

УДК 343.21:004.4

Трибюк В.О., здобувач

Фант М.О., к.філол.н.

Державний університет «Житомирська політехніка»

ЮРИДИЧНІ ТА ЕТИЧНІ АСПЕКТИ ВЕБ-СКРАПІНГУ

На сьогоднішній день інформація – це один з головних ресурсів людства. Отримання та обробка цієї інформації вручну займає дуже багато часу, що для нас є проблемою у світі, який швидко розвивається і де кожна хвилина має значення. Саме тому зростає потреба в автоматизованих методах збору даних, одним із яких є веб-скрапінг. Веб-скрапінг – це автоматизована техніка, що дозволяє програмно витягувати дані із вебсайтів та формувати на їх основі власні інформаційні ресурси, каталоги або аналітичні вибірки [2].

Однак у процесі застосування, можуть виникнути низка проблем та обмежень, такі як етичність та легальне отримання інформації. Серед найбільш поширених труднощів варто виокремити – динамічні інтерфейси, захист від ботів, зміна структури сторінок, а також пов'язані з авторським правом, персональними даними та умовами користування веб ресурсами.

Тому перш ніж розпочинати веб-скрапінг певного контенту, доцільно перевірити, чи надає ресурс офіційні способи доступу до даних – зокрема:

- власний API;
- партнерські програми або комерційні ліцензії на контент.

За відсутності таких механізмів необхідно уважно ознайомитися з «Умовами надання послуг» та файлом robots.txt, який визначає дозволені або заборонені для автоматичного збору розділи сайту, а також може містити параметр Crawl-delay, який визначає мінімальний проміжок часу між запитамі, що надсилаються до сайту.

Особливо важливим є врахування юридичних ризиків, які можуть призвести до штрафів та судових позовів. Основні проблемні зони:

1. Ігнорування умов надання послуг.

Багато веб-ресурсів забороняють автоматизований збір даних. Порушення цих умов вважається недотриманням контракту та злочином, що може призвести до блокування доступу або юридичних претензій з боку власника сайту.

2. Збір приватної або персональної інформації без правових підстав.

Збір персональних даних, таких як електронні адреси, номери телефонів, IP-адреси або інші ідентифікаційні відомості, регулюється спеціальними законодавчими актами. В Україні цей процес регламентується Законом України «Про захист персональних даних» [1]. На міжнародному рівні, застосовується Регламент ЄС про захист даних (GDPR) [4]. Несанкціоноване збирання або обробка персональних даних може кваліфікуватися як порушення правових норм та спричиняти юридичну відповідальність.

3. Створення надмірного навантаження на сервери через занадто часті запити.

Невеликий інтервал між HTTP-запитами може призвести до підвищеної завантаженості сервера та уповільнення роботи веб ресурсу, що розцінюється як неетична або навіть шкідлива поведінка. Таке порушення може призвести до блокування IP або судових позовів.

4. Копіювання захищеного авторським правом контенту без отримання дозволу.

Тексти, зображення, описи товарів, огляди та інші матеріали є інтелектуальною власністю. Масове копіювання контенту без ліцензійного дозволу або без посилань на власника може кваліфікуватися як порушення авторського права, навіть якщо сайт знаходиться у відкритому доступі [3].

Підводячи підсумки після проведеного дослідження можна виокремити, що в більшості випадків веб-скрапінг – це «сіра зона». Адже навіть, якщо інформація знаходиться у відкритому доступі, вона може захищатися різними юридичними законами, що ускладнює належне збирання даних. Тому найбільш доцільним є використання офіційних джерел доступу до даних або укладання домовленостей із власниками веб ресурсів.

Список використаних джерел:

1. Про захист персональних даних. Офіційний вебпортал парламенту України [Електронний ресурс] – Режим доступу до ресурсу: <https://zakon.rada.gov.ua/laws/show/2297-17#Text>

2. Introduction to Web Scraping - GeeksforGeeks. GeeksforGeeks [Електронний ресурс] – Режим доступу до ресурсу: <https://www.geeksforgeeks.org/web-scraping/introduction-to-web-scraping/>

3. Web scraping - legal or illegal? - geeksforgeeks. GeeksforGeeks [Електронний ресурс] – Режим доступу до ресурсу: <https://www.geeksforgeeks.org/python/web-scraping-legal-or-illegal/>

4. What is GDPR, the eu's new data protection law? - gdpr.eu. GDPR.eu [Електронний ресурс] – Режим доступу до ресурсу: <https://gdpr.eu/what-is-gdpr/>