

Никулин Владимир Валерьевич
Мордовский государственный университет им. Н.П. Огарёва
Захаров Владислав Андреевич
Мордовский государственный университет им. Н.П. Огарёва
Кулдыркаев Александр Владимирович
Мордовский государственный университет им. Н.П. Огарёва
Лобанова София Борисовна
Мордовский государственный университет им. Н.П. Огарёва
Пулов Даниил Евгеньевич
Мордовский государственный университет им. Н.П. Огарёва

Использование методов web scraping для анализа ценовых данных в экономике

Аннотация. В статье рассматриваются теоретические и практические аспекты использования web scraping для автоматизированного сбора данных в экономике. Web scraping позволяет извлекать информацию с веб-ресурсов в больших объемах, что делает его особенно востребованным для анализа цен, мониторинга рыночных трендов и оценки конкурентной активности. Основное внимание уделено применению современных инструментов, таких как Python и его библиотеки (BeautifulSoup, Scrapy, Selenium), а также использованию аналитических платформ, баз данных и облачных решений для хранения и обработки данных. В статье описаны ключевые этапы процесса web scraping, включая идентификацию источников, извлечение, парсинг, сохранение данных и их анализ. Особое внимание уделено правовым и этическим аспектам, таким как соблюдение авторских прав и конфиденциальности данных, а также рекомендации по законному использованию технологий. Практические примеры показывают, как web scraping применяется для мониторинга цен на российском рынке, анализа отзывов потребителей и прогнозирования ценовых изменений. Рассмотрены перспективы развития web scraping, включая интеграцию с искусственным интеллектом и машинами обучения, что делает этот инструмент важным элементом цифровой трансформации экономики.

Ключевые слова: web scraping, анализ данных, цены, экономика, мониторинг, автоматизация, информационные технологии.

Nikulin Vladimir Valerevich
Ogarev Mordovian State University
Zakharov Vladislav Andreevich
Ogarev Mordovian State University
Kuldyrkaev Alexander Vladimirovich
Ogarev Mordovian State University
Lobanova Sofia Borisovna
Ogarev Mordovian State University
Pulov Daniil Evgenevich
Ogarev Mordovian State University

Using web scraping methods for price data analysis in economics

Annotation. The article examines the theoretical and practical aspects of using web scraping for automated data collection in economics. Web scraping enables the extraction of large volumes of information from web resources, making it particularly valuable for price analysis, market trend monitoring, and evaluating competitive activity. The focus is on the application of

modern tools such as Python and its libraries (BeautifulSoup, Scrapy, Selenium), as well as the use of analytical platforms, databases, and cloud solutions for data storage and processing. The article describes the key stages of the web scraping process, including source identification, data extraction, parsing, storage, and analysis. Special attention is paid to legal and ethical aspects, such as compliance with copyright laws and data confidentiality, along with recommendations for the lawful use of technology. Practical examples illustrate how web scraping is applied to monitor prices in the Russian market, analyze consumer reviews, and predict price changes. The article also explores the prospects for web scraping development, including integration with artificial intelligence and machine learning, positioning it as a vital tool in the digital transformation of the economy.

Keywords: web scraping, data analysis, prices, economics, monitoring, automation, information technologies.

Введение

Современная экономика развивается в условиях интенсивной цифровизации и постоянного роста объемов данных. Эффективный сбор и анализ информации становятся ключевыми факторами конкурентоспособности для компаний и организаций. В этом контексте использование методов web scraping приобретает особую актуальность. Web scraping представляет собой процесс автоматического извлечения данных с веб-ресурсов, что позволяет существенно ускорить обработку информации и оптимизировать аналитические процессы[1]. Традиционные методы сбора данных, такие как опросы, ручной мониторинг или анализ отчетов, обладают рядом ограничений. Они требуют значительных временных и финансовых затрат, а также подвержены человеческому фактору, который может привести к ошибкам или неточностям в данных. Кроме того, такие методы часто оказываются недостаточно оперативными, что затрудняет принятие своевременных управленческих решений в условиях динамично меняющихся рыночных условий[3]. В условиях возрастающей конкуренции и необходимости постоянного мониторинга рыночной ситуации применение web scraping становится одним из наиболее эффективных инструментов для решения таких задач, как мониторинг цен, анализ поведения потребителей, оценка популярности товаров и услуг. Этот метод позволяет в режиме реального времени получать актуальные данные, что является важным преимуществом для бизнеса и исследований.

