

Технологии информационного общества**СБОР ИНТЕРНЕТ-ДАННЫХ: ПОТЕНЦИАЛ ДЛЯ БИЗНЕСА И
РОБОТИЗИРОВАННАЯ АВТОМАТИЗАЦИЯ ПРОЦЕССА**

Статья рекомендована к публикации членом редакционного совета К. К. Колиным 17.09.2025.

Гадасина Людмила Викторовна

*Кандидат физико-математических наук
Санкт-Петербургский государственный университет
Кафедра информационных систем в экономике
Доцент
Санкт-Петербург, Российская Федерация
l.gadasina@spbu.ru*

Иванова Виктория Валерьевна

*Кандидат экономических наук
Санкт-Петербургский государственный университет
Кафедра информационных систем в экономике
Доцент
Санкт-Петербург, Российская Федерация
v.ivanova@spbu.ru*

Зырянова Дарья Александровна

*Санкт-Петербургский государственный университет
Магистрант направления подготовки «Бизнес-информатика»
Санкт-Петербург, Российская Федерация
darastarovatova@gmail.com*

Аннотация

Статья посвящена вопросам автоматизации сбора данных в интернете. Систематизированы типы источников и особенности сбора данных с них. Представлен обзор современных технологий сбора данных, особое внимание уделено технологии роботизированной автоматизации процессов Robotic Process Automation (RPA). Приведен практический кейс, иллюстрирующий конкретные шаги по созданию автоматизированных процессов сбора данных на популярных маркетплейсах Wildberries и Ozon для логистической компании.

Ключевые слова

технология RPA; интернет-данные; скрапинг

Введение

В настоящее время у бизнеса имеется понимание, что информация является основой их деятельности. Сырьем для получения информации являются данные, которые нужно уметь получать, сохранять и обрабатывать. Работа с внутренними данными компании подчиняется правилам компаний и их регламентам в области управления данными. Внутренние данные являются базой регламентированной отчетности и, естественно, аналитики компании.

© Гадасина Л. В., Иванова В. В., Зырянова Д. А., 2026

Производство и хостинг журнала «Информационное общество» осуществляется Институтом развития информационного общества.

Данная статья распространяется на условиях международной лицензии Creative Commons «С указанием авторства - С сохранением условий версии 4.0 Международная» (Creative Commons Attribution – ShareAlike 4.0 International; CC BY-SA 4.0). См. <https://creativecommons.org/licenses/by-sa/4.0/legalcode.ru>

https://doi.org/10.52605/16059921_2026_04_147

Однако, собственных данных недостаточно для решения всех управленческих задач компании, в том числе таких как состояние конкурентной среды, разработка сценариев развития компании, управление инвестициями и оценка рисков, оптимизация запасов и логистика, прогнозирование показателей (например, продаж и закупок), анализ спроса и удовлетворенности клиентов, управление рекламой, развитие персонала.

В соответствии с рекомендацией по стандартизации Р 50.1.089- 2014 «Использование внешних источников данных» источниками внешних данных по отношению к организации являются факторы ближнего и дальнего окружения. Внешние источники данных с содержательной (семантический) точки зрения могут быть экономическими, политическими, рыночными, технологическими, природно-климатическими, информацией причастных сторон [6].

С организационно-технологической точки зрения можно идентифицировать следующие типы источников:

