

ВЕБ-СКРЕЙПИНГ И ЭТИЧЕСКИЕ АСПЕКТЫ СБОРА ИНФОРМАЦИИ

Гимпель К.А.

*Белорусский государственный университет информатики и радиоэлектроники,
г. Минск, Республика Беларусь*

Научный руководитель: Нестеренков С.Н. – к.т.н., доцент, доцент кафедры ПОИТ

Аннотация. В статье рассматриваются технологические и правовые основы веб-скрейпинга как метода автоматизированного сбора структурированных данных из открытых интернет-источников. Проводится анализ этических и юридических рисков, возникающих при его применении.

Ключевые слова: веб-скрейпинг, сбор данных, персональные данные.

Введение. В условиях цифровизации общества и стремительного роста объёмов информации, размещаемой в сети Интернет, возрастает потребность в эффективных методах её сбора и анализа. Значительная часть таких данных представлена в неструктурированном виде и предназначена прежде всего для восприятия человеком, что существенно усложняет их последующую автоматизированную обработку.

Основная часть. Одним из подходов к решению задачи является веб-скрейпинг – технология автоматизированного извлечения (сбора) структурированных данных с веб-страниц с помощью программного обеспечения, которое имитирует действия человека в веб-браузере для получения доступа к содержимому и его дальнейшей обработки. Веб-скрейпинг находит широкое применение в анализе рынков, мониторинге цен, академических исследованиях и многих других областях. Его использование позволяет формировать массивы данных, пригодные для последующего статистического анализа и машинного обучения.

Вместе с тем расширение практики использования веб-скрейпинга актуализирует вопросы правомерности и допустимости его применения. Автоматизация процесса сбора информации снижает порог доступа к данным, однако одновременно повышает риск нарушения прав субъектов информации и владельцев интернет-ресурсов. Это требует комплексного рассмотрения не только технических возможностей, но и сопутствующих юридических и этических ограничений.

В Республике Беларусь информация по категории доступа делится на общедоступную и информацию, распространение и (или) предоставление которой ограничено. Так вот скрейпинг информации, доступ к которой ограничен, является незаконным. К такой информации относятся сведения о частной жизни физических лиц и персональные данные, государственные секреты, служебная информация ограниченного распространения, а также коммерческая, банковская и иная охраняемая законом тайна. Соответственно сбор общедоступных данных, таких как открытые цены в интернет-магазинах или публикации в новостных лентах, как правило, признается правомерным. Проблема возникает при автоматическом сборе телефонных номеров, адресов электронной почты, фамилий и имен пользователей с форумов, досок объявлений или профилей в социальных сетях без явного согласия субъектов данных. Такие действия нарушают законодательство о защите персональных данных.

Кроме того, правовые аспекты веб-скрейпинга тесно связаны с соблюдением технических и договорных ограничений, устанавливаемых владельцами веб-ресурсов. В соответствии с положениями статьи 20 Закона «Об информации, информатизации и защите информации», «Порядок распространения и (или) предоставления информации ... определяется соглашением субъектов соответствующих информационных отношений, если иное не установлено законодательными актами» [1]. Почти у каждого сайта присутствует текстовый файл robots.txt, который содержит инструкции для поисковых роботов, о том какие страницы сайта они могут сканировать и индексировать, а какие нет. Таким образом содержание этого файла – это прямое техническое указание владельца ресурса, запрещающее доступ для роботов (включая скрейперов) на конкретные страницы. Его игнорирование может

рассматриваться как нарушение установленного «порядка предоставления информации». Дополнительно многие сервисы включают в пользовательские соглашения явные запреты на автоматизированный сбор данных. Нарушение таких условий способно повлечь гражданско-правовую ответственность, связанную с неисполнением договорных обязательств.

Помимо строгого следования букве закона, ответственное использование веб-скрейпинга предполагает соблюдение ряда этических принципов. К ним относится уважение к ресурсам целевого веб-сайта. Интенсивные автоматизированные запросы могут создавать избыточную нагрузку на сервер, приводя к замедлению работы или даже отказу в обслуживании для обычных пользователей. Поэтому важно реализовывать механизмы, ограничивающие частоту запросов и собирать только те данные, которые действительно необходимы для достижения заявленной цели. С технической стороны это может выражаться в настройке произвольных пауз между запросами, использовании кэширования уже полученных данных и работе в периоды наименьшей нагрузки на сервер.

Этический подход также подразумевает прозрачность целей сбора данных. В случаях, когда скрейпинг проводится в исследовательских или академических целях, целесообразно по возможности уведомлять администраторов ресурсов о характере и масштабах сбора данных. Это не только минимизирует риски конфликта, но и способствует формированию доверия между исследователями и владельцами цифровых платформ. Подобная коммуникация может выявить существующие интерфейсы прикладного программирования (API), использование которых является более предпочтительным и легитимным методом получения данных, поскольку они специально предназначены для программного взаимодействия и часто имеют встроенные ограничения по частоте запросов.

Необходимо также учитывать долгосрочные последствия скрейпинга для целостности информационной экосистемы. Массовый автоматизированный сбор может искажать аналитику, доступную владельцам ресурсов, например, данные о трафике, и влиять на их бизнес-решения. Более того, он может провоцировать ответные меры – усиление технических барьеров, таких как внедрение CAPTCHA, блокировка IP-адресов или усложнение структуры сайта, что в конечном счете усложняет доступ к информации для всех сторон, включая законных пользователей.

Заключение. Веб-скрейпинг представляет собой инструмент для работы с большими объемами информации, применение которого требует взвешенного подхода на стыке технологий, права и этики. Правомерность сбора данных определяется не только их формальной доступностью, но и соблюдением законодательных норм, в частности касающихся защиты персональных данных и коммерческой тайны, а также уважением к техническим ограничениям, установленным владельцами ресурсов.

Список литературы

1. Об информации, информатизации и защите информации [Электронный ресурс] : Закон Респ. Беларусь, 10 нояб. 2008 г., № 455-3 // ЭТАЛОН. Законодательство Республики Беларусь / Нац. центр правовой информ. Респ. Беларусь. – Минск, 2026. – Режим доступа: <https://pravo.by/document/?guid=3871&p0=H10800455>. – Дата доступа: 11.02.2026.

UDC 004.62:34.096

WEB SCRAPING AND ETHICAL ASPECTS OF INFORMATION COLLECTION

Gimpel K.A.

Belarusian State University of Informatics and Radioelectronics, Minsk, Republic of Belarus

Nesterenkov S.N. – Cand. of Sci., associate professor of the department of ITS

Annotation. The article discusses the technological and legal foundations of web scraping as a method of automated collection of structured data from open Internet sources. The ethical and legal risks arising from its application are analyzed.

Keywords: web scraping, data collection, personal data.