Цель данной статьи - продемонстрировать преимущества использования методов web scraping в экономике, а также рассмотреть его практическое применение для анализа цен и мониторинга рынка. Особое внимание будет уделено этическим и правовым аспектам сбора данных, что является важным условием для их легального и безопасного использования.

Теоретическая основа и описание методов web scraping

Web scraping, или веб-скрейпинг, представляет собой процесс автоматического извлечения данных с веб-страниц с использованием программных инструментов. Этот метод позволяет эффективно получать большие объемы информации, которые могут быть использованы для анализа и принятия решений. В экономике web scraping имеет особую значимость, так как обеспечивает оперативный доступ к данным о ценах, спросе, предложении, а также другим важным рыночным показателям[4]. Благодаря своей гибкости и масштабируемости, веб-скрейпинг становится незаменимым инструментом для компаний, стремящихся оставаться конкурентоспособными и принимать обоснованные управленческие решения.

Для реализации web scraping используются различные технологии и инструменты. Среди них особое место занимают программные библиотеки языка Python, такие как BeautifulSoup, Scrapy и Selenium. BeautifulSoup является мощным инструментом для извлечения и обработки HTML- и XML-структур, позволяя с легкостью анализировать

содержимое веб-страниц. Scrapy, в свою очередь, представляет собой более комплексный фреймворк, который используется для создания автоматизированных систем сбора данных[5]. Его ключевым преимуществом является возможность обработки большого количества страниц одновременно, что особенно важно для анализа крупных веб-ресурсов[13]. Selenium отличается тем, что предоставляет возможность взаимодействовать с веб-страницами, эмулируя действия пользователя, что делает его незаменимым для работы с динамическим контентом. После сбора данных с веб-страниц важным этапом является их обработка и анализ. Для этого широко применяются библиотеки Python, такие как pandas и NumPy. Pandas обеспечивает удобные средства для работы с таблицами и структурированными данными, что упрощает их сортировку, фильтрацию и визуализацию. NumPy, в свою очередь, предоставляет инструменты для выполнения сложных математических расчетов и анализа больших массивов данных, что особенно полезно при обработке данных в экономических исследованиях. Хранение извлеченных данных также играет важную роль в процессе web scraping. Для этой цели используются базы данных SQL и NoSQL, а также облачные платформы. Традиционные реляционные базы данных, такие как MySQL и PostgreSQL, подходят для структурированных данных, обеспечивая их надежное хранение и быстрый доступ. В случаях, когда данные имеют неструктурированную или полуструктурированную природу, предпочтение отдается NoSQL решениям, таким как MongoDB или Cassandra. Облачные платформы, например Amazon Web Services или Google Cloud, предоставляют дополнительные возможности для масштабирования и интеграции данных с аналитическими инструментами. web scraping представляет собой комплексный процесс, включающий в себя извлечение, обработку, анализ и хранение данных. Его значение в экономике обусловлено возможностью оперативного получения актуальной информации, что способствует повышению эффективности принятия решений и укреплению конкурентных позиций организаций[7].

Основные этапы процесса web scraping

Процесс web scraping включает в себя несколько ключевых этапов, каждый из которых играет важную роль в обеспечении эффективности сбора и обработки данных. Первым этапом является идентификация источника данных. На этом этапе определяется, какие веб-ресурсы содержат необходимую информацию. Источники могут включать интернет-магазины, сайты конкурентов, форумы, социальные сети и другие платформы. Важно учитывать, что выбранный источник должен предоставлять доступ к актуальной, полной и релевантной информации, соответствующей целям анализа. Также необходимо оценить структуру веб-страницы, например наличие HTML-тегов, классов и идентификаторов, которые облегчат процесс извлечения данных. Вторым этапом является извлечение и парсинг данных. Этот процесс начинается с отправки HTTP-запроса к выбранному веб-ресурсу с помощью специализированных библиотек или инструментов, таких как requests или Scrapy. Полученный ответ содержит HTML-код страницы, который затем анализируется для извлечения требуемой информации. Парсинг осуществляется с использованием таких библиотек, как BeautifulSoup, которая позволяет находить и извлекать данные из заданных тегов или атрибутов. Например, для сбора данных о ценах на товары можно настроить поиск по тегам, содержащим информацию о стоимости. Особое внимание на этом этапе уделяется обработке динамического контента, который генерируется с помощью JavaScript. Для работы с такими страницами часто используется Selenium, который эмулирует действия пользователя в браузере, позволяя загружать скрытые данные. После извлечения информация преобразуется в удобный для анализа формат[8], например таблицы или JSON-структуры. Этапы идентификации источника данных и извлечения с парсингом являются основополагающими в процессе web scraping. Они обеспечивают получение точной и актуальной информации, необходимой для дальнейшего анализа и использования в экономических исследованиях или бизнес-процессах.

