

Что такое Octoparse? Это инструмент для парсинга, не требующий навыков кодинга. В этом **обзоре Octoparse** вы узнаете про его функционал, а также о том, как **веб скрапинг в Octoparse** позволяет автоматически извлекать информацию с сайтов.

Что такое Octoparse

Если вы хоть поверхностно крутитесь в сфере приложений для сбора данных или софта для автоматизации процессов, велик шанс, что Octoparse уже давно и прочно засел в вашем арсенале ультраполезных утилит, значительно облегчающих веб скрапинг и сопутствующие ему задачи. Его уже взяло на вооружение немало специалистов по стратегическому планированию продвижения продуктов из областей маркетинга и трейдинга, в чьи задачи входит анализ конкурентности рынка и рекламный таргетинг. Но для непосвящённых неофитов данного дела мы подготовили детальный разбор, в котором вы почерпнёте для себя множество полезной информации о том, как использовать Octoparse и как данный, без преувеличения, революционный софт выводит процесс веб-скрапинга на качественно новый уровень.

Особенности Octoparse

Сбор большого количества информации и соответствующее объёмам полученных данных масштабирование задач уже не первый год как прочно влились в сферу онлайн коммерции. Рынок переполнен огромным количеством готовых решений, позволяющих на полную катушку задействовать цифровые ресурсы. Эффективность их разнится, но основным препятствием в данной комплексной сфере было и остаётся достаточно большое количество барьеров на пути к заветным данным. Тут мы и подходим к, пожалуй, главному преимуществу, делающему сбор данных в Octoparse максимально гладким и быстрым. Его способность вести себя максимально “человечно” по сравнению с альтернативными инструментами подобного толка позволяет как обойти ограничение запросов, так и максимально расширить функции и возможности скрапера, в норме ограниченными солидным количеством проверок, аутентификаций, а также прочих цифровых препятствий и разного рода лимитов.

Среди главных особенностей инструмента для сбора данных Octoparse стоит в первую очередь отметить:

- **Мультинаправленность.** В какой бы области или индустрии вы не работали, от соцмедиа и коммерции до образования и новостей, найдётся готовое решение на буквально любой случай. Перейдя в необходимый раздел, вам останется лишь несколько шагов настройки, прежде чем вы сможете заставить данные работать на себя, а не себя на данные.
- **Интуитивность и удобство.** Комфортный и многофункциональный интерфейс, совмещающий в себе лаконичность и утилитарность, даёт вам возможность быстро настроить автоматизацию рутинных задач, оставляя вам максимум времени на обдумывание глобальных вопросов.

- **Неприкосновенность данных.** Безопасность в Octoparse обеспечивается целым рядом наиболее современных цифровых протоколов шифрования. Это позволяет удостовериться, что только вы имеете доступ к полученной информации, и не волноваться о сохранности ваших данных.
- **Киллер фичи.** Прорывные технологические новшества, такие как способность Octoparse обходить captcha, что служит одним из самых сложных препятствий для большинства подобных утилит, позволяет увеличить количество запросов, таким образом в разы расширив ваш радиус возможностей.
- **Всеобъёмность.** В то время как большинство инструментов для веб-скрапинга часто спотыкается об ограничения в доступности того или иного типа данных, Octoparse предоставляет вам обширный инструментарий для сбора, способный извлекать не только текст, но также и разного рода медиафайлы, а также HTML-ссылки, URL-адреса, метаданные и многое другое.
- **Планирование и автоматизация.** Время крайне ценно, особенно в цифровом мире, где прорывы и смены парадигм случаются порой чуть ли не каждую неделю, и каждая секунда на счету. Благо Octoparse позволяет по максимуму автоматизировать ваши рабочие процессы, заранее назначая необходимые проверки и сканирования.
- **Встроенный VPN.** Региональные ограничения уже не первый год ставят палки в колёса пользователям по всему миру. Особенно остро вопрос как обойти ограничения запросов стоит именно в области веб-парсинга. Именно для этих целей внедрена тонкая настройка прокси в Octoparse, делающая процесс беспрепятственным и быстрым.
- **Интеграция Adblock.** При современном обилии попапов, баннеров, видео и прочей рекламной мишуры, процесс экстракции данных может заметно затянуться. К счастью продуманный API Octoparse имеет встроенный блокировщик рекламы, что позволяет оптимизировать процессы и не волноваться, что парсинг запнётся в самый неожиданный момент.

