

ОГЛЯД МЕТОДІВ ТА АЛГОРИТМІВ ЗБОРУ ДАНИХ ДЛЯ ПРОЦЕСУ ПЛАНУВАННЯ ВИКОНАННЯ ДОГОВОРІВ

Кожушко І. В., Грибков С. В.

Національний університет харчових технологій, Київ, Україна

E-mail: illyakozshushko@gmail.com, sergio_nuft@i.ua

Overview of Data Collection Methods and Algorithms for the Contract Simulation Process

Methods and algorithms of data collection for the process of modeling the contract execution plan are described. The field of research is analyzed. The concept of mining data for data collection and analysis is covered. The use of a heuristic approach during web scraping is described. The use of the server-oriented programming language Node.js is covered. Describes ways to combat web scraping, as well as methods to circumvent them. It is proposed to create a forecast for the possible development of the contract life cycle.

Планування виконання договорів із клієнтами на підприємстві можна розглядати як складну систему з ієрархічно зв'язаними між собою елементами. Підходи до планування виконання договорів на різних підприємствах розглянуто в роботі [1]. В сучасних умовах проблема прийняття рішень в області планування виконання договору пов'язана не тільки з дослідженням і точним знанням усієї доступної інформації перед плануванням виконання договору, моделі договору, її властивостей, але й із вибором методів і засобів реалізації збору даних для успішного моделювання такого складного процесу. Затверджувати план на виконання договору та створення моделі виконання договору можна тільки в тому випадку, коли є надійна інформація, яка точно відображає властивості моделі договору, поставлені цілі та задачі плану, розроблено програму дій виконання договору. Тому разом зі стандартним життєвим циклом договору доцільно розглянути можливість створення прогнозу на можливий розвиток життєвого циклу договору, аби відкрити вікно приблизного розуміння майбутнього виконання договору для особи, яка приймає рішення. Таким чином можна забезпечити впевнену підтримку планування виконання договорів.

Проведено огляд методів збору даних в мережі Інтернет, визначено особливості алгоритмів і методів, які забезпечують мінімізацію часу на збір даних, для отримання найбільш точного відображення моделі під час планування виконання договорів.

Для збору та аналізу великої кількості даних з'явилося поняття «дата майнінг» (від англ. data mining — збір даних), яке також відоме як процес KDD (KDD — knowledge discovery in databases) [2–3]. Для завершення процесу моделювання виконання договору використовуються різні методики. Інтелектуальний аналіз даних використовує вже створені інструменти, щоби виявити корисні приховані закономірності, тенденції. За допомогою цих

методів можна отримати й прогнози майбутнього.

Більшість працюючих методів і алгоритмів збору даних у мережі Інтернет базуються на унікальних і евристичних алгоритмах із особистим підходом до кожного джерела даних.

Найрозповсюдженішим методом збору даних є використання серверно-орієнтованих мов програмування, які найкраще всього справляються з багатопоточною роботою системи — наприклад, Node.js. Метод полягає в тому, що веб-скрапер [4] виконує запит на сервер (будь-який сайт завжди розміщено на сервері), отримує відповідь, зберігає дані або формує з них інформацію та надає користувачу. На практиці в більшості випадків достатньо використання POST- і GET-запитів на сервер (рис. 1).