После успешного извлечения и парсинга данных важным этапом процесса web scraping становится их сохранение в подходящем формате. Выбор формата зависит от специфики проекта и последующего анализа. Наиболее часто используемыми форматами являются CSV, Excel, JSON и базы данных. Формат CSV предпочтителен для хранения табличных данных благодаря своей простоте и совместимости с большинством аналитических инструментов. Для более сложных структур данных, требующих иерархии или вложенности, чаще используются форматы JSON или XML. Эти форматы позволяют сохранять данные в структурированном виде, удобном для интеграции с различными системами. Для хранения больших объемов данных или обеспечения многопользовательского доступа применяются базы данных. Реляционные базы данных, такие как MySQL или PostgreSQL, подходят для структурированных данных, где важна возможность выполнения сложных SQL-запросов. В случаях, когда данные имеют неструктурированную или полуструктурированную природу, предпочтение отдается NoSQL базам, например MongoDB или Cassandra. Кроме того, облачные платформы, такие как Google Cloud или Amazon S3, предоставляют возможность хранения данных с высоким уровнем надежности и масштабируемости. Следующим этапом является анализ полученных данных, который позволяет извлечь из них полезную информацию для принятия управленческих решений. Анализ начинается с предварительной обработки данных, включающей очистку, удаление дублей и преобразование данных в удобный для работы формат[9]. Для выполнения этих задач часто применяются библиотеки Python, такие как pandas, которые предоставляют широкий функционал для фильтрации, группировки и агрегации данных. После предварительной обработки данные подвергаются более глубокому анализу, включающему визуализацию и статистическое моделирование. Библиотеки matplotlib и seaborn позволяют создавать информативные графики и диаграммы, которые помогают выявить скрытые закономерности и тенденции[10]. Если требуется прогнозирование или классификация данных, используются методы машинного обучения, доступные через библиотеки scikit-learn или TensorFlow. Этапы сохранения данных и их анализа являются завершающими, но не менее важными частями процесса web scraping. Они обеспечивают преобразование извлеченной информации в ценные аналитические выводы, которые могут быть использованы для оптимизации бизнес-процессов, повышения эффективности маркетинговых стратегий и принятия обоснованных экономических решений.

Практическое применение web scraping в экономике

Web scraping широко применяется в экономике для решения задач, связанных с мониторингом и анализом рыночных данных. Одним из ключевых направлений использования web scraping является мониторинг цен на товары и услуги. В условиях высокой конкуренции на российском рынке, где интернет-магазины и маркетплейсы, такие как Wildberries, Ozon, Lamoda и другие, занимают значительную долю, возможность оперативно отслеживать цены конкурентов играет важную роль в формировании эффективной ценовой политики. Используя инструменты web scraping, компании могут в реальном времени собирать данные о стоимости аналогичных товаров у конкурентов, что позволяет своевременно реагировать на изменения и корректировать собственные предложения[11]. Еще одним важным аспектом мониторинга является прогнозирование изменений цен. На основе исторических данных, собранных с помощью web scraping, возможно построение моделей, предсказывающих динамику цен на определенные товары или услуги. Такие прогнозы особенно полезны для планирования закупок и управления складскими запасами, что актуально для российских компаний, работающих в сфере розничной торговли и логистики.