Это был лишь краткий обзор Octoparse. Теперь же давайте детальнее ознакомимся с данной утилитой.

Возможности парсера Octoparse

Аналитика данных в бизнесе не терпит упущений и просчётов. Именно поэтому веб-скрапинг в Octoparse как нельзя лучше подходит в качестве инструмента экстракции данных. Вот лишь несколько важных аспектов, что помогут вам всегда оставаться на острие.

Извлечение IP-адреса

Веб-скрапинг с Octoparse позволяет извлекать IP-адреса, фильтровать данные и экспортировать результаты для дальнейшего использования. Определение IP-адресов помогает обнаружить вредоносный трафик, отследить киберугрозы, предотвратить

мошенничество и проанализировать схемы доступа к серверу для улучшения безопасности. Это хороший метод, чтобы проверить безопасность данных при использовании, удостоверившись в надёжности системы.

Извлечение адреса электронной почты

Если вы работаете в области маркетинга или лидогенерации, парсер имэйлов просто незаменим для конкурентного анализа, а также составления рассылочных списков и выстраивания продуманных маркетинговых стратегий. Готовые решения для сбора данных в Octoparse позволят вам занять в своё распоряжение URL, адреса имэйлов и номера телефонов, а возможность дальнейшей интеграции с такими сервисами как Zapier подарит вам возможность оперативно пустить всю приобретённую информацию в ход без необходимости в сложном упорядочивании.

Извлечение веб-данных

Настройка прокси Octoparse в сочетании с глубокой интеграцией API позволяет оптимизировать процесс парсинга, а утилизация облачных сервисов даёт возможность для автоматизированного скрапинга 24/7. Всегда имея на руках только актуальную информацию, вы сможете на лету корректировать стратегию продвижения и оперативно реагировать на любые перемены.

Извлечение изображения

Парсинг изображений позволяет быть в курсе текущих трендов в визуальном дизайне, что немаловажно в сферах электронной коммерции и маркетинга. Скраппер Octoparse позволяет настроить пагинацию, что даст вам возможность сразу заполучить множество URL с разных страниц вместо того, чтобы тратить часы на сбор изображений вручную.

Извлечение номера телефона

Octoparse упрощает извлечение телефонных номеров благодаря автоматическому определению на основе искусственного интеллекта и продвинутым алгоритмам обхода капчи. Массовое извлечение делает этот метод эффективным для сбора больших объемов данных, а функция ротации IP-прокси для предотвращения блокировок попросту неоценима. Плотная интеграция со всеми актуальными источниками, такими как Yelp и LinkedIn, позволяет не только значительно сэкономить время, но и заполучить сопутствующую информацию, в т.ч. имена и адреса абонентов.

Разнородный сбор данных

API Octoparse делает возможным всестороннюю агрегацию данных благодаря новейшим методам обхода captcha. Даже сложные скрипты и прочие ухищрения наподобие вложенных структур не способны помешать скраппингу. Дальнейшая интеграция Octoparse с Python, в частности такими библиотеками, как BeautifulSoup и Selenium, позволяет чётко систематизировать и структурировать полученные данные.

Тарифы Octoparse

Цены Octoparse разнятся в зависимости от выбранного пакета. Перейдя в раздел Pricing на сайте Octoparse и введя диапазон необходимых вам сайтов и частоту парсинга, вы сможете определить наиболее подходящий вам тариф. Здесь же мы предоставим вам краткий обзор.

Free

Бесплатный тариф для небольших проектов:

- доступ к десктопному приложению
- неограниченное количество страниц для скреппинга
- назначение до 10 тасков
- экспорт до 10000 строк данных за раз (до 50000 в месяц)

Standard

Стандартный план для командных проектов. Включает в себя все преимущества бесплатного пакета плюс:

- более 500 шаблонов для готовых решений
- назначение до 100 тасков
- облачный сервис (до 6 параллельных процессов)
- неограниченный экспорт данных
- ротация IP
- премиум прокси
- автопрохождение капчи
- загрузка файлов, в т.ч. изображений
- автоэкспорт
- автозапуск процессов по графику

Professional

План для бизнесов среднего масштаба. Включает в себя все преимущества стандартного тарифа плюс:

- назначение до 250 тасков
- облачный сервис (до 20 параллельных процессов)
- экспорт в Google Sheets, Google Drive или Dropbox
- автоматический бэкап данных
- продвинутый API
- приоритетная поддержка

Также советуем обратить внимание, что цены на Octoparse варьируются в зависимости от размера оплаты. Предоплатив выбранный пакет сразу на год вперёд, вы сэкономите до 16% от суммы.