1. Опросы и интервью, проводимые компанией или по ее заказу. Здесь можно выделить такие особенности сбора данных: большие организационные, временные затраты [21] и часто низкая мотивация респондентов [9].
2. Сайты и веб-страницы. В настоящее время данные ресурсы являются дешевым и неформальным источником данных. На сайтах компаний публикуется обширный для анализа конкурентной среды объем сведений, в том числе данные о продуктах и разработках, партнерствах в проектах и другие. В настоящее время существует большое количество готовых решений сбора таких данных [4]. Алгоритмы постоянно совершенствуются и являются гибкими и адаптируемыми, часто несложными для самостоятельной разработки. К недостаткам можно отнести возможное противоречие законодательным нормам, сбои в корректной работе при изменении структуры сайта.
3. Социальные сети. Социальные сети — серьезный источник данных для аналитических задач различных направлений: маркетинг, социология, политология. В настоящее время алгоритмы и программы сбора и обработки данных хорошо развиты, но есть серьезные ограничения, такие как конфиденциальная политика самих сетей, законодательные ограничения, не всегда доступные качественные программы-парсеры.
4. Данные партнеров и конкурентов, в том числе специализированные платформы, такие как маркетплейсы. На таких ресурсах доступны данные об ассортименте товаров (или услуг) компаний, цены, акции и скидки, доступные остатки. Сбор таких данных позволяет получить базу для решения задач ценообразования, оптимизации предложений собственных товаров [5]. Второй группой данных, собираемой с маркетплейсов, являются отзывы клиентов как на товары (или услуги) конкурентов, так и на собственный ассортимент. Анализ отзывов и оценок дают возможность выявить слабые места продукта или услуги, отслеживать настроения клиентов. Можно выделить следующие проблемы при сборе данных с маркетплейсов: динамические обновления приводят к сбоям в работе программ-парсеров, механизмы защиты от парсеров со стороны маркетплейсов, сложные форматы данных, правовые ограничения.
5. Государственные и коммерческие базы данных, в том числе патентные базы — это открытые данные, формируемые государственными или специализированными информационно-аналитическими структурами, а также частными компаниями. К ним, в том числе, относятся: базы данных объектов недвижимости (Росреестр — rosreestr.gov.ru, Циан — cian.ru), данные ресурсов проверки контрагентов (СПАРК, Контур.Фокус), патентные базы (База Федерального института промышленной собственности (ФИПС), PATENTSCOPE Всемирной организации интеллектуальной собственности (ВОИС)), научно-технические библиотеки (ГПНТБ России, ВИНТИ РАН, ИНИОН РАН), базы данных рабочих платформ (HeadHunter, SuperJob), данные транспортно-логистических платформ (НЦТЛП «ГосЛог», TRANS.RU, Roolz , «МегаФон-КАРГО»). При работе по сбору данных с каждой из платформ возникают специфические проблемы. Так, например, многие предлагаемые решения по сбору данных для патентных систем и научно-технических ресурсов могут иметь ограниченную функциональность, приводить к неполным или некорректным данным из-за изменений в форматах и структуре патентной информации, что способствует постоянному обновлению алгоритмов и механизмов систем парсинга для сбора информации патентных систем и научных библиотек [3].

1 Особенности сбора веб-данных

Можно выделить следующие особенности сбора веб-данных:

- данные могут оказаться недостоверными. Это происходит в случае, когда данные собираются из сомнительных (непроверенных) источников. Таким образом, необходима проверка источников данных;
- данные могут собираться незаметно для их поставщика [10];
- веб-ресурс может быть сложно устроен, что может затруднять сбор с него необходимой информации [13];
- данные могут собираться одноразово или регулярно с определенной, заданной бизнес-процессом периодичностью. Эти два подхода требуют разной стратегии сбора данных.

Если данные необходимо собрать одноразово, то возможно использовать ручной способ сбора. Такой способ повышает контролируемость процесса и снижает количество ошибок. Однако и в случае одноразового сбора может оказаться, что возникнет необходимость перерабатывать огромное количество информации. Это может отнимать много времени и снижать производительность сотрудников, задействованных в этом процессе [20, 23]. В таких случаях и в случаях, когда сбор данных необходимо осуществлять регулярно, возникает необходимость автоматизации такого процесса, ввиду его большой трудоемкости [4, 12, 23].

Одним из современных подходов является автоматический сбор данных с последующей их обработкой с помощью методов машинного обучения для упрощения процессов принятия решений [11].

Для автоматического сбора данных с веб-ресурсов используют:

- данные, предоставляемые ресурсами в машиночитаемых форматах, таких как CSV-файлы, JSON или XML;
- ключи API;
- парсинг (скрапинг), веб-сканирование;
- скрапинг с помощью технологий Robotic Process Automation (RPA).

Первый способ является самым простым и легко автоматизируемым. Кроме того, такой способ не требует сложной верификации качества данных.

Использование ключей API является самым удобным способом сбора данных сложной структуры. Однако, далеко не все ресурсы предоставляют такую возможность. Условия использования ключей API могут оказаться финансово невыгодными для тех, кто осуществляет сбор.

Веб-сканирование – это процедура автоматического извлечения данных с веб-сайтов с помощью программного обеспечения. Это технология, которая позволяет извлекать структурированные данные из текста, такого как HTML. Заметим, что следует различать понятия парсинг и скрапинг сайтов. Парсинг данных – это процесс преобразования собранных необработанных неструктурированных данных в структурированный формат, например JSON, XML или CSV. Понятие парсинг часто относится к лингвистическому анализу, например, разбору предложений на синтаксические компоненты [19]. В то же время веб-скрапинг определяется шире – как автоматизированный сбор данных с веб-страниц с использованием различных программных инструментов, за исключением методов, основанных на API [20]. Цель веб-скрапинга – преобразование неструктурированных данных и их сохранение в упорядоченных базах данных. Веб-скрапинг полезен в ситуациях, когда данные не предоставляются в веб-ресурсе в машиночитаемом формате [1, 2, 14]. К процедурам веб-скрапинга относятся программирование HTTP, парсинг DOM и HTML парсеры.

Бизнес всегда хочет оптимизировать время на свои процессы, в том числе на сбор данных. В этих условиях возможно применение технологий RPA, которые позволяют имитировать поведение пользователя [17, 22, 23].