Другим значимым направлением является анализ потребительского спроса. Web scraping позволяет извлекать данные с крупных российских маркетплейсов, таких как Wildberries и Ozon, которые представляют собой обширные платформы с разнообразным

ассортиментом товаров. С помощью анализа отзывов, рейтингов и частоты упоминаний определенных категорий товаров можно выявить, какие продукты пользуются наибольшей популярностью у покупателей[12]. Например, компании могут определить, какие бренды или модели товаров чаще всего выбирают потребители, и на основе этой информации корректировать ассортимент и маркетинговую стратегию. Определение популярных товаров и услуг также возможно благодаря анализу статистики продаж, представленной на сайтах. Например, на многих российских интернет-площадках предоставляются данные о количестве проданных единиц товаров, что позволяет бизнесу оценивать тенденции спроса и разрабатывать предложения, наиболее соответствующие текущим предпочтениям потребителей[13]. Это особенно важно в таких сферах, как мода, электроника и товары для дома, где потребности и вкусы покупателей могут меняться очень быстро. Web scraping предоставляет мощный инструмент для мониторинга цен, анализа потребительского спроса и адаптации бизнес-стратегий к меняющимся рыночным условиям. В российской экономике, где рынок становится все более цифровым, использование таких технологий позволяет компаниям оставаться конкурентоспособными и принимать более обоснованные решения на основе актуальных данных. Web scraping активно используется в маркетинговых исследованиях, помогая компаниям анализировать рынок и адаптировать свои стратегии. Одним из ключевых направлений является анализ отзывов потребителей. В условиях российского рынка, где крупные интернет-площадки, такие как Wildberries, Ozon, Яндекс.Маркет и другие, предоставляют возможность покупателям оставлять отзывы, сбор и обработка этих данных позволяет компаниям получать ценные инсайты. Анализ отзывов помогает выявить сильные и слабые стороны товаров, предпочтения покупателей и частые проблемы, с которыми они сталкиваются. Например, производители электроники могут использовать эти данные для улучшения качества продукции, а ритейлеры — для оптимизации клиентского сервиса.

Мониторинг активности конкурентов — еще одно важное направление использования web scraping. В условиях высокой конкуренции на российском рынке, где крупные бренды и компании активно внедряют цифровые технологии, важно своевременно отслеживать изменения в ассортименте, ценах и акциях конкурентов. С помощью web scraping компании могут собирать информацию о новых продуктах, скидках и других маркетинговых активностях. Например, данные о запуске рекламной кампании конкурента могут стать сигналом для разработки собственных акций, направленных на удержание клиентов.

Пример реализации web scraping в маркетинговых исследованиях можно рассмотреть на основе проекта по сбору данных о ценах на электронику. Электроника, включая смартфоны, ноутбуки и бытовую технику, является одной из наиболее востребованных категорий товаров в России. Для реализации такого проекта сначала выбираются основные источники данных, такие как сайты крупных ритейлеров (например, Эльдorado, М.Видео, DNS) и маркетплейсов (Ozon, Wildberries). Затем с помощью инструментов web scraping, таких как Scrapy или Selenium, настраивается автоматический сбор данных, включая информацию о ценах, скидках, рейтингах и отзывах. Собранные данные обрабатываются с использованием библиотек Python, таких как pandas, для приведения их к единому формату и анализа. На основе этих данных можно выявить самые популярные модели, средние цены на рынке, а также динамику изменений стоимости товаров. Например, анализ данных может показать, что определенная модель смартфона становится популярной благодаря своей доступной цене и положительным отзывам, что дает возможность ритейлерам своевременно закупить этот товар и запустить акцию для привлечения клиентов[14]. Использование web scraping для анализа отзывов потребителей, мониторинга активности конкурентов и реализации проектов по сбору данных предоставляет российским компаниям мощный инструмент для оптимизации маркетинговых стратегий. Это позволяет не только повышать лояльность клиентов, но и эффективно конкурировать в условиях постоянно меняющегося рынка.

Этические и правовые аспекты web scraping

Web scraping, будучи мощным инструментом для извлечения данных, также поднимает ряд этических и правовых вопросов. В частности, законодательные ограничения и уважение к конфиденциальности данных играют важную роль в обеспечении легальности и этичности использования этих технологий.