Интерфейс Octoparse

Интуитивно понятный интерфейс делает веб скрапинг с Octoparse тривиальной задачей. Вам понадобится лишь несколько нажатий на клавиши и пара минут времени. Вот краткий обзор интерфейса Octoparse, который поможет вам быстрее сориентироваться и приступить к выполнению задуманного.

Профиль пользователя

Нажав на иконку с аватаром в левом верхнем (в приложении) или правом верхнем (в веб версии) углу экрана, вы перейдёте на вкладку My Account. Тут вы можете:

- редактировать профиль
- проверить или сменить текущий тарифный план
- проверить текущий баланс кредитов и купонов
- увидеть запланированные транзакции
- подвязать сторонние аккаунты

Помните, что вы можете присоединить до пяти аккаунтов. К примеру, если вы хотите добавить ваш аккаунт Apple, Microsoft или Google в Octoparse, вы можете добавить все три, тем самым упростив доступ к приложению в будущем.

Создание новой задачи

Добавление задач осуществляется по нажатию на кнопку New прямо под аватаром. Тут вы можете как создать новый таск, так и выбрать уже готовое решение или импортировать уже готовые таски. Подробнее об этом читайте ниже в разделе “Как создать новую задачу в Octoparse”.

Dashboard - информационная панель

Непосредственно под кнопкой New вы найдёте кнопку My Tasks, нажав на которую вам откроется информационная панель, где вы сможете увидеть все текущие, выполненные и запланированные задачи. Удобный фильтр по словам позволяет мгновенно выискивать нужные задачи среди десятков и сотен других. Например, введя слово Twitter, Octoparse отфильтрует для вас все таски связанные с этой социальной сетью. Различные типы сортировки и переключение между облачными и локальными процессами ещё больше упрощает процесс менеджмента задач.

Шаблоны

Под вкладкой Task List вы увидите вкладку Templates, где вы сможете найти множество готовых решений на любые случаи:

- **Сбор информации с соцсетей.** Используя предзаготовленные шаблоны, вам станет доступна быстрая и плотная интеграция Octoparse с Instagram, Facebook и другими популярными сайтами данной направленности.

- **Лидогенерация.** Такие инструменты, как Google Search Scraper и Yellow Pages Scraper станут незаменимыми помощниками в деле поиска потенциальных клиентов и выстраивания грамотной маркетинговой стратегии.
- **Электронная коммерция.** При работе с онлайн площадками Amazon, eBay, Etsy и Uber решения для веб скрапинга попросту необходимы. Шаблоны по типу Amazon Scraper и eBay Scraper помогут при поиске критично важной информации о пользователях, товарах и ценах.
- **Рынок работы.** Такие популярные шаблоны, как Indeed Job Scraper и Town Work Job Scraper дают возможность для агрегации списков компаний, их сотрудников, джоб дескрипшнов, открытых вакансий и прочей информации, представляющей интерес для всех, кто задействован в сферах рекрутинга и аналитики рынка работы.

Помните, что не все шаблоны доступны в Octoparse бесплатно. Подключив пакет от стандартного и выше, вы получите неограниченный доступ к 500+ шаблонам. Цены Octoparse вполне демократичны, а готовые решения позволят вам сэкономить уйму сил и времени.