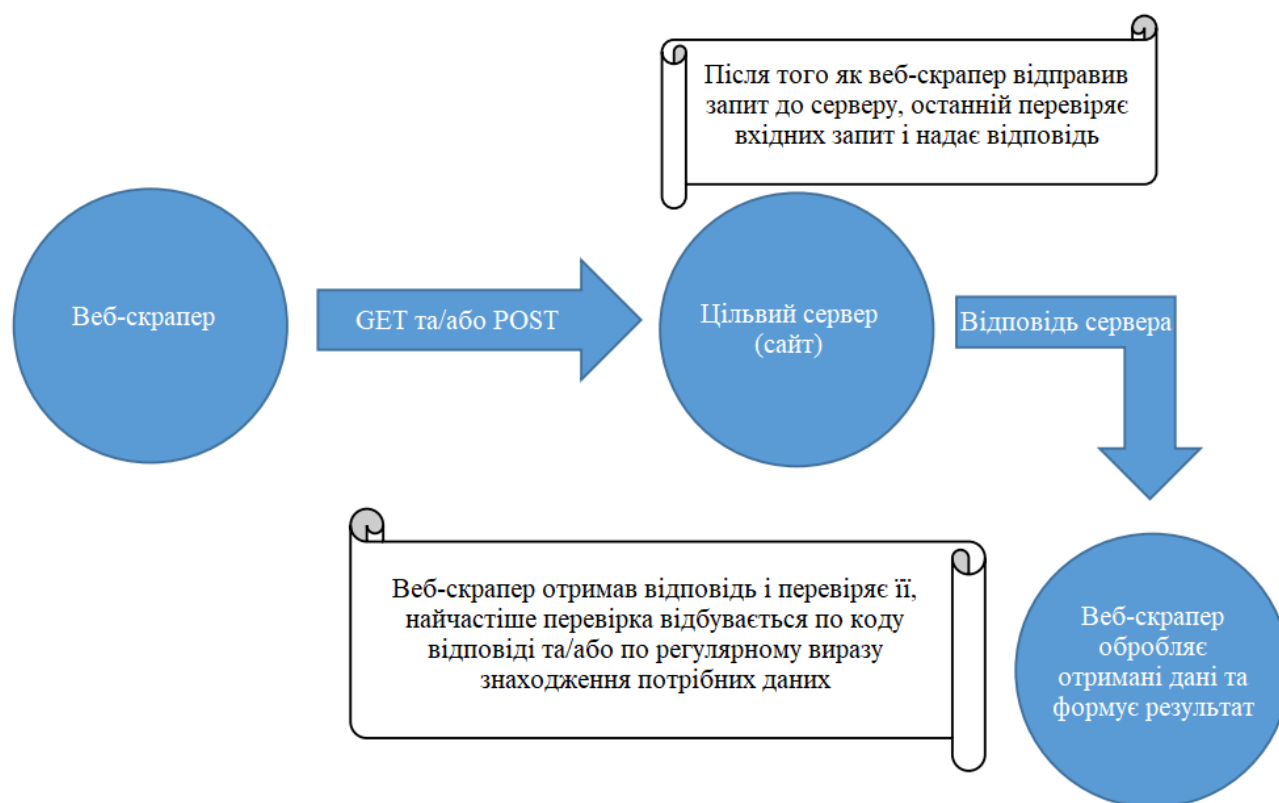


Рис. 1. Метод збору даних веб-скрапером

Веб-скрапери постійно розвиваються. Своєю чергою, сайти, які найчастіше стають ціллю роботи веб-скраперів, також розвивають свій захист і під'єднують різні інструменти для протидії веб-скрапінгу. Якщо говорити про особливі випадки, які почали масово з'являтися зовсім нещодавно, можна виділити інструменти Cloudflare [5] або перевірку за допомогою JavaScript.

Cloudflare надає сайту мережевій послуги захисту контенту, пом'якшення DDoS-атак, роботи служб безпеки в Інтернеті. Головна функція Cloudflare — це роль зворотного проксі для веб-трафіку. Cloudflare підтримує нові веб-протоколи, включно з SPDY і HTTP/2. Також Cloudflare пропонує підтримку HTTP/2 Server Push та проксі Websockets.

Існує метод веб-скрапінгу, за допомогою якого можливо отримувати дані з мережевих джерел, які використовують вищезгадані інструменти для протидії

веб-скрапінгу — емуляція браузера Chrome. В Node.js для цього можна використати модуль Puppeteer [6] (рис. 2). Puppeteer — це бібліотека Node, яка надає високорівневий API для керування Chrome або Chromium за протоколом DevTools.