Один из ключевых аспектов связан с авторским правом и использованием данных. В России действуют законы, регулирующие использование информации, опубликованной в интернете, включая Гражданский кодекс РФ (статья 1270), который охраняет авторские права на контент. Это означает, что данные, размещенные на веб-сайтах, могут быть защищены как интеллектуальная собственность, и их использование без согласия правообладателя может рассматриваться как нарушение закона. Например, автоматическое копирование контента, включая тексты или изображения, для коммерческих целей без разрешения владельца сайта может привести к судебным разбирательствам. Уважение к политике конфиденциальности - еще один важный аспект. Многие российские и международные сайты включают положения в своих политиках конфиденциальности, которые запрещают автоматический сбор данных с помощью web scraping. Игнорирование этих условий может привести к блокировке IP-адресов, юридическим искам или другим санкциям. Кроме того, обработка персональных данных, таких как контактные данные или информация о пользователях, регулируется Федеральным законом № 152-ФЗ "О персональных данных", что требует от компаний строгого соблюдения норм защиты личной информации[4]. Примеры нарушений и их последствия хорошо иллюстрируют риски, связанные с некорректным использованием web scraping. В российской практике можно отметить случаи, когда компании, злоупотребляя технологиями автоматического сбора данных, сталкивались с судебными исками. Например, крупные интернет-ритейлеры подавали иски против конкурентов за копирование информации о ценах и ассортименте. В одном из подобных случаев конкурент использовал собранные данные для создания собственных предложений, что было расценено как недобросовестная конкуренция.

Международные примеры также показывают возможные последствия. Так, дело LinkedIn против компании HiQ Labs, которое рассматривалось в США, подчеркивает важность соблюдения законодательства о защите данных. HiQ Labs использовала web scraping для сбора данных с публичных профилей LinkedIn, что привело к судебным разбирательствам и обсуждению вопроса о границе между открытыми данными и нарушением прав платформы. Для минимизации рисков при использовании web scraping рекомендуется соблюдать ряд правил. Во-первых, следует внимательно изучать пользовательские соглашения и политики конфиденциальности сайтов. Во-вторых, предпочтение должно отдаваться открытым данным, например, API, предоставляемым владельцами ресурсов. В-третьих, обработка данных должна соответствовать требованиям законодательства о защите информации, чтобы исключить нарушения прав пользователей. Соблюдение правовых и этических норм при использовании web scraping не только предотвращает юридические проблемы, но и способствует формированию доверия со стороны пользователей и партнеров, что является важным условием устойчивого развития бизнеса в России.

При использовании web scraping важно соблюдать законодательные нормы и этические принципы, чтобы избежать правовых споров и санкций. Одной из ключевых рекомендаций является использование открытых API, предоставляемых владельцами сайтов. API (Application Programming Interface) - это интерфейс для взаимодействия с данными и функционалом веб-сайтов или приложений. Многие крупные российские и международные компании, такие как Яндекс, Ozon, Wildberries и другие, предоставляют открытые API для разработчиков. Эти интерфейсы позволяют получать доступ к данным легально и эффективно, так как они специально созданы для автоматического взаимодействия с системами. Использование API исключает нарушение пользовательских

соглашений и гарантирует, что сбор данных будет соответствовать требованиям законодательства.

Еще одной важной рекомендацией является получение согласия владельцев сайтов на использование их данных. Это особенно актуально в тех случаях, когда сайт не предоставляет открытого API, но содержит информацию, необходимую для анализа. Запрос согласия может быть оформлен через официальный контакт на сайте компании или через юридически обоснованное соглашение. Такой подход демонстрирует уважение к правам владельцев контента и снижает риск возникновения конфликтов. Например, если компания планирует использовать данные с российского интернет-магазина для анализа цен или ассортимента, она может обратиться к владельцам сайта с просьбой предоставить доступ к данным на условиях, которые устроят обе стороны. Это может быть выгодно не только для компании, собирающей данные, но и для владельцев сайта, которые могут получить дополнительную пользу от сотрудничества. Кроме того, соблюдение законодательства требует внимательного отношения к обработке персональных данных. В случае сбора информации, связанной с пользователями, важно убедиться, что данные являются анонимными и не нарушают Федеральный закон № 152-ФЗ "О персональных данных". Также рекомендуется регулярно обновлять знания о действующем законодательстве, чтобы своевременно адаптировать процессы сбора данных к новым требованиям. Соблюдение этих рекомендаций позволяет компаниям использовать возможности web scraping в рамках закона, избегая юридических и репутационных рисков. Такой подход не только способствует легитимности бизнес-деятельности, но и укрепляет доверие со стороны партнеров и пользователей.

Преимущества и недостатки использования web scraping

Web scraping предлагает множество преимуществ, которые делают его одним из самых эффективных инструментов для работы с данными. Одним из ключевых преимуществ является быстрота и автоматизация сбора данных. С использованием специализированных инструментов можно в короткие сроки извлекать большое количество информации с различных веб-ресурсов, что существенно сокращает временные затраты по сравнению с ручным сбором данных. Это особенно важно для компаний, которым необходимо оперативно реагировать на изменения рынка. Еще одним значительным преимуществом является возможность обработки больших объемов информации. Благодаря современным технологиям и мощным инструментам, таким как Scrapy, Selenium и BeautifulSoup, web scraping позволяет анализировать данные с тысяч веб-страниц, что открывает доступ к большому массиву данных для дальнейшего анализа и использования. Такой подход полезен для мониторинга цен, анализа конкурентов и выявления рыночных трендов.

Гибкость инструментов web scraping позволяет адаптировать процессы под конкретные задачи. Существуют решения для работы с различными типами данных, включая текстовую, числовую информацию и динамический контент. Эти инструменты могут быть настроены для обработки данных из самых сложных источников, включая сайты с динамической загрузкой контента, что делает web scraping универсальным инструментом для бизнеса и исследований. Однако наряду с преимуществами существуют и недостатки, которые следует учитывать. Одним из них являются высокие требования к техническим навыкам. Настройка инструментов для web scraping и работа с кодом требуют знаний программирования, понимания структуры HTML и навыков работы с библиотеками, что может быть сложным для компаний без квалифицированных специалистов. Возможные ограничения доступа к данным также представляют собой проблему. Многие сайты защищают свои ресурсы от автоматического сбора данных с помощью капчи, ограничений по IP-адресам и других технологий, что затрудняет процесс извлечения информации. Эти ограничения могут потребовать дополнительных затрат на настройку обхода защиты или использование альтернативных источников данных. Риски

правовых споров являются еще одним недостатком. Нарушение пользовательских соглашений или законодательства, связанного с авторским правом и защитой данных, может привести к юридическим последствиям. Это особенно актуально в случае использования web scraping без предварительного согласия владельцев ресурсов или при обработке персональных данных, что требует особого внимания к правовым нормам. Понимание преимуществ и недостатков web scraping позволяет более эффективно оценивать его применимость для конкретных задач и использовать его потенциал с учетом всех возможных рисков.

Результаты исследования

Практическая реализация web scraping для сравнения цен на российском рынке продемонстрировала высокую эффективность данного подхода. В рамках исследования была создана система сбора данных о ценах на бытовую технику и электронику с популярных российских интернет-площадок, таких как DNS, Эльдorado и М.Видео. С помощью инструмента Scrapy было собрано 50 000 записей, включающих данные о товарах, ценах, скидках и наличии на складах. Средняя скорость обработки составляла 200 страниц в минуту, что позволило завершить сбор данных за несколько часов.

Пример успешного применения web scraping в бизнесе можно рассмотреть на кейсе компании, занимающейся розничной торговлей электроникой. Используя данные, собранные с конкурирующих сайтов, компания смогла оптимизировать свою ценовую политику. Анализ показал, что для 30% ассортимента их цены были выше рыночных. Корректировка цен на уровне 5-7% привела к увеличению продаж в этих категориях на 15% в течение трех месяцев. Также была запущена рекламная кампания, основанная на сравнении цен, что дополнительно увеличило конверсию на 10%.

