

2. Gatignon, Hubert, Statistical Analysis of Management Data, Third Edition, New York, NY: Springer Science+Business Media, LLC
3. Evans J.R. Business Analytics: Methods, Models, and Decisions/ 2nd. Global Ed. Pearson, 2017. 653 p
4. David R. Anderson, Dennis J. Sweeney, Thomas A. Williams, Jeffrey D. Camm, James J. Cochran Statistics for Business & Economics/ Cengage Learning, 2016 . 1120 p
5. Fawcett F. T. Data Science for Business: What You Need to Know about Data Mining and Data-Analytic Thinking/ O'Reilly Media, Inc., 2013. 414 p

ЗАСТОСУВАННЯ WEB SCRAPING ДЛЯ ФОРМУВАННЯ ІНДЕКСУ СПОЖИВЧИХ ЦІН

Горобець Олена Олександрівна,

кандидат економічних наук, доцент
кафедри статистики, ІТ
та математичних методів в економіці;

Червона Світлана Петрівна,

кандидат економічних наук, доцент,
доцент кафедри статистики,
ІТ та математичних методів в економіці;
Національна академія статистики, обліку та аудиту

На сьогодні офіційна статистика України практично позбавлена усіх основних джерел статистичних даних, а саме: первинних даних, отримуваних за допомогою статистичних спостережень та обстежень (від суб'єктів комерційної діяльності, фізичних осіб, домашніх господарств, а також державних структур, суспільних об'єктів, тощо), вторинних даних з адміністративних джерел, а також інформації, що збиралася спеціалізованими урядовими агенціями. Таким чином можна стверджувати, що в Україні зараз практично повністю не функціонує система збирання первинних статистичних даних [1].

Важливим макроекономічним показником (у аналізі та прогнозуванні цінових процесів, розробці державних соціальних програм, індексації грошових доходів населення та ін.) є Індекс споживчих цін.

До споживчого набору входять різноманітні товари та послуги за основними розділами, наприклад, 120 позицій за розділом «Продукти харчування та безалкогольні напої», 57 позицій за розділом «Одяг та взуття», 18 позицій за розділом «Охорона здоров'я», 23 позиції за розділом «Транспорт» та ін. [2].

Очевидно, що в умовах війни якісне збирання даних з метою формування Індексу є неможливим.

Альтернативним рішенням є Web Scraping.

Web Scraping – це вилучення даних із веб-сайтів за допомогою автоматизованих програмних засобів [3].

Загалом, Web Scraping можна організувати за допомогою шести основних кроків (рис.).

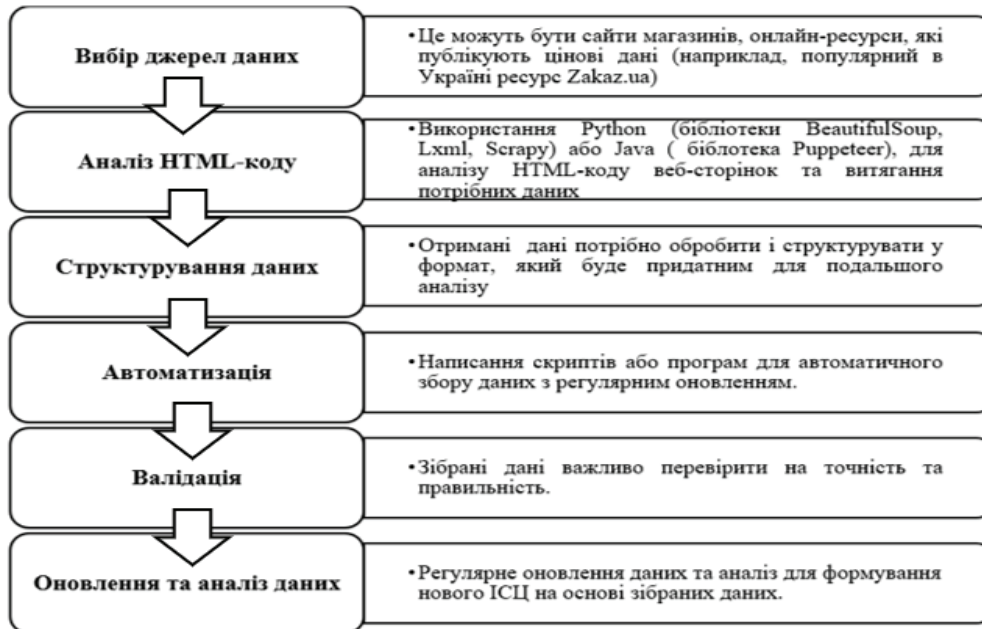


Рис. 1. Шість основних кроків в організації Web Scraping

Джерело: сформовано авторами.

Найпопулярнішою мовою програмування для Web Scraping є Python однак іноді можуть виникати ситуації, коли інші мови, такі як JavaScript (за допомогою Puppeteer) або Ruby (бібліотека Nokogiri), можуть бути більш підходящими залежно від конкретного проєкту чи потреби взаємодії з певними веб-сайтами або технологіями.

Наприклад, JavaScript надає можливість автоматизації взаємодії із веб-сторінками; Ruby дозволяє зручно парсити HTML та XML-документи, а також видобувати необхідні дані; мову програмування R варто використовувати у тих випадках, коли виникає необхідність у статистичному аналізі та обробці даних, також R має пакети, які дозволяють використовувати Web Scraping та отримувати дані із сайтів.

Варто зауважити, на певних правилах Web Scraping. Так, Web Scraping не повинен заборонятися користувачкою угодою сайту, персональні дані користувачів повинні збиратися виключно за їх згодою, не порушується законодавство країни у якій хоститься сайт, що є об'єктом Web Scraping. Деякі веб-сайти забороняють або обмежують Web Scraping, однак ця інформація зазначається у політиці конфіденційності сайту, умовах використання або файлі robots.txt, який забороняє індексацію сторінок пошуковими системами.

Таким чином, Web Scraping є своєрідною альтернативою збирання необхідних даних для формування Індексу споживчих цін у цифровізованому світі.

Список використаних джерел

1. Osaulenko O., Horobets O. (2023). Using Big Data by Ukrainian official statistics when martial law applies: problems and solutions. *Statistics in Transitions. New series*. vol. 24, 1, pp. 29-43.
2. Індекс споживчих цін. URL: https://ukrstat.gov.ua/operativ/prostat/pererahunki/ind_cin/isc.pdf
3. What is Web Scraping and How Can It Benefit Your Business? (2023). Retrieved from: <https://dataforest.ai/blog/what-is-web-scraping-and-how-can-it-benefit-your-business>

КРАУДСОРСИНГ ДАНИХ – СУТНІСТЬ ТА ПЕРЕВАГИ ЗАСТОСУВАННЯ

Гринчак Наталія Анатоліївна,
кандидат економічних наук, доцент,
доцент кафедри менеджменту,
маркетингу та публічного управління;
Національна академія статистики, обліку та аудиту

Краудсорсинг стає все більш популярним методом збору даних. Використовуючи можливість залучення великої групи людей до вирішення певної проблеми, компанії та організації можуть швидко й дешево збирати великі обсяги даних. Як правило, краудсорсинг реалізується через онлайн-спільноти, що дозволяє збирати інформації від великої спільноти користувачів, незалежно від їх місця розташування.

Краудсорсинг даних — це процес отримання даних із великої кількості джерел. Прикладом краудсорсингу даних є використання онлайн-анкет для збору відгуків від клієнтів. Краудсорсинг даних можна використовувати для покращення обслуговування клієнтів, розуміння потреб клієнтів і прийняття кращих бізнес-рішень.

Стрімке зростання популярності краудсорсингу пояснюється його численними перевагами [1]:

– Краудсорсинг даних може надати підприємствам точні та своєчасні дані. Дані є гнучкими та можуть бути змінені відповідно до потреб бізнесу. Підприємство може платити за використання даних або отримувати сповіщення в реальному часі, коли трафік перевантажений.

– Краудсорсинг даних може допомогти пришвидшити процес пошуку потрібних даних, дозволяючи великій кількості людей швидко та дешево надавати дані. Це гарантує швидке виконання завдань з обробки даних із дотриманням високих стандартів якості.

– Краудсорсинг дає компаніям доступ до великої кількості кваліфікованих збирачів даних з усього світу. Високоякісний набір даних важливий для успіху