

МЕТОД ОТРИМАННЯ ЗОБРАЖЕНЬ С ВЕБ-РЕСУРСУ ШЛЯХОМ ПАРСИНГУ

Збирати дані з веб-сайтів (тобто займатися веб-скрейпінгом) можна або через готовий АРІ, або шляхом парсинга. Самостійно парсити сторінки не завжди просто: багато сайтів не люблять скрейперов і намагаються блокувати їх.

Метою доповіді є огляд методу отримання даних з веб-ресурсу шляхом парсингу.

Веб-скрейпінг шляхом парсингу став важливим інструментом для автоматизованого збору інформації в Інтернеті. Він є частиною маркетингових інформаційних систем (MIS) для формування баз даних або банків даних, завдяки яким потрібні відомості надаються у формі табличних звітів.

Веб-скрейпінг шляхом парсингу є формою автоматичного копіювання даних, в якій певна необхідна для отримання інформація збирається з Інтернету і акумулюється, як правило, в центральну локальну базу даних або електронну таблицю для подальшого вилучення та аналізу. Парсер для скрейпінга веб-сторінок може отримати доступ до всесвітньої павутини безпосередньо через веб-браузер.[1]

Отриманий вміст сторінки, зокрема зображення, можна аналізувати, копіювати в електронну таблицю і так далі. [2]

В основному веб-скрейпери вирішують наступні завдання:

- Пошук необхідної інформації;
- Копіювання даних з Інтернету;

- Моніторинг оновлень на сайтах.

Веб-скрейпінг може бути як самостійним інструментом і служити для цільового пошуку інформації, також він може стати компонентом веб-розробок, які використовуються для веб-індексації, веб-майнінг та інтелектуального аналізу даних, онлайн-моніторингу, зміни цін і їх порівняння, для спостереження за конкуренцією, і другого збору даних. [3]

Література

1. Ткачов В.М. Програмний кластер для паралельної обробки великих обