

**ПИТАННЯ КРИМІНАЛЬНОГО ПРАВА,  
КРИМІНОЛОГІЇ ТА КРИМІНАЛЬНО-ВИКОНАВЧОГО ПРАВА**

**КОРОСТЕЛЬОВА Л. А.,**  
ад'юнкта кафедри кримінально-  
правових дисциплін  
(Луганський державний університет  
внутрішніх справ імені Е. О. Дідоренка)

**УДК 343.9**

**DOI <https://doi.org/10.32842/2078-3736/2022.2.29>**

**ІННОВАЦІЇ У ЗБОРІ ДАНИХ ДЛЯ КРИМІНОЛОГІЧНИХ ДОСЛІДЖЕНЬ**

У статті розглянуто сучасні можливості для розвитку кримінологічних досліджень. Доведено, що однією із інновацій у зборі даних для кримінологічних досліджень є метод вебскрейпінгу. Розкрито суть поняття вебскрейпінгу, а також розглянуто основні інструменти цього методу. Приділено увагу розгляду ролі вебскрейпінгу у кримінології, а також зокрема і в кримінологічних дослідженнях як інноваційного методу збору інформації.

Висловлено авторську позицію про те, що одним із напрямків оновлення методів збору інформації для кримінологічних досліджень є вебскрейпінг. Аргументовано, що автоматизація збору даних інформації (вебскрейпінг), створила нові можливості для кримінологічних досліджень. Наведено приклади використання зазначеного методу у зарубіжній кримінології, а також на прикладі закордонних досліджень було проведене власне дослідження щодо протидії маніпуляцій і дезінформації у Луганській області. У роботі визначено комплекс переваг та недоліків зазначеного методу збору даних для кримінологічних досліджень, які згруповані навколо методологічних, технічних, правових, фінансових факторів.

Обґрунтовано висновок про те, що у ХХІ столітті збільшуються комунікаційні потоки і глобалізація поширення Інтернету та мобільного зв'язку суттєво вплинуло на соціальні зміни. В умовах цифрової сучасності, кримінологи зобов'язані розуміти важливість цифрової культури та набуття навичок у галузі інформаційно-комунікаційних технологій, які необхідні для запобігання та протидії злочинності. Використання зазначеного методу у зборі даних у вітчизняну кримінологію як методу підвищує ефективність пошуку інформації у дослідницькій діяльності. У статті доведено, що розвиток відкритих даних у світі, зокрема і в Україні оновив можливості вебскрейпінгу, що у подальшому буде дозволяти передбачати і запобігати майбутнім тенденціям злочинності.

**Ключові слова:** інноваційні методи, збір даних, кримінологія, кримінологічні дослідження, вебскрейпінг.

**Korostelova L. A. Innovations in data collection for criminological research**

The article considers modern opportunities for the development of criminological research. It is proved that one of the innovations in data collection for criminological research is the method of web-scraping. The essence of the concept of web scraping is revealed, and also the basic tools of this method are considered. Attention is paid to the role of webscraping in criminology, as well as in criminological research as an innovative method of collecting information.



The author's position is expressed that one of the directions of updating the methods of collecting information for criminological research is web-scraping. It is argued that the automation of data collection (web-scraping) has created new opportunities for criminological research. Examples of the use of this method in foreign criminology are given, as well as on the example of foreign studies conducted its own study on combating manipulation and misinformation in the Luhansk region. The paper identifies a set of advantages and disadvantages of this method of data collection for criminological research, which are grouped around methodological, technical, legal, financial factors. The conclusion that communication flows are increasing in the 21st century and the globalization of the Internet and mobile communications has significantly affected social change is substantiated. In today's digital world, criminologists must understand the importance of digital culture and the acquisition of information and communication technology skills needed to prevent and combat crime. The use of this method in data collection in domestic criminology as a method increases the efficiency of information retrieval in research activities. The article proves that the development of open data in the world, in particular in Ukraine, has updated the possibilities of web-scraping, which in the future will allow to predict and prevent future trends in crime.

**Key words:** *innovative methods, data collection, criminology, criminological research, web scraping.*

**Вступ.** Актуальним напрямом забезпечення процесу протидії злочинності є усунення безсистемності в зборі інформації [1, с. 206] у кримінологічних дослідженнях.

