Колледж Автономной некоммерческой образовательной организации высшего образования

«Научно-технологический университет «Сириус»

**Реферат на тему**

**«Парсинг данных»**

Работу подготовил:

Студент группы К0709-24/1 Гладченко Б.И

Преподаватель:

Яковлева С. В.

# Оглавление

Оглавление

[Оглавление 2](#_Toc191226070)

[Введение 3](#_Toc191226071)

[1. Парсинг в современном мире 4](#_Toc191226072)

[1.1. Определение и принцип работы. 4](#_Toc191226073)

[1.2 Что о парсинге говорит закон? 4](#_Toc191226074)

[1.2.1 Законодательное регулирование парсинга в Российской Федерации 4](#_Toc191226075)

[1.2.2 Ограничения и правовые риски при использовании парсинга 5](#_Toc191226076)

[1.3 Область применения парсеров 6](#_Toc191226077)

[1.3.1 Сбор оперативной информации 6](#_Toc191226078)

[1.3.2 Автоматизация сбора больших объемов данных 6](#_Toc191226079)

[1.3.3 Репликация и агрегация контента 7](#_Toc191226080)

[1.3.4 Применение парсеров в маркетинговых исследованиях 7](#_Toc191226081)

[Вывод 9](#_Toc191226082)

[Список литературы 11](#_Toc191226083)

# Введение

В современном мире информация играет одну из ключевых ролей. Как сказал Натан Ротшильд: “Кто владеет информацией, тот владеет миром”, и это действительно так. [1] Тот, кто быстрее и качественнее соберет данные, тот раньше получит аналитику и сможет ее использовать. С появлением интернета, появилась возможность анализировать информацию. Результат научно-технического прогресса привел к тому, что человек не в состоянии обработать огромное количество информации вручную. А в современных условиях надо быстро обрабатывать огромное количество информации. И тут на помощь приходит парсинг. Это быстро развивающаяся ветвь в программировании. В своем проекте я покажу, как использовать парсинг, его возможности. Смысл парсинга в том, чтобы ускорить рутинную работу. Чтобы собрать и сохранить в таблице список из тысячи статей на сайте, человек потратит часы. Парсер сделает эту работу за несколько минут. Парсер ускоряет работу в сотни раз и допускает меньше ошибок, чем человек. Для написания парсеров подходят любые языки программирования, на которых создаются программы для работы в интернете.

# 1. Парсинг в современном мире

## 1.1. Определение и принцип работы.

Парсинг — это процесс автоматизированного сбора информации с различных веб ресурсов, например, с веб сайтов. В этом случае плюсом является то, что вы сами определяете, откуда и какую информацию собирать, как формировать набор данных. Вы даже можете формировать разные наборы данных и затем объединять их по определенному признаку, например географическому. [1] Парсить можно данные с сайтов, поисковой выдачи, форумов и социальных сетей, порталов и агрегаторов. Сами же парсеры – скрипты, с помощью которых и осуществляется сбор и систематизация данной информации. [3]. Они работают так:

-ищут источники по заданным параметрам — например, парсеру можно дать список сайтов, а он найдёт на них страницы с ценами;

-извлекают из источников нужную информацию — несколько строк текста, ссылку или сумму; -преобразуют информацию — например, парсер может взять фрагмент из HTML-документа и преобразовать его в текст без кода;

-сохраняют информацию в нужном формате — например, в виде списка или таблицы в Excel.

## 1.2 Что о парсинге говорит закон?

### *****1.2.1 Законодательное регулирование парсинга в Российской Федерации*****

Парсинг, или автоматизированный сбор данных с веб-сайтов, сам по себе является законным, так как представляет собой всего лишь инструмент для автоматизации обработки открытых данных. Данные, которые парсеры получают автоматически, в большинстве случаев можно было бы собрать вручную, а законодательство Российской Федерации не запрещает использование таких методов. Однако важно учитывать ряд ограничений и рисков, связанных с этим процессом.

### *1.2.2 Ограничения и правовые риски при использовании парсинга*

Несмотря на то, что сам факт автоматизированного сбора информации не противоречит закону, существуют определённые случаи, когда использование парсеров может привести к административной или даже уголовной ответственности.

#### Перегрузка серверов и нарушение работы сайта

Использование парсинга таким образом, что он создаёт чрезмерную нагрузку на сервер сайта, может быть квалифицировано как одна из форм DDoS-атаки. Если в результате работы парсера сайт перестаёт нормально функционировать, это может быть расценено как неправомерный доступ к компьютерной информации. Такие действия подпадают под:

**Статью 272 УК РФ («Неправомерный доступ к компьютерной информации»)**, наказание по которой включает штраф до 200 тысяч рублей или лишение свободы до 2 лет;

**Статью 273 УК РФ («Создание, использование и распространение вредоносных компьютерных программ»)**, предусматривающую штраф до 500 тысяч рублей и лишение свободы на срок до 7 лет.

#### Использование данных для рассылки спама

Полученные при парсинге данные, такие как контактные адреса, не могут быть использованы для массовой рассылки рекламы без согласия пользователей. Это подпадает под **Закон «О рекламе» (ФЗ-38)**, который предусматривает:

Штраф за несанкционированную рассылку спама в размере до **500 тысяч рублей** за каждое незаконное рекламное сообщение.

#### Нарушение авторских прав и плагиат

Если парсер используется для копирования и повторного использования контента без разрешения правообладателя, это может квалифицироваться как нарушение авторских прав. В этом случае применяются нормы **статьи 146 УК РФ («Нарушение авторских и смежных прав»)**, согласно которым:

За незаконное использование объектов авторского права грозит штраф до **500 тысяч рублей** или иные меры наказания.

#### Обработка и распространение персональных данных

Если в ходе парсинга собираются и передаются персональные данные пользователей (ФИО, телефон, адрес, электронная почта и т. д.), то такие действия подпадают под **ФЗ-152 «О персональных данных»**. Распространение таких данных без согласия субъектов является нарушением закона и может повлечь:

Административную ответственность с штрафами до **18 миллионов рублей** для юридических лиц.

## 1.3 Область применения парсеров

Парсинг, или автоматизированный сбор данных с веб-сайтов, широко применяется в различных сферах благодаря своей эффективности и способности обрабатывать большие объемы информации. Рассмотрим подробнее основные области использования парсеров:

### *1.3.1 Сбор оперативной информации*

В условиях, когда данные быстро устаревают, парсеры позволяют мгновенно получать и обновлять информацию. Это особенно важно в следующих случаях:

* **Финансовые рынки**: Мониторинг курсов валют, цен на акции и другие финансовые показатели в реальном времени. Автоматический сбор таких данных обеспечивает своевременное принятие решений в торговле и инвестициях.
* **Новости и события**: Отслеживание свежих новостных материалов и событий, где информация теряет актуальность уже через несколько минут. Парсеры помогают агрегировать новости из различных источников для создания актуальных лент.

### *1.3.2 Автоматизация сбора больших объемов данных*

Ручной сбор информации с множества веб-страниц может быть трудоемким и неэффективным. Парсеры решают эту проблему, автоматизируя процесс и снижая затраты времени и ресурсов. Примеры включают:

* **Мониторинг цен и ассортимента**: Сбор данных о ценах и наличии товаров на различных платформах электронной коммерции для анализа конкурентов и динамического ценообразования.
* **Сбор контактной информации**: Извлечение адресов электронной почты, номеров телефонов и других контактных данных для формирования баз потенциальных клиентов или партнеров.

### *1.3.3 Репликация и агрегация контента*

Парсеры используются для копирования и последующего размещения материалов с других сайтов, что может быть полезно в следующих ситуациях:

* **Создание агрегаторов**: Сбор информации с различных ресурсов для формирования единой базы данных, например, агрегаторы новостей, объявлений или товаров.
* **Резервное копирование**: Создание локальных копий веб-сайтов для архивирования или переноса контента на другие платформы.

### *1.3.4 Применение парсеров в маркетинговых исследованиях*

В маркетинге парсеры являются незаменимыми инструментами для сбора и анализа данных, что способствует принятию обоснованных решений и разработке эффективных стратегий. Основные направления использования парсеров в маркетинговых исследованиях:

* **Анализ конкурентов**: Сбор информации о ценах, акциях, ассортименте и других маркетинговых активностях конкурентов. Это позволяет выявлять их стратегии и адаптировать собственные действия на рынке.
* **Изучение целевой аудитории**: Сбор данных о поведении пользователей, их предпочтениях и отзывах. Это помогает сегментировать аудиторию и разрабатывать персонализированные предложения.
* **Мониторинг репутации бренда**: Отслеживание упоминаний компании или продукта в интернете, включая социальные сети, форумы и новостные сайты. Это позволяет своевременно реагировать на отзывы и управлять репутацией.
* **Трендовый анализ**: Выявление актуальных тем и тенденций на основе анализа поисковых запросов, хэштегов и обсуждений в сети. Это помогает создавать востребованный контент и предлагать продукты, соответствующие текущим интересам потребителей.