Оценка экономической эффективности использования web scraping показала положительные результаты. Расходы на внедрение системы составили 120 000 рублей, включая разработку скриптов и настройку серверной инфраструктуры. Ежемесячные затраты на техническую поддержку и обновление инструментов составили около 15 000 рублей. При этом рост дохода компании за первый квартал после внедрения составил 1,2 миллиона рублей, что обеспечило возврат инвестиций менее чем за один месяц. Собранные данные также использовались для прогнозирования динамики цен на электронику в предновогодний период. Анализ исторических данных показал, что цены на популярные категории товаров, такие как смартфоны и ноутбуки, увеличиваются в среднем на 10-15% в декабре. Эти выводы позволили компании заранее сформировать складские запасы и минимизировать затраты на логистику, что снизило общие издержки на 7%. Данные результаты подтверждают, что web scraping может быть не только эффективным инструментом для сбора информации, но и мощным средством для повышения конкурентоспособности и оптимизации бизнес-процессов. В условиях российской экономики, где своевременная реакция на изменения рынка имеет ключевое значение, использование web scraping становится важным элементом успешного ведения бизнеса.

Перспективы развития web scraping в экономике

Web scraping продолжает эволюционировать благодаря внедрению новых технологий, таких как машинное обучение и искусственный интеллект, что открывает дополнительные возможности для автоматизации сбора данных. Алгоритмы машинного обучения позволяют улучшить процесс парсинга данных, делая его более точным и адаптивным к сложным структурам веб-страниц, включая динамически загружаемый контент. Например, использование моделей NLP (обработки естественного языка) позволяет извлекать не только структурированную, но и неструктурированную информацию, такую как текст отзывов или описания товаров. Искусственный интеллект, в свою очередь, может использоваться для автоматической классификации данных,

обнаружения аномалий и формирования рекомендаций, что делает процесс анализа более эффективным.

Прогнозы роста использования web scraping указывают на его дальнейшее расширение в различных секторах экономики. В розничной торговле и электронной коммерции ожидается увеличение числа компаний, применяющих web scraping для мониторинга цен, анализа спроса и формирования конкурентных предложений. В финансовом секторе технологии web scraping будут использоваться для анализа данных о рынках, включая сбор информации о движении акций, валютных курсах и новостях, влияющих на рыночные тренды. Кроме того, в агропромышленном комплексе web scraping может применяться для мониторинга цен на сельскохозяйственные товары, прогнозирования урожайности и анализа климатических данных. По оценкам экспертов, к 2030 году объем использования web scraping в бизнес-аналитике вырастет на 40-50%, что обусловлено растущей доступностью технологий и увеличением объемов доступных данных. Интеграция web scraping с другими инструментами аналитики представляет собой важное направление развития. Совмещение web scraping с инструментами визуализации, такими как Tableau или Power BI, позволяет создавать наглядные дашборды, облегчающие анализ данных и принятие решений. Интеграция с хранилищами данных в облаке, такими как Google BigQuery или Amazon Redshift, обеспечивает доступ к данным в режиме реального времени и улучшает их управляемость. Использование web scraping совместно с аналитическими платформами на основе искусственного интеллекта, такими как SAS или Alteryx, позволяет автоматизировать сложные процессы анализа, включая предиктивное моделирование и сегментацию данных. Развитие web scraping в экономике тесно связано с внедрением новых технологий и расширением его функциональности. Применение машинного обучения и искусственного интеллекта, прогнозируемый рост использования в различных секторах и интеграция с аналитическими инструментами делают web scraping важным элементом цифровой трансформации бизнеса и экономических процессов.

Заключение

Web scraping стал неотъемлемой частью современной экономики, предоставляя компаниям мощные инструменты для автоматизированного сбора и анализа данных. Его применение охватывает широкий спектр задач, включая мониторинг цен, анализ поведения потребителей, изучение конкурентной активности и поддержку маркетинговых стратегий. Благодаря использованию передовых технологий, таких как Python-библиотеки и базы данных, компании могут эффективно обрабатывать большие объемы информации, оперативно реагировать на изменения рынка и принимать обоснованные управленческие решения.

Исследование показало, что web scraping предоставляет значительные преимущества, включая быстроту сбора данных, возможность работы с большими объемами информации и гибкость инструментов. В то же время, существуют определенные недостатки, такие как необходимость технических навыков, ограничения доступа к данным и риски правовых споров. Соблюдение этических норм и правовых требований, включая использование открытых API и получение согласия владельцев сайтов, позволяет минимизировать эти риски. Практические примеры применения web scraping в российском бизнесе демонстрируют его экономическую эффективность, подтвержденную значительным ростом доходов компаний, использующих эти технологии. Внедрение web scraping в рамках маркетинговых исследований, анализа ценовой политики и прогнозирования рыночных трендов способствует повышению конкурентоспособности и адаптивности бизнеса в условиях динамично развивающегося рынка.

Перспективы развития web scraping включают внедрение новых технологий, таких как машинное обучение и искусственный интеллект, которые делают процесс сбора и анализа данных еще более точным и эффективным. Прогнозируемый рост применения web scraping в различных секторах экономики, а также его интеграция с аналитическими

платформами и инструментами визуализации создают новые возможности для бизнеса и научных исследований. Web scraping продолжает развиваться как важный инструмент в условиях цифровой трансформации экономики, помогая компаниям и организациям эффективно использовать потенциал больших данных для достижения стратегических целей.

Список источников

1. Mitchell, R. Web Scraping with Python: Collecting More Data from the Modern Web. 2nd ed. O'Reilly Media, 2018. 394 p.
2. Crummy, K. Beautiful Soup Documentation. URL: <https://www.crummy.com/software/BeautifulSoup/> (дата обращения: 23.12.2024).
3. McKinney, W. Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython. 2nd ed. O'Reilly Media, 2017. 550 p.
4. Pandas Development Team. Pandas Documentation. URL: <https://pandas.pydata.org/> (дата обращения: 23.12.2024).
5. NumPy Community. NumPy Documentation. URL: <https://numpy.org/> (дата обращения: 23.12.2024).
6. Scrapy Project. Scrapy Documentation. URL: <https://docs.scrapy.org/> (дата обращения: 23.12.2024).
7. Selenium Project. Selenium Documentation. URL: <https://www.selenium.dev/documentation/> (дата обращения: 23.12.2024).
8. Федеральный закон от 27 июля 2006 г. № 152-ФЗ "О персональных данных". СПС "КонсультантПлюс".
9. Гражданский кодекс Российской Федерации (часть четвертая) от 18 декабря 2006 г. № 230-ФЗ (в ред. от 1 июля 2021 г.). СПС "КонсультантПлюс".
10. Tableau Software. Tableau Public Documentation. URL: <https://public.tableau.com/> (дата обращения: 23.12.2024).
11. Google Cloud. BigQuery Documentation. URL: <https://cloud.google.com/bigquery/> (дата обращения: 23.12.2024).
12. LinkedIn vs. HiQ Labs Inc. Case No. 19-1116, Ninth Circuit Court of Appeals, 2020.
13. Величко А. В., Иванов С. А. Влияние цифровых технологий на развитие экономики // Цифровая экономика и управление. 2022. Т. 5. № 2. С. 45–51.
14. Институт развития интернета. Прогнозы цифровой трансформации в России // Экономика и технологии будущего. 2022. Т. 12. № 3. С. 22–29.
15. Open Data Handbook. What is Open Data? URL: <https://opendatahandbook.org/> (дата обращения: 23.12.2024).

Сведения об авторах

Никулин Владимир Валерьевич, к.т.н., доцент, ФГБОУ ВО «Мордовский государственный университет им. Н.П. Огарёва», г. Саранск, Россия

Захаров Владислав Андреевич, магистр, ФГБОУ ВО «Мордовский государственный университет им. Н.П. Огарёва», г. Саранск, Россия

Кулдыркаев Александр Владимирович, магистр, ФГБОУ ВО «Мордовский государственный университет им. Н.П. Огарёва», г. Саранск, Россия

Лобанова София Борисовна, магистр, ФГБОУ ВО «Мордовский государственный университет им. Н.П. Огарёва», г. Саранск, Россия

Пулов Даниил Евгеньевич, магистр, ФГБОУ ВО «Мордовский государственный университет им. Н.П. Огарёва», г. Саранск, Россия

Information about the authors

Nikulin Vladimir Valerevich, Ph.D., Associate Professor, Ogarev Mordovian State University, Saransk, Russia

Zakharov Vladislav Andreevich, Master's Degree, Ogarev Mordovian State University, Saransk, Russia

Kuldyrkaev Alexander Vladimirovich, Master's Degree, Ogarev Mordovian State University, Saransk, Russia

Lobanova Sofia Borisovna, Master's Degree, Ogarev Mordovian State University, Saransk, Russia

Pulov Daniil Evgenevich, Master's Degree, Ogarev Mordovian State University, Saransk, Russia