RPA – это технология автоматизации бизнес-процессов с использованием программных роботов для выполнения рутинных, повторяющихся задач. Концепция RPA-технологии заключается в передаче программным роботам задач с высокой повторяемостью, которые просто стандартизировать. Заметим, что эта технология требует минимальных изменений в ИТ-инфраструктуре компании, так как роботы, имитирующие действия человека в цифровой среде не

требуют изменения существующих информационных систем, поэтому может быть внедрена быстро и относительно просто.

Сбор данных с помощью технологий RPA является более медленной процедурой, чем традиционный скрапинг, но тем не менее автоматизация позволяет высвободить время для других задач. Преимущества использования RPA — создает виртуальную рабочую силу, которая:

- автоматизирует стандартизированные задачи со 100%-ной точностью и последовательностью;
- может имитировать взаимодействие человека с клавиатурой и мышью в пользовательском интерфейсе, как если бы это делал настоящий человек. Это позволяет обходить защиту сайтов от автоматизированного скрапинга данных, которая включает отслеживание определенной активности на сайте и изменение структуры сайта.

Укрупненные этапы сбора данных с веб-ресурсов с помощью RPA:

1. Постановка поисковой задачи.
2. Оценка частоты потребности в данных.
3. Анализ структуры информации и выбор способа ее представления при выгрузке.
4. Первичный анализ ресурсов на наличие информации и ее видимость на ресурсе.
5. Проверка необходимости авторизации на ресурсе.
6. Разработка робота.
7. Запуск робота и сбор информации в файл.

В настоящее время в России активно развивается программное обеспечение для RPA. Согласно последнему опубликованному исследованию издания TAdviser [7] первой пятеркой крупнейших поставщиков RPA-систем в России по выручке за 2023 год являются PIX Robotics (Пикс Роботикс), Robin (Робин), Primo RPA (Примо РПА), Первый Бит и RAMAX Group.

2 Кейс: холодный поиск клиентов для компании

Особенности RPA-технологии для сбора данных с площадок одного типа, но разной организационной политики рассмотрим на примере практического кейса решения бизнес-задачи. В качестве инструментальной среды использовалась платформа PIX RPA. Платформа PIX входит в реестр отечественного ПО и отвечает требованиями государственных регуляторов.

2.1 Описание кейса

Заказчиком решения является франчайзи компании СДЭК — российского оператора транспортных услуг. Проблема: менеджеры по продажам вынуждены тратить много рабочего времени на поиск клиентов для заключения договорных отношений. Цель: оптимизация работы отдела продаж.

В компании выделяется два типа поиска клиентов: горячий и холодный. Горячий поиск осуществляется по внутренним данным о продажах компании. Холодный поиск — поиск юридических лиц, у которых со СДЭК нет договорных отношений. Холодный поиск предполагает работу с внешними источниками данных. В открытых интернет-ресурсах собираются данные о компаниях, индивидуальных предпринимателях. Необходимы следующие атрибуты: официальные наименования, ИНН, сфера деятельности. Проверяется наличие упоминаний о сотрудничестве с СДЭК в сфере доставки. Полученные сведения о клиентах проходят процедуру верификации в системе на предмет соответствия требованиям компании.

Источником для поиска компаний выбраны маркетплейсы Wildberries и Ozon. Выбор источника обусловлен тем, что компании, работающие через маркетплейсы, осуществляют доставку своего товара, а следовательно, потенциально могут быть клиентами компании СДЭК.

2.2 Сбор данных с маркетплейса Wildberries

Обобщенная схема сбора данных на маркетплейсе Wildberries:

- выбрать наименование товара;
- провести анализ структуры запроса в браузере (Google Chrome) в инструментах разработчика;
- идентифицировать API-запрос, обеспечивающий поиск товаров согласно введенному поисковому запросу;

- определить местонахождение уникального идентификатора продавца;
- осуществить переход на страницу продавца и извлечь необходимые данные из контактной информации.

Заметим, что на веб-странице продавца на маркетплейсе Wildberries имеется статический JSON-файл, содержащий полные данные этого продавца. Таким образом, для сбора данных создается двухэтапный алгоритм обработки. На первом этапе система последовательно обходит все страницы поисковой выдачи, извлекая идентификаторы продавцов. На втором этапе для каждого найденного идентификатора формируется соответствующий запрос к JSON-файлу, что позволяет получить структурированные данные о продавце. Такой подход обеспечивает автоматизированный сбор информации о всех продавцах, представленных в результатах поиска по заданному товарному запросу.

Для реализации алгоритма сбора данных может использоваться как парсер, так и сервисный робот. Укрупненная схема для сервисного робота, реализованная на платформе PIX RPA, представлена на рис. 1.