Збір даних у кримінологічних досліджень це трудомісткий процес, адже на практиці кримінологи використовують різні методи від методу опитування до методи експерименту.

У сучасних умовах є нагальним змінити традиційний підхід до збору інформації про злочинність за яким визнають ряд недоліків та неточностей, умови для відомчого впливу, викривлення даних, що впливає на об'єктивність інформації про стан злочинності в державі та про результати роботи правоохоронної системи [1, с. 206-207].

Невід'ємною складовою сучасної кримінології має стати виявлення, розпізнання, аналіз реальних та потенційних загроз з метою створення пропозицій щодо швидкого реагування на них держави, органів охорони правопорядку, міжнародного співтовариства [2, с. 169].

**Постановка завдання.** Метою статті є розгляд інноваційного методу збору даних для кримінологічних досліджень (вебскрейпінгу) та його можливості для сучасної кримінології.

**Результати дослідження.** Більшість учених кримінологів наполягають на впровадженні інноваційних методів збору кримінологічної інформації.

Так, Жан-Луї ван Гельдер і Стін Ван Даеле, у своєму науковому дослідженні «Інноваційні методи збору даних у кримінологічних дослідженнях», говорять, про те, що за допомогою нових технологій, розроблені нові методи аналізу та збору даних, і більша частина з них доступна на рівні користувачів, що дозволяє збирати та досліджувати дані способами, що виходять за межі стандартних досліджень [3, с. 2].

Слід погодитися із зазначеними ученими, що сучасна трансформація кримінологічної науки обумовлює розвиток інноваційних методів збору даних для кримінологічних досліджень. На нашу думку, перспективним напрямом для сучасної української кримінології і кримінологічних досліджень зокрема, є використання такого методу збору даних – вебскрейпінгу.

Загалом, вебскрейпінг – це автоматичне вилучення даних із вебресурсів. Якщо перекладати з англійського *scraping* – «зішкрібання, зачищення», в сучасних роботах вебскрейпінг не перекладають в більшості випадках, оскільки переклад українською звучить не дуже благозвучно. В Мережі поширені різні засоби і інструменти для вебскрейпінгу. Серед них є бібліотеки, додатки, online-сервіси, хмарні сервіси, сервіси типу DaaS, плагіни до браузерів.



зерів. Скрейпери в цих інструментах часто поєднуються з краулерами (пошуковими системами), що дозволяє витягувати інформацію не тільки з конкретної вебсторінки, але і зі сторінок, на які виводять гіперпосилання з цієї сторінки. У цих випадках процес отримання інформації може займати досить багато часу. Деякі інструменти надають можливість запускати скрейпери за розкладом з певною періодичністю. Оскільки сучасні вебсторінки часто бувають динамічними, частина інструментів для вебскрейпінга надають можливість отримувати інформацію з *html*-коду, який виходить після відпрацювання скриптів [4, с. 12].

Варто зазначити, що сам метод скрейпінгу має кілька етапів для витягнення потрібної інформації із сайтів:

1. Збір контенту. Пошукова машина скрейпінгу завантажує код сторінки сайту, під час такого процесу до нього долучається спеціальний скрипт, який розбиває весь код на частини та аналізує інформацію яка потрібна досліднику.

2. Витяг інформації. Вся інформація, яка витягується зі сторінки, може бути не потрібна користувачеві, адже його цікавить тільки конкретне, наприклад, корисні дані які можна використати їх у своєму науковому дослідженні. Скрейпінг буде знаходити ті сторінки чи розділи, де розміщена інформація про дану категорію, витягнувши все в кінцевий файл тільки тексти цих даних.

3. Збереження результатів. Коли вся потрібна інформація витягнута з сайту, її потрібно зберегти. Зазвичай вносяться записи в базу даних, адже так зручніше аналізувати інформацію з інших даних. Запропонована методика дасть можливість дослідити та проаналізувати дану інформацію, чи потрібна вона дослідникові для внесення її у свою роботу [5, с. 154].