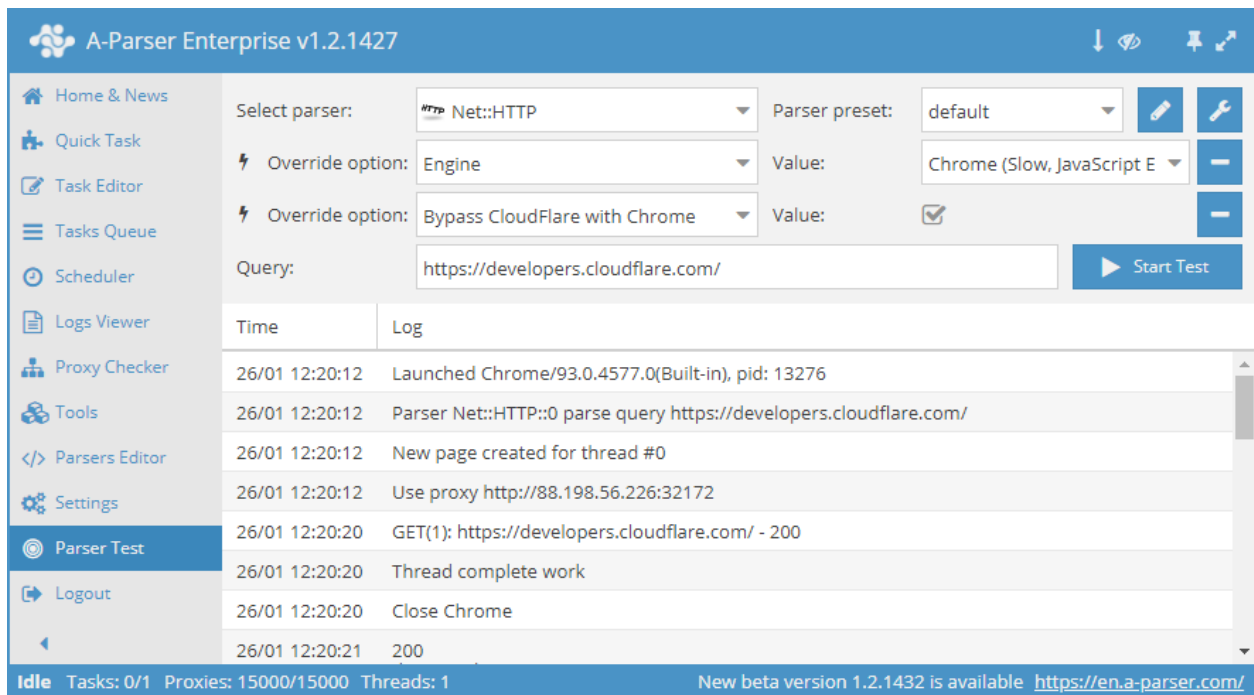


Рис. 2. Приклад використання Puppeteer інструментом A-Parser

Одержані результати буде використано для збору даних, що будуть використовуватись як початкові для процесу планування виконання договорів.

Література

1. Hrybkov S., Kharkianen O., Ovcharuk V., Ovcharuk I. (2020) 'Development of Information Technology for Planning Order Fulfillment at a Food Enterprise', *Eastern-European Journal of Enterprise Technologies*, vol. 1, no. 3(103), pp. 62–73.
2. Fayyad U., Piatetsky-Shapiro G., Smyth P. (1996) *Knowledge Discovery and Data Mining* [online]. URL: https://www.academia.edu/8638675/Knowledge_Discovery_and_Data_Mining_Towards_a_Unifying_Framework.
3. Rehman N. (2017) 'Data Mining Techniques, Methods, Algorithms, and Tools', *Int. j. comput. sci. inf. technol. & mobile computing*, vol. 6, is. 7, July 2017, pp. 227–231.
4. Грибков С. В., Кожушко І. В. (2021) 'Використання автоматизованого збору даних для управління виконанням договорів', *Матер. 87 Міжнар. наук. конф. молодих уч., аспір. і студ. «Наукові здобутки молоді — вирішенню проблем харчування людства у XXI ст.»*, 15–16 квітня 2021 р., К.: НУХТ, ч. 2., с. 370.
5. Cloudflare (2022) *Cloudflare Developers* [online]. URL: <https://developers.cloudflare.com>.
6. *Puppeteer* [online] (2022) URL: <https://pptr.dev>.