Использование парсеров в маркетинговых исследованиях позволяет автоматизировать рутинные процессы, повышать точность данных и оперативно реагировать на изменения рынка, что в конечном итоге способствует повышению конкурентоспособности бизнеса. Парсинг данных продолжает развиваться, находя всё новые области применения и становясь всё более совершенным инструментом для анализа информации. В условиях цифровой трансформации и роста объемов данных парсеры помогают компаниям и исследователям быстро находить, структурировать и анализировать информацию. Однако при использовании парсинга важно учитывать не только его преимущества, но и юридические аспекты, чтобы избежать правовых рисков и этических нарушений.

# Вывод

Парсинг данных — это мощный инструмент, позволяющий автоматизировать сбор, обработку и анализ информации из различных источников. В современном мире, где объемы данных растут с огромной скоростью, технологии парсинга помогают компаниям, исследователям и аналитикам эффективно работать с большими массивами информации.

Одним из ключевых преимуществ парсинга является его универсальность: он применяется в самых разных сферах — от финансовых рынков и маркетинговых исследований до анализа конкурентной среды и мониторинга новостей. Благодаря парсерам компании могут оперативно обновлять базы данных, отслеживать изменения цен, исследовать поведенческие факторы аудитории и оптимизировать бизнес-стратегии. В условиях высококонкурентных рынков такие технологии позволяют принимать взвешенные и своевременные решения, что является важным фактором успеха.

Несмотря на очевидные преимущества, парсинг также связан с рядом ограничений и рисков. В первую очередь это вопросы правового регулирования. Законодательство в области автоматизированного сбора данных неоднозначно и зависит от характера извлекаемой информации, способов её использования и уровня вмешательства в работу веб-ресурсов. Например, перегрузка серверов, использование парсинга для спама или копирование контента без разрешения правообладателей может повлечь за собой юридическую ответственность. Особенно важным аспектом является защита персональных данных: нарушение норм ФЗ-152 "О персональных данных" может привести к серьёзным штрафам и судебным разбирательствам.

Кроме юридических аспектов, существуют и технические ограничения. Многие сайты стремятся защитить свои данные от парсеров, используя механизмы CAPTCHA, динамическую загрузку контента, блокировки IP-адресов и другие методы противодействия. Это требует от разработчиков парсеров постоянного совершенствования алгоритмов работы, использования прокси-серверов и обхода различных защит.

Таким образом, парсинг остаётся важным инструментом, но его применение требует внимательного подхода. При его использовании необходимо соблюдать баланс между эффективностью сбора данных и соблюдением правовых норм, а также учитывать этическую сторону вопроса. В будущем можно ожидать дальнейшего развития технологий парсинга, интеграции с искусственным интеллектом и машинным обучением, что позволит ещё более точно и глубоко анализировать информацию. Однако одновременно с этим будут ужесточаться меры защиты данных, что сделает процесс парсинга более сложным и требовательным к техническим решениям.

# Список литературы

1. Парсинг, как инструмент для участников рынка недвижимости [Электронный ресурс] // Diggernaut. — 2025. — URL: <https://www.diggernaut.ru/blog/parsing-kak-instrument-dlya-uchastnikov-rynka-nedvizhimosti/> (дата обращения: 23.02.2025).
2. Что такое парсер (граббер)? [Электронный ресурс] // MyBlaze. — 2025. — URL: <https://myblaze.ru/chto-takoe-parser-grabber/> (дата обращения: 23.02.2025).
3. Как парсить сайт? [Электронный ресурс] // Cossa. — 2025. — URL: <https://www.cossa.ru/imarketing/261951/> (дата обращения: 23.02.2025).
4. Что такое парсинг и что о нём обязательно нужно знать маркетологу [Электронный ресурс] // Skillbox. — 2025. — URL: <https://skillbox.ru/media/marketing/chto-takoe-parsing-i-chto-o-nyem-obyazatelno-nuzhno-znat-marketologu/#stk-1> (дата обращения: 23.02.2025).