных частей мира и избежать блокировки.

Большинство плагинов для защиты от веб-скрейпинга может обнаружить парсер, если запросы идут из аналогичной подсети или находятся в непрерывной последовательности. Например: 178.176.84.1, 178.176.84.2, 178.176.84.3, 178.176.84.4 находятся в одной и той же последовательности.

Еще одним методом исключения блокировки является медленный обход страницы. При использовании парсеров, программа считывает данные с нечеловеческой скоростью. Такие действия легко обнаруживаются защитой сайта. Добавляя случайные задержки в действиях программы, может сделать данные действия похожими на человеческие, чтобы плагины защиты сайтов их не обнаруживали. Слишком частая отправка запросов может привести к сбою сервера сайта. Необходимо ограничить частоту запросов, чтобы не перегружать ресурс и не подавать лишние подозрения для блокировки IP-адреса парсера.

User-Agent, или агент пользователя – это компьютерная программа, представляющая пользователя и выполняющая действия от его лица, например, браузер в контексте всемирной паутины [3]. Некоторые сайты блокируют user-agent, если они не из основного браузера. Если же user-agent не установлен, многие веб-сайты не разрешают доступ к контенту. Решение этой проблемы заключается в том, что нужно либо создать список пользовательских агентов, либо использовать библиотеки, такие как fake-useragent (python). Спам-боты, менеджеры закачек и некоторые браузеры нередко шлют подложные UA-строки, чтобы выдать себя за других клиентов. Эта ситуация известна под названием подмена или подделка пользовательского агента (user agent spoofing).

Headless браузеры

Проблема при парсинге некоторых веб-сайтов заключается в том, что содержимое отображается JavaScript кодом, а не HTML. Для сканирования таких сайтов может потребоваться

использовать какой-либо headless браузер. Автоматизированные браузеры, используемые такими инструментами как Selenium и Puppeteer, так же можно использовать для контроля и парсинга динамических веб-сайтов. Такой способ более затратный, но наиболее эффективный.

Большинство веб-сайтов используют CAPTCHA для обнаружения трафика ботов. Для их устранения можно использовать сервисы для обхода капчи. Нужно иметь в виду, что данные услуги предоставляются платно, а также могут увеличить время парсинга страниц. Поэтому следует учитывать дополнительное время и расходы, которые, возможно, придется понести, если использовать сервис для обхода капчи.

Избегание ловушек сайтов

Многие сайты пытаются обнаружить веб-парсеры и вводят невидимые элементы на страницу, которые может использовать только поисковик. Необходимо определить имеет ли элемент стили display: none или visibility: hidden, и избегать действий с ними, иначе это может идентифицировать парсер.

Список использованных источников

1. Что такое сайты-агрегаторы [Электронный ресурс]. – Режим доступа: <https://blog.calltouch.ru/chto-takoe-sajt-agregator-opredelenie-sozdanie-i-prodvizhenie-spisok-sajtov-agregatorov-tovarov-i-uslug>. – Дата доступа: 15.03.2023.
2. Что такое парсер и как с ним работать [Электронный ресурс]. – Режим доступа: <https://romi.center/ru/learning/article/what-is-data-parsing>. – Дата доступа: 15.03.2023.
3. Агент пользователя [Электронный ресурс]. – Режим доступа: https://developer.mozilla.org/ru/docs/Glossary/User_agent. – Дата доступа: 16.03.2023.

УДК 004

КАНБАН-МЕТОДОЛОГИЯ И ЕЁ ИСПОЛЬЗОВАНИЕ ДЛЯ ПОВЫШЕНИЯ ЭФФЕКТИВНОСТИ БИЗНЕС-ПРОЦЕССОВ ОРГАНИЗАЦИИ

Сосновская А.И., студ., Пунчик З.В., к.соц.н., доц.

Белорусский государственный экономический университет, г. Минск, Республика Беларусь

Реферат. В статье рассмотрена Канбан-методология, ее особенности и преимущества. Также рассмотрены способы повышения эффективности бизнес-процессов в организациях за счет внедрения данной методологии.

Ключевые слова: Канбан-методология, S.T.A.T.I.K, визуализация процесса, оптимизация бизнес-процессов, управление потоком работы, улучшения процесса.

Канбан-методология – это метод управления процессами, который использует визуальное отображение потока работы и управление рабочими задачами. Он был разработан японской компанией Toyota в 1950-х годах, и с тех пор применяется в различных отраслях и компаниях по всему миру.

Первоначально Канбан-методология была изобретена как часть системы производства Toyota топ-менеджерами компании во главе с ее президентом Таичи Оно. Они поставили цель выпускать автомобили с максимальной скоростью, а складские запасы свести к минимуму. Канбан-система использовала карточки, которые применялись для передачи информации от одного этапа производства к другому. Эта система позволила компании выполнить поставленные ими цели за небольшой период времени. Впоследствии Канбан-методология была адаптирована для использования в других отраслях и бизнес-процессах. Сегодня Канбан используется в различных областях, включая разработку программного обеспечения, маркетинг, управление проектами и т. д. [1].

Ключевые принципы Канбан-методологии

Первый принцип заключается в том, чтобы визуализировать процесс работы. Для этого используется доска Kanban, на которой отображаются все задачи, их текущий статус и