Инструменты

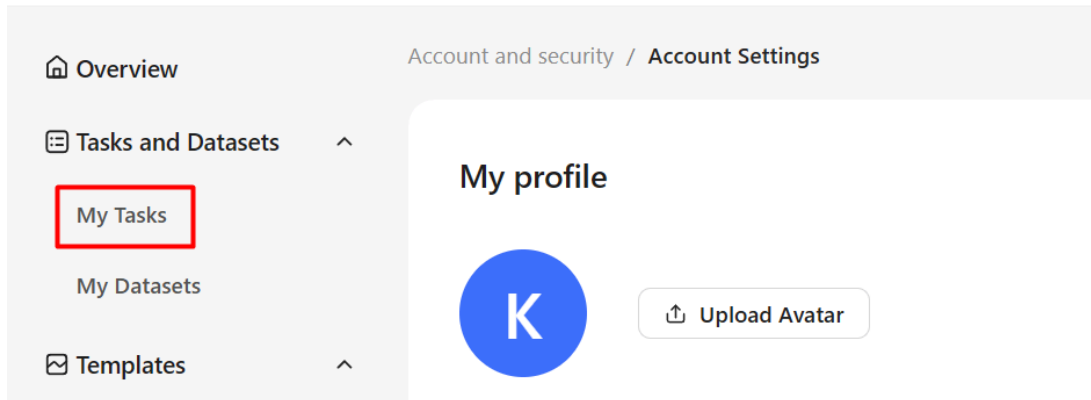
Финальная вкладка Tools даёт вам доступ к утилите RegEx, позволяющей создавать регулярные выражения, используемые для поиска, извлечения и обработки текста, а также Database Auto-export Tool, с помощью которой вам станет доступна автоматизированная выгрузка данных непосредственно в базы Excel или MySQL.

Как создать новую задачу в Octoparse

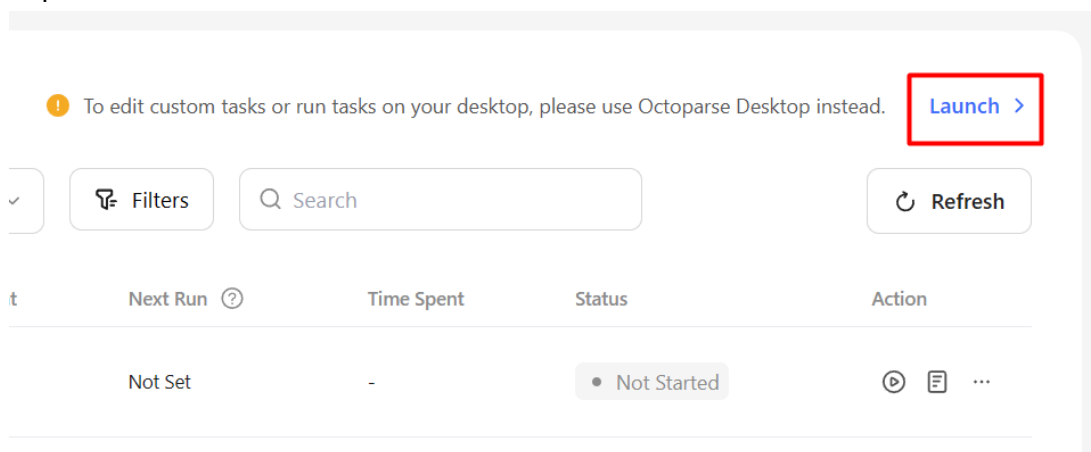
Комплексный софт порой вызывает отторжение в первую очередь нагромождением непонятных опций и кнопок, что часто вызывает неприятие со стороны пользователей. Обширный выбор tutorиалов Octoparse на Youtube, как от самой компании, так и сторонних обозревателей, поможет вам детально вникнуть во все нюансы его работы. Ну а для тех, кто ещё колеблется над вопросом пробовать или нет, мы приведём здесь сжатую версию, чтобы вы смогли наглядно убедиться в удобстве и простоте управления программой.

Предположим, вы уже зарегистрировали свой аккаунт. Чтобы создавать новые задачи (задачи) вам потребуется установить десктопное приложение.

- Если вы находитесь на странице своего аккаунта (User Center) в веб версии, просто нажмите кнопку My Tasks в левом верхнем углу (она находится во вкладке Tasks And Datasets, сразу под Overview).

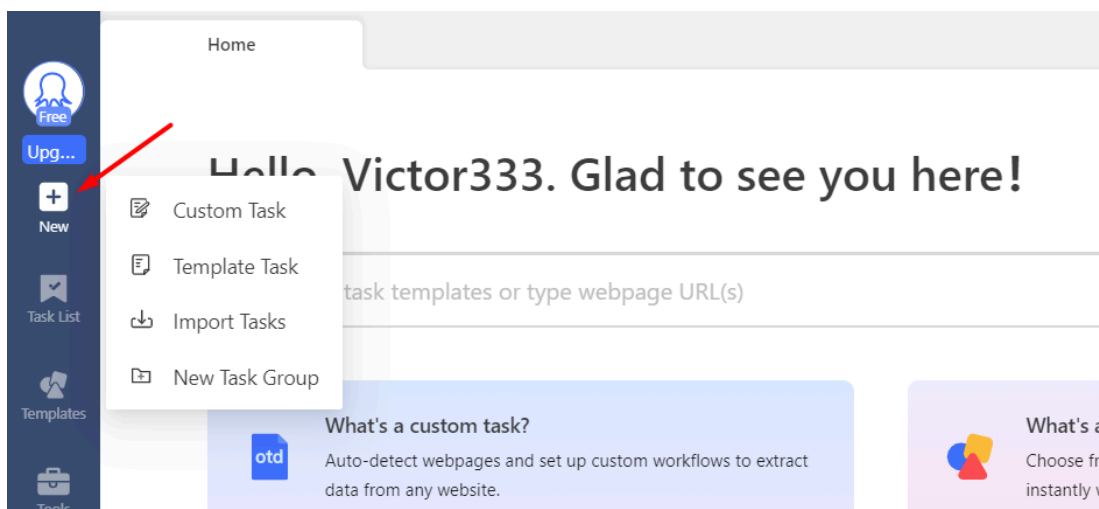


- В правом верхнем углу открывшейся страницы вы увидите кнопку Launch, при нажатии на которую вам будет предложен выбор либо перейти в десктопное приложение (в случае если оно уже установлено), либо скачать это самое приложение.



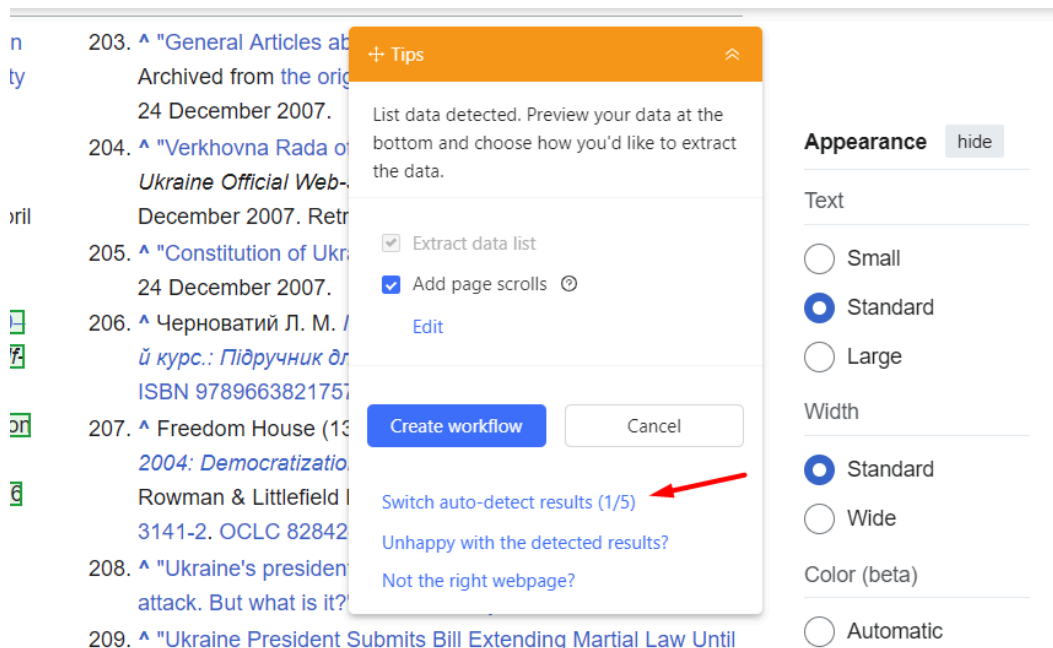
Установив приложение Octoparse в любое удобное для вас место, открываем его с рабочего стола. Для вашего удобства наличествуют удобные тултипы-подсказки, которые можно при желании отключить. Далее приводим для вас пошаговую инструкцию как использовать Octoparse при запуске нового таска.

1. **Добавление таска.** Жмём на кнопку New в левом верхнем углу (сразу под иконкой пользователя). В выпавшем меню вам будет доступно сразу несколько опций:

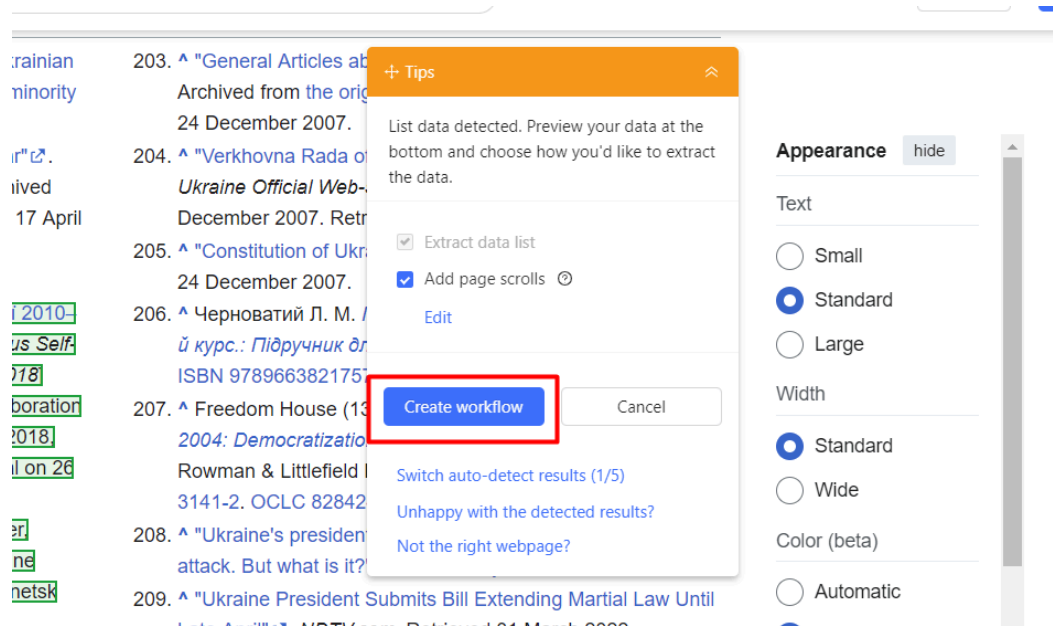


- A. Custom task. В открывшемся окне вводим адрес сайта и нажимаем Save. Браузеру понадобится несколько секунд на анализ данных, после чего вы сможете продолжить по своему усмотрению.
- B. Template task. При нажатии вам будет представлен выбор из множества готовых шаблонов для самых разных сайтов и задач. К примеру, если вам нужно интегрировать парсер Octoparse в Google Play или Twitter, вы можете выбрать уже готовое решение, тем самым сэкономив своё время и усилия.
- C. Import tasks. Если вы ранее уже сохраняли проекты под другим аккаунтом или на другом устройстве, вы можете импортировать их через эту вкладку и продолжить работу. Если у вас уже есть готовая группа, вы можете легко добавить к ней свой новый проект.
- D. New Task Group. Создание групп позволит вам легко отслеживать все ваши проекты, удобно рассортированные в соответствии с вашими предпочтениями или направленностью. К примеру, вы можете создать группу с интеграцией Octoparse в Instagram, Facebook и другие социальные сети, в то время как другая группа будет сосредоточена на электронной почте или онлайн-магазинах.

2. Оценка и сортировка данных. После автоматического анализа данных, у вас появится возможность переключаться между различными типами данных. К примеру, предоставив для анализа страницу в Facebook, Octoparse даст вам возможность отдельно увидеть ссылки, фото, статистику и т.д. Все данные удобно представлены в виде полей внизу экрана. При необходимости вы можете удалить нерелевантные поля.



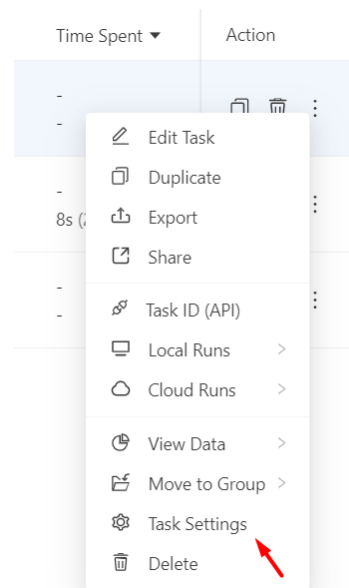
3. Запуск процесса. Когда вы убедились, что все данные подобраны верно, можно начинать настройку рабочего процесса. На плавающей вкладке Tips нажимаем Create Workflow. После настройки деталей по типу различной пагинации или бесконечного скролла, нажимаем Run Task Now. Вам будет предложен выбор - запустить процесс локально (доступно в Otoparse бесплатно) или же в облаке (премиум опция). После вам будет предложено экспортировать полученные данные в любом удобном вам формате (HTML, Google Tabs, Excel и пр.)