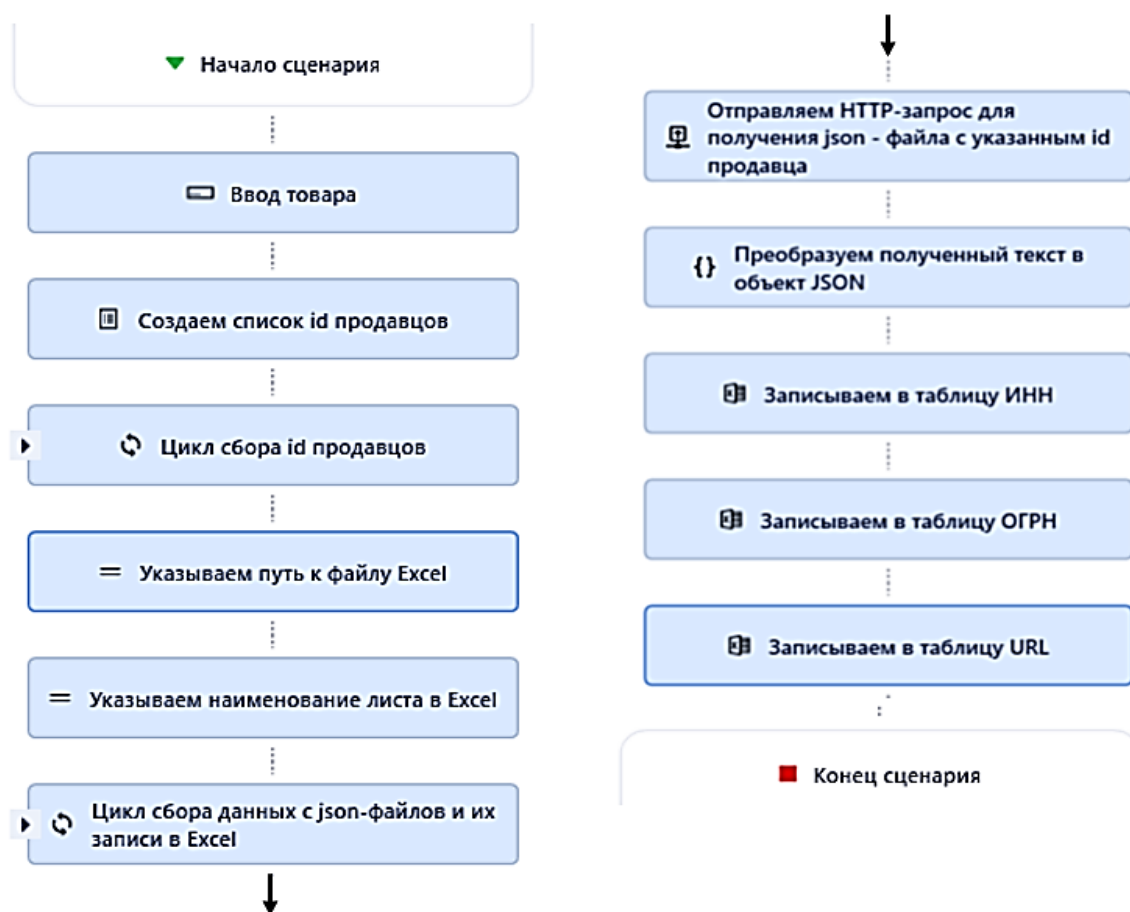


Рис. 1. Обобщенная схема основных шагов алгоритма – интерфейс платформы PIX RPA.

2.3 Сбор данных с маркетплейса Ozon

Веб-ресурс маркетплейса Ozon обладает рядом особенностей: нет URL для подключения к API, подходящего для адаптации для автоматической обработки данных, структура страниц ресурса регулярно меняется, изменяя позиции ее элементов в CSS. Это приводит к необходимости адаптации предыдущего алгоритма сбора данных под особенности платформы.

Обобщенная схема сбора данных на маркетплейсе Ozon:

- выбрать наименование товара;
- найти элементы, содержащие сведения о продавце. Используется язык запросов XPath, позволяющий искать элементы по тексту, а не по структуре в CSS;
- перейти в раздел «Продавец»;

- при наличии кнопки «Посмотреть все», нажать на неё. В результате прогрузится полный список продавцов данного товара;
- перейти на страницу продавца и извлечь необходимые данные из контактной информации.

Особенности реализации робота:

- необходимо внедрять в код робота активности «ожидание» или «ожидание веб-элемента», которые приостановят программу на указанное время, чтобы веб-элементы успели полностью загрузиться;
- страница продавца может потребовать прокрутки роботом для извлечения необходимых данных, которые находятся внизу станицы, с помощью языка запросов XPath. Такая прокрутка может быть реализована с помощью программного кода, созданного на языке JavaScript и встраиваемого в основной код робота;
- если в момент работы робота товар закончился, и его карточка отсутствует на веб-ресурсе, то может возникнуть ошибка. На платформе PIX RPA предусмотрена специальная процедура «Try/Catch/Finally» обработки таких ошибок. Ее следует реализовать для разрабатываемого робота.