Серед інструментів для вебскрейпінга є і ті, які не вимагають програмування, є інструменти з відкритим кодом, вільно-поширювані, умовно-безкоштовні і повністю комерційні [4, с. 12].

До найбільш популярних інструментів, які не потребують програмування для скрейпінгу можна віднести такі програмні продукти:

1) *Scrapingbee* – це веб-API для збору даних, який обробляє браузер та управління проксі. Він може виконувати *Javascript* на сторінках і обробляти проксі-сервери для кожного запиту, щоб отримати необроблену *HTML*-сторінку, не заблокувавши її. Вони також мають спеціальний *API* для пошуку *Google*. Особливості даної програми полягають по-перше, у підтримки візуалізації *JavaScript*; по-друге, вона забезпечує автоматичну обробку проксі; по-третє, можна безпосередньо використовувати цю програму на *Google Sheet*, а також зазначений додаток можна використовувати з веббраузером *chrome* і він відмінно підходить для збору *Amazon* і пошуку *Google* [6].

2) *Octoparse* – це вебінструмент для збору даних, простий у використанні як для кодерів, так і для некодерів, і популярний для вилучення даних електронної комерції. Він може масштабувати веб-дані (до мільйонів) і зберігати їх у структурованих файлах, таких як *Excel*, *CSV*, *JSON* для завантаження. Основними можливостями, якими володіє програма є: витяг хмари з обертаннями IP для обходу *captcha* та блокування; вбудований інструмент *Regex* для автоматичного очищення даних; заплануйте вишкрібання та регулярно отримуйте оновлення даних; підключення API для налаштування конвеєру даних безпосередньо до вашої бази даних, а також підтримка систем *Windows* і *Mac*. [6]

3) Інструмент *Scraper API* допомагає керувати проксі-серверами, браузерами та *CAPTCHA*. Це дозволяє отримати *HTML* з будь-якої веб-сторінки за допомогою простого виклику API. Його легко інтегрувати, оскільки вам просто потрібно надіслати запит *GET* кінцевій точці API із вашим ключем API та URL-адресою. За допомогою нього можна налаштувати заголовки кожного запиту, а також тип запиту. Інструмент також пропонує геолокацію обертових проксі і неперевершену швидкість та надійність, що дозволяє створювати масштабовані вебскрепери [6].

4) *Scraping-Bot.io* – це ефективний інструмент для сканування даних з URL-адреси. Він надає API, адаптовані до ваших потреб у скануванні: загальний API для отримання необробленого *HTML* сторінки, API, що спеціалізується на розтиранні веб-сайтів, та API для



скасування з веб-сайтів даних. Основними можливостями зазначеного інструменту є: візуалізація *JS (Chrome)*, високоякісні проксі, повна сторінка *HTML*, дозволяє виконувати великі об'ємні збору даних [6].

5) *FMiner* – ще один популярний інструмент для вебскрепінгу, який полягає у вилученні даних, скануванні скрепінгу екрану, макросів та веб-підтримки для *Windows* і *Mac OS*. Особливості зазначеного програмного продукту полягають у тому, що програма дозволяє розробити проект вилучення даних за допомогою простого у використанні візуального редактора, а також надає можливість детально проаналізувати сторінки веб-сайту за допомогою комбінації структур посилань відповідності шаблону *URL*-адреси. Також варто зазначити, що зазначена програма може витягувати дані з важких для сканування динамічних веб-сайтів *Web 2.0* та зокрема, націлює захист вебсайту *CAPTCHA* за допомогою сторонніх автоматизованих служб декапча або введення вручну [6].

Метод скрейпінгу має кілька етапів для витягнення потрібної інформації із сайтів:

1. Збір контенту. Пошукова машина скрейпінгу завантажує код сторінки сайту, під час такого процесу до нього долучається спеціальний скрипт, який розбиває весь код на частини та аналізує інформацію яка потрібна досліднику.

2. Витяг інформації. Вся інформація, яка витягується зі сторінки, може бути не потрібна користувачеві, адже його цікавить тільки конкретне, наприклад, корисні дані які можна використати їх у своєму науковому дослідженні. Скрейпінг буде знаходити ті сторінки чи розділи, де розміщена інформація про дану категорію, витягнувши все в кінцевий файл тільки тексти цих даних.

3. Збереження результатів. Коли вся потрібна інформація витягнута з сайту, її потрібно зберегти. Зазвичай вносяться записи в базу даних, адже так зручніше аналізувати інформацію з інших даних. Запропонована методика дасть можливість дослідити та проаналізувати дану інформацію, чи потрібна вона дослідникові для внесення її у свою роботу [5, с. 154].

Такі сучасні технічні можливості істотно прискорюють процес роботи і в подальшому надають можливість дослідити та проаналізувати дані.

У кримінології метод вебскрейпінгу призначений для збору даних із соціальних мереж і мережевих вебсайтів, таких як *Twitter*, *Facebook*, *LinkedIn* і інші, ставши популярними інструментами для завантаження твітів, статусів та повідомлень. Хоча ці інструменти націлені на вебсайти соціальних мереж, аналіз веб-сторінок націлений на базовий текстовий *HTML*-код, що становить як ці, так і інші веб-сайти. Функції *R* можуть розпізнавати шаблони в коді *HTML* (наприклад, початкові та кінцеві теги *HTML*), витягувати необроблений текст зі сторінок *HTML* і перетворювати цей текст на набір даних, який використовують у статистичному аналізі [7].

Американські кримінологи активно використовують вебскрейпінг у впливі даних про погоду з *Weather Underground* та об'єднанні їх із даними про арешт. Дослідження показали, що насильницькі злочини, включаючи розбійні напади, домашнє насильство та меншою мірою, вбивства, відбуваються частіше з підвищенням температури в атмосферному повітрі. Знання, отримані внаслідок таких проєктів, допомагають правоохоронним органам краще планувати свою діяльність [7].

Ще одним прикладом використання цього методу – журналістський проєкт, створений *Reveal*, який проаналізував екстремістські групи у *Facebook* та порівняв їх списки учасників зі списками членів правоохоронних груп у *Facebook* де виявив багато збігів [8].

На прикладах різних наукових розвідок, нами було проведене власне дослідження із використанням вебскрейпінгу як методу збору інформації щодо протидії маніпуляцій і дезінформації у Луганській області. Було проаналізовано 1500 коментарів і лайків до 20 відео-роліків із *YouTube* каналів новин Луганської області. За результатами нашого дослідження з використанням вище зазначеного методу було отримано наступні результати дослідження, 60% населення Луганської області підтримує і дивиться новини РФ, 40% населення підтримує українські новини і українську владу. Використовуючи такий інноваційний метод збору інформації у сучасних дослідженнях, наприкладі нашого дослідження щодо протидії



маніпуляціям і дезінформації, він дозволяє отримати якісні дані для подальшого розвідувального аналізу. Поєднання сучасних методів із використанням технологій і традиційних методів дослідження маніпуляцій і дезінформації стане основним підґрунтям для подальшої боротьби із такими явищами [9, с. 30-31].

Варто зазначити, що вебскрейпінг має комплекс переваг та недоліків, які згруповані навколо методологічних, технічних, правових, фінансових факторів [10, с. 165].

Якщо розглядати переваги цього методу у кримінологічних дослідженнях: це, по-перше, повнота даних з використанням даного методу збору інформації. Можна наголосити на генеральній сукупності об'єкта дослідження, за умови, що інтернет дані повні для роботи з досліджуваною проблемою. Відкриті дані як не основне джерело будуть доповнювати опис об'єкта.

По-друге, скрейпінг на відмінно від класичних методів збору даних для кримінологічних досліджень, не пов'язаний з так званим «людським фактором», наприклад, неправильною інтерпретацією питання в анкеті опитаного. Виключивши при цьому коректність відбору респондентів та поставлених питань для них. Також у кримінолога з'являється шанс зафіксувати реальну поведінку індивіда, мінімізуючи призму сприйняття. А також за його допомогою можна визначити кримінологічну характеристику.

По-третє, кримінологи виходять не з теорій, а із доступних даних про вебплатформи, які є маркерами найзначніших соціальних подій. Розширюються можливості використання інтернет-даних, якщо не як самостійний метод дослідження, то як відправна точка поглиблення методології збору даних.

Використовуючи дані з інтернет-платформ, вебскрейпінг сприяє об'єктивному розумінню сучасної картини світу, оскільки все більше і більше сучасна людина проводить час у віртуальному світі. А також може стати корисним тоді, коли офіційна статистика не дозволяє описати злочини або взагалі відсутня, це в певному сенсі стосується і латентних злочинів.

Одночасно з можливостями використання даного методу для збору інформації в кримінологічному дослідженні виникає ряд обмежень.

Методологічні обмеження вебскрейпінгу полягають у об'єкті дослідження. Обмежена галузь застосування: підходить для дослідження об'єктів / явищ, які можуть бути описані виключно інтернет-даними (блогосфера, взаємодія через соцмережі, цифрові біржі), або дослідження буде передбачати методику співвіднесення онлайн- та офлайн-статусів індивідів. Технічні обмеження полягають у тому, що платформа може встановити обмеження на пагінацію, обсяг даних, швидкість завантаження під час вебскрейпінгу. Також платформа може внести зміни до структури даних, що призведе до необхідності перетворення коду або конфігурації програмного забезпечення, що використовується для скрейпінгу. [10, с. 167].

**Висновки.** Підсумовуючи вище зазначене, можна стверджувати про таке, цифрова революція вплинула на соціальні зміни. Впровадження вебскрейпінгу у вітчизняну кримінологію як одного із методів збору інформації для кримінологічного дослідження дозволяє оновити методику кримінологічних досліджень і в майбутньому ефективно протидіяти кримінально протиправним діям. Використання зазначеного методу у кримінології потребує підтримки наукової спільноти та перепідготовки кримінологів для вдосконалення знань із програмування, а також вміння роботи із неструктурованою інформацією.

#### Список використаних джерел:

1. Бесчастний В.М. Методи інформаційного забезпечення як інструментарій збору та обробки інформації про злочинність. *Право і суспільство*. Дніпро, 2017. С. 206-212.
2. Головкін Б.М. Теперішнє і майбутнє кримінології. *Проблеми законності*. 2020. Вип. 149. С. 168-184.
3. Jean-Louis van Gelder & Stijn Van Daele. Innovative data collection methods in criminological research: editorial introduction. *Crime Science*. 2014. No 6. P.1 – 4.



4. Китаев Е. Л., Скорнякова Р. Ю. Скрейпинг «на лету» внешних веб ресурсов, управляемый разметкой HTML страницы. *Препринты ИПМ им. М.В. Келдыша*. 2019. № 20. С. 2-31. DOI: <http://doi.org/10.20948/prepr-2019-20>
5. Вакуленко Ю.О., Щербіна О.С. Застосування методу парсингу для ефективного пошуку інформації у дослідницькій діяльності. *Вісник студентського наукового товариства ДонНУ імені Василя Стуса*. 2019. Том 2 № 11. С.153-156.
6. 15 найкращих інструментів для вилучення веб-матеріалів для вилучення даних у 2021 році. URL : <https://uk.csstricks.net/8225197-15-best-web-scraping-tools-for-data-extraction-in-2021> (дата звернення 12.11.2021).
7. Creative Solutions To Elusive Data: Web-Scraping Online Police Reports To Map Co-Offending Networks In US Cities. URL: <https://www.behr.ufl.edu/networks/website-article/creative-solutions-elusive-data-web-scraping-online-police-reports-map-co/> (дата звернення 12.10.2021).
8. Why Web Scraping Is Vital to Democracy, URL : <https://themarkup.org/news/2020/12/03/why-web-scraping-is-vital-to-democracy> (дата звернення 21.10.2021).
9. Коростельова Л.А. Протидія маніпуляціям і дезінформації із використанням інноваційних методів дослідження. *Об'єднані наукою: перспективи міждисциплінарних досліджень*: тези доп. наук.-практ.конф. (м. Київ, 18-19 лист.2021р.). Київ, 2021. С. 30-31.
10. Вилкова О.В. К вопросу о научной осмысленности применения Веб-скрейпинга как метода сбора данных в социологических исследованиях. *Вестн. Томск. гос. ун-та. Философия. Социология. Политология*. 2020. № 54. С. 163–175.