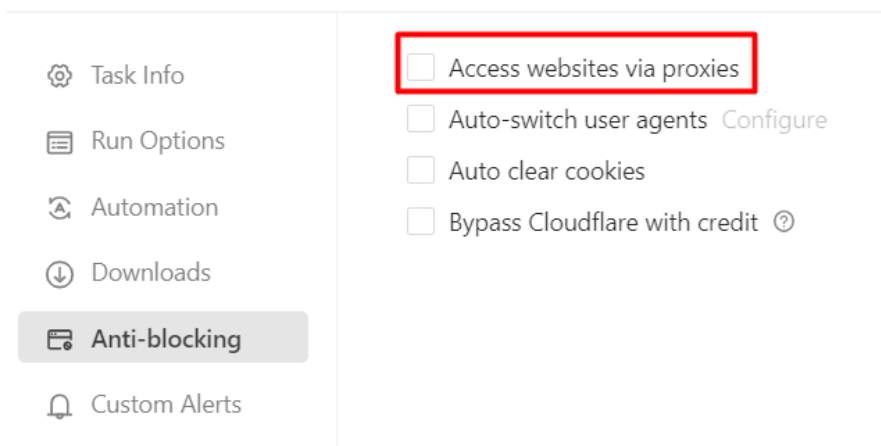
4. Настройка прокси. Otoparse внедрили продвинутые алгоритмы VPN, прекрасно осознавая, как немалое количество пользователей по всему миру страдает от блокировок.

Если вы подозреваете, что блокировка может помешать вам или же просто хотите обезопасить себя, рекомендуем заранее настроить прокси Octoparse. Вот как сделать это в несколько несложных шагов:

- A. Выберите необходимую задачу во вкладке Task List. Нажмите на три точки справа, в выпавшем меню выберите опцию Task Settings.



- B. В новом меню перейдите на вкладку Anti-blocking. Помимо прочих важных опций, таких как автоматическая очистка куки для пресечения отслеживания, автосмена агентов (браузеров, через которые программа будет выполнять вход на страницу) и обход защиты Cloudflare, вы увидите опцию Access website via proxies. Именно она нам и нужна.



- C. По клику вам будет предложено несколько вариантов стран и частот интервалов ротации прокси, но при желании вы можете использовать и собственные адреса.

☒ Access websites via proxies

☐ Use Octoparse proxies [?](#)

Account Credit: \$1 [Add Credit](#) [Refresh](#)

Low account credit. Add credit now to avoid interruption of services.

Select Country/Region

Default

▼

Rotate every

10mins

▼

☒ Use my own proxies [Configure](#)

☐ Auto-switch user agents [Configure](#)

D. Кликнув на [Configure](#) рядом с опцией Use My Own Proxies, вы сможете настроить интервал переключений, а также внести собственный список желаемых IP-адресов.

☒ Access websites via proxies

☐ Use Octoparse proxies [?](#)

Account Credit: \$1 [Add Credit](#) [Refresh](#)

Low account credit. Add credit now to avoid interruption of services.

Select Country/Region

Default

 ▼

Rotate every

10mins

 ▼

☒ Use my own proxies [Configure](#) 

E. По завершению всех необходимых манипуляций жмём [Confirm](#) и [Save](#). Теперь вы можете не волноваться, что какие-либо данные ускользнут от вас из-за блокировки.

Proxy Settings

×

Switch every

60s

IP Proxies

Type in your own list of proxy servers and put them in separate lines.
Separate each IP address and its port number with a colon(:).
e.g.
1.1.1.1:8000
2.2.2.2:9000
...

Cancel

Confirm

Cancel

Save

Надеемся, что наш краткий сегодняшний экскурс на тему что такое Octoparse поможет вам сориентироваться и обуздать потенциал веб скрапинга. Если вы ищете дополнительную информацию, вы всегда можете найти её на канале Octoparse в Youtube или посетить крайне информативный блог на официальном сайте приложения.