Укрупненная схема для сервисного робота, реализованная на платформе PIX RPA, представлена на рис. 2.

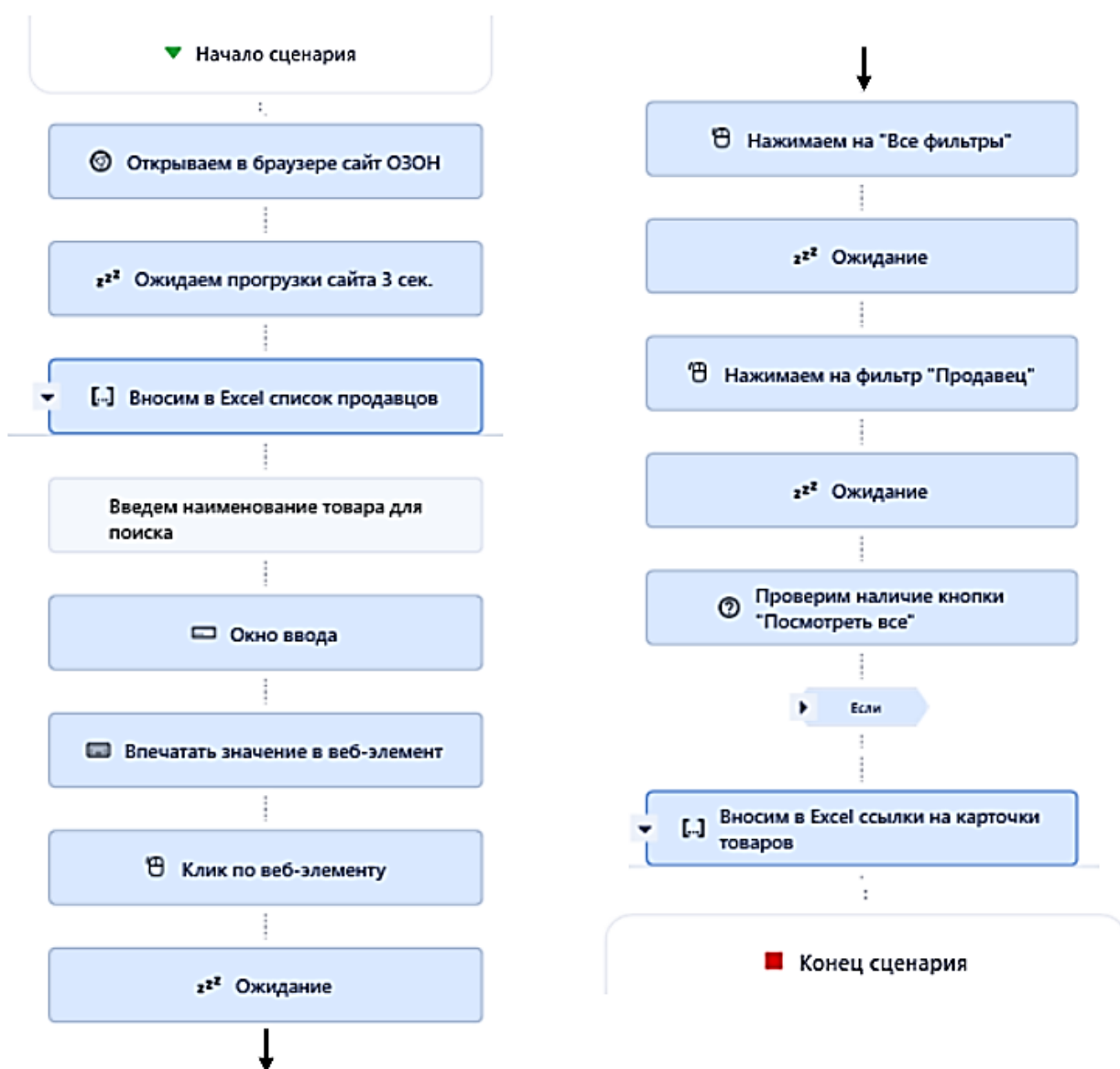


Рис 2. Обобщенная схема основных шагов алгоритма – интерфейс платформы PIX RPA.

Блок «Если» обобщенной схемы, представленной на рис. 2, реализует условия поиска продавцов при наличии и отсутствии на веб-ресурсе кнопки раскрытия сведений о компании-продавце товара.

Оценка временных затрат на сбор данных из интернета должна включать временные затраты на разработку робота, поддержку его работы и время работы скрапера. Усредненная оценка для описанного кейса включает 4 часа и 7 часов работы специалиста для написания программных кодов для маркетплейсов Wildberries и Ozon соответственно, 30 минут еженедельной отладки кода под изменения сайта. Сбор данных для одного прохода в среднем составляет 15 минут. Заметим, что для оценки финансовой целесообразности необходимо учитывать также и стоимость лицензии на одного робота или ежемесячные затраты для соответствующего программного обеспечения.

Описанный кейс позволил подтвердить следующие положения: сокращается время на ручной поиск данных, единый сценарий позволяет его использовать многократно (для разных объектов сбора), учитываются особенности веб-организации площадок-источников.

Реализация скрапинга посредством применения программного робота обладает следующими особенностями:

- во время работы робота браузер должен оставаться открытым;
- некоторые события веб-элемента не могут быть вызваны роботом, в связи с чем необходимо использовать JavaScript код или передать управление мышью роботу, что блокирует любые другие действия пользователем;
- необходимо выставлять задержки между действиями, чтобы веб-страницы успели полностью загрузиться.

Заключение

Для эффективного сбора данных компании прежде всего нужно четко поставить поисковую задачу. На втором этапе необходимо провести тщательную оценку имеющихся источников информации, в том числе их структуру и особенности, а также учесть правовые ограничения при работе с внешними данными.

Сбор данных актуальна и в то же время сложная для бизнеса задача. Хотя технологий реализации такой задачи существует много, российский рынок программного обеспечения в большей степени сосредоточен на развитии технологий RPA, нежели методов веб-скрапинга. Это связано с тем, что технологии RPA имеют более широкое назначение и активно применяются для роботизации различных бизнес-процессов компании. Переход на эти технологии для процесса сбора данных способствует его естественной интеграции в общую архитектуру бизнес-процессов компании.

Благодарности

Исследование выполнено при поддержке Фонда Потанина.

Литература

1. Егармин П.А., Панов Р.Е., Ахматшин Ф.Г., Егармина А.П., Золотухина И.Т. Технология парсинга данных с применением нейросети и алгоритма web-драйвера. Современные наукоемкие технологии. 2024. № 5-1. С. 26-30.
2. Ежова Л.А., Сергеев С.А. Инновации в госзакупках: парсинг для снижения затрат и повышения прозрачности. РИСК: Ресурсы, Информация, Снабжение, Конкуренция. 2024. № 4. С. 187-192.
3. Козина С.А., Кулинченко И.А., Коробкин Д.М., Фоменков С.А. Концепция и архитектура парсинга и хранения единой базы патентов и научных журнальных публикаций. Моделирование, оптимизация и информационные технологии. 2024;12(4).
4. Меньшиков Я.С. Преимущества автоматического сбора данных в сети интернет над ручным сбором данных. Universum: технические науки. 2022. № 10-1 (103). С. 33-36.
5. Напольская Ю.В. Парсинг цен В FMCG: инструмент для контроля, анализа и прогнозирования. Цифровое моделирование экономики. 2025. Т. 2. № 1. С. 170-181.

6. Р 50.1.089-2014. Менеджмент риска. Использование внешних источников данных. Рекомендации по стандартизации. М: Стандартинформ. 2015.
7. Российский рынок RPA-систем 2014. URL: https://www.tadviser.ru/index.php/Статья:Российский_рынок_RPA-систем_2024 (дата обращения 15.08.2025).
8. Barbera, G., Araujo, L., & Fernandes, S. The value of web data scraping: An application to TripAdvisor// Big Data and Cognitive Computing, 2023 7(3), Art. 121. <https://doi.org/10.3390/bdcc7030121>
9. Bavdaž M, Snijkers G, Sakshaug JW, et al. Business data collection methodology: Current state and future outlook. Statistical Journal of the IAOS, 2020. 36(3), P. 741-756. DOI: 10.3233/SJI-200623
10. Boegershausen, J., Datta, H., Borah, A., Stephen, A. T. (2022). Fields of gold: Scraping web data for marketing insights// Journal of Marketing, 2022. 86(5), P. 1-20. <https://doi.org/10.1177/00222429221100750>
11. Chugh, A., Jain, C., Arora, Y. Integrating Robotic Process Automation and Machine Learning for Web Scraping// In Intelligent Systems and Smart Infrastructure. 2023. P. 287-293. CRC Press. DOI 10.1201/9781003357346-32
12. Florentina, M. Web data extraction with robot process automation. study on LinkedIn web scraping using UiPath Studio// Annals of 'Constantin Brancusi' University of Targu-Jiu. Engineering Series (1). 2020. P. 14.
13. Jin, C., Fan, H., Huang, Z., Dou, Y. Wcsr: a robotic process automation data collection method based on web content structure recognition// In 2022 6th International Conference on Robotics and Automation Sciences (ICRAS). 2022. P. 73-80. DOI: 10.1109/ICRAS55217.2022.9842218.
14. Khder, M. A. Web scraping or web crawling: State of art, techniques, approaches and application// International Journal of Advances in Soft Computing & Its Applications. 2021. 13(3), P. 144-168. DOI: 10.15849/ijasca.211128.11
15. Li, Y. M., Li, T. Y. Deriving market intelligence from microblogs // Decision Support Systems, 2013. P. 206-217. <https://doi.org/10.1016/j.dss.2013.01.023>
16. Markovich, A., Efrat, K., Raban, D. R., Souchon, A. L. Competitive intelligence embeddedness: Drivers and performance consequences// European Management Journal, 2019. 37(6), P. 708-718. <https://doi.org/10.1016/j.emj.2019.04.003>
17. Mishra, A., Mishra, S., Kumar, N. S. Data Analysis using Robot Process Automation Study on Web Scraping using UI Path Studio// In 2022 4th International Conference on Advances in Computing, Communication Control and Networking (ICAC3N). 2022. P. 2221-2225. DOI: 10.1109/ICAC3N56670.2022.10074502
18. Munot, A. V., Bora, P. P., Durgude, S. An Intelligent and Automated Web Data Extraction System for E-commerce// In International Conference on Smart Computing and Communication. 2024. (P. 329-337). Singapore: Springer Nature Singapore. https://doi.org/10.1007/978-981-97-1313-4_28
19. Petrov, S., McDonald, R. Overview of the 2012 shared task on parsing the web// In Notes of the first workshop on syntactic analysis of non-canonical language (sancl). 2024. Vol. 59, P. 4-44.
20. Singrodia, V., Mitra, A., & Paul, S. A review on web scrapping and its applications// In 2019 international conference on computer communication and informatics (ICCCI) 2019. P. 1-6 IEEE. DOI: 10.1109/ICCCI.2019.8821809.
21. Snijkers, Ger, et al. Designing and conducting business surveys. John Wiley & Sons, 2013.
22. Tathe K., Sharma S., Kharade J. Data collection using web scrapping with robotic process automation// International Research Journal of Modernization in Engineering Technology and Science. 2022. Volume: 04. Issue: 06.
23. Usman, K., Hayath, T., Shafiulla, M. Robotics Process Automation Technologies for Data Collection and Analysis // 2024 15th International Conference on Computing Communication and Networking Technologies (ICCCNT), Kamand, India, 2024, pp. 1-7, DOI: 10.1109/ICCCNT61001.2024.10725006.

INTERNET DATA COLLECTION PROCESSES: POTENTIAL FOR BUSINESS AND RPA

Gadasina, Lyudmila V.

*Candidate of physical and mathematical sciences
Saint Petersburg State University
Department of Information Systems in Economics
Associate Professor
Saint Petersburg, Russian Federation
l.gadasina@spbu.ru*

Ivanova, Victorija V.

*Candidate of economic sciences
Saint Petersburg State University
Department of Information Systems in Economics
Associate Professor
Saint Petersburg, Russian Federation
v.ivanova@spbu.ru*

Zyrianova, Daria A.

*Saint Petersburg State University
Master's Student in Business Informatics
Saint Petersburg, Russian Federation
darastarovatova@gmail.com*

Abstract

The article is devoted to the issues of automation of data collection on the Internet. The types of sources and the features of data collection from them are systematized. An overview of modern technologies of data collection is presented, with a special focus on the technology of robotic automation of processes Robotic Process Automation (RPA). A practical case is provided, illustrating specific steps for creating automated data collection processes on popular marketplaces Wildberries and Ozon for a logistics company.

Keywords

RPA technology; internet data; scraping

References

1. Egarmin P.A., Panov R.E., Ahmatshin F.G., Egarmina A.P., Zolotuhina I.T. Tekhnologiya parsinga dannyh s primeneniem nejroseti i algoritma web-drajvera. *Sovremennye naukoemkie tekhnologii*. 2024. № 5-1. S. 26-30.
2. Ezhova L.A., Sergeev S.A. Innovacii v goszakupkah: parsing dlya snizheniya zatrat i povysheniya prozrachnosti. *RISK: Resursy, Informaciya, Snabzhenie, Konkurenciya*. 2024. № 4. S. 187-192.
3. Kozina S.A., Kulinchenko I.A., Korobkin D.M., Fomenkov S.A. Konceptiya i arhitektura parsinga i hraneniya edinoj bazy patentov i nauchnyh zhurnal'nyh publikacij. *Modelirovanie, optimizaciya i informacionnye tekhnologii*. 2024; 12(4).
4. Men'shikov Ya.S. Preimushchestva avtomaticheskogo sbora dannyh v seti internet nad ruchnym sborom dannyh. *Universum: tekhnicheskie nauki*. 2022. № 10-1 (103). S. 33-36.
5. Napol'skaya Yu.V. Parsing cen V FMCG: instrument dlya kontrolya, analiza i prognozirovaniya. *Cifrovoe modelirovanie ekonomiki*. 2025. T. 2. № 1. S. 170-181.
6. R 50.1.089-2014. Menedzhment riska. Ispol'zovanie vneshnih istochnikov dannyh. Rekomendacii po standartizacii. M: Standartinform. 2015.

7. Rossijskij rynok RPA-sistem 2014.URL: https://www.tadviser.ru/index.php/Статья:Российский_рынок_RPA-систем_2024 (accessed on 15.08.2025).
8. Barbera, G., Araujo, L., & Fernandes, S. The value of web data scraping: An application to TripAdvisor// Big Data and Cognitive Computing, 2023 7(3), Art. 121. <https://doi.org/10.3390/bdcc7030121>
9. Bavdaž M, Snijkers G, Sakshaug JW, et al. Business data collection methodology: Current state and future outlook. Statistical Journal of the IAOS, 2020. 36(3), P. 741-756. DOI: 10.3233/SJI-200623
10. Boegershausen, J., Datta, H., Borah, A., Stephen, A. T. (2022). Fields of gold: Scraping web data for marketing insights// Journal of Marketing, 2022. 86(5), P. 1-20. <https://doi.org/10.1177/00222429221100750>
11. Chugh, A., Jain, C., Arora, Y. Integrating Robotic Process Automation and Machine Learning for Web Scraping// In Intelligent Systems and Smart Infrastructure. 2023. P. 287-293. CRC Press. DOI 10.1201/9781003357346-32
12. Florentina, M. Web data extraction with robot process automation. study on LinkedIn web scraping using Uipath Studio//Annals of 'Constantin Brancusi' University of Targu-Jiu. Engineering Series (1). 2020. P. 14.
13. Jin, C., Fan, H., Huang, Z., Dou, Y. Wcsr: a robotic process automation data collection method based on web content structure recognition// In 2022 6th International Conference on Robotics and Automation Sciences (ICRAS). 2022. P. 73-80. DOI: 10.1109/ICRAS55217.2022.9842218.
14. Khder, M. A. Web scraping or web crawling: State of art, techniques, approaches and application// International Journal of Advances in Soft Computing & Its Applications. 2021. 13(3), P. 144-168. DOI: 10.15849/ijasca.211128.11
15. Li, Y. M., Li, T. Y. Deriving market intelligence from microblogs // Decision Support Systems, 2013. P. 206-217. <https://doi.org/10.1016/j.dss.2013.01.023>
16. Markovich, A., Efrat, K., Raban, D. R., Souchon, A. L. Competitive intelligence embeddedness: Drivers and performance consequences// European Management Journal, 2019. 37(6), P. 708-718. <https://doi.org/10.1016/j.emj.2019.04.003>
17. Mishra, A., Mishra, S., Kumar, N. S. Data Analysis using Robot Process Automation Study on Web Scraping using UI Path Studio//In 2022 4th International Conference on Advances in Computing, Communication Control and Networking (ICAC3N). 2022. P. 2221-2225. DOI: 10.1109/ICAC3N56670.2022.10074502
18. Munot, A. V., Bora, P. P., Durgude, S. An Intelligent and Automated Web Data Extraction System for E-commerce// In International Conference on Smart Computing and Communication. 2024. (P. 329-337). Singapore: Springer Nature Singapore. https://doi.org/10.1007/978-981-97-1313-4_28
19. Petrov, S., McDonald, R. Overview of the 2012 shared task on parsing the web// In Notes of the first workshop on syntactic analysis of non-canonical language (sancl). 2024. Vol. 59, P. 4-44.
20. Singrodia, V., Mitra, A., & Paul, S. A review on web scrapping and its applications// In 2019 international conference on computer communication and informatics (ICCCI) 2019. P. 1-6 IEEE. DOI: 10.1109/ICCCI.2019.8821809.
21. Snijkers, Ger, et al. Designing and conducting business surveys. John Wiley & Sons, 2013.
22. Tathe K., Sharma S., Kharade J. Data collection using web scrapping with robotic process automation// International Research Journal of Modernization in Engineering Technology and Science. 2022. Volume: 04. Issue: 06.
23. Usman, K., Hayath, T., Shafiulla, M. Robotics Process Automation Technologies for Data Collection and Analysis // 2024 15th International Conference on Computing Communication and Networking Technologies (ICCCNT), Kamand, India, 2024, pp. 1-7, DOI: 10.1109/ICCCNT61001.2024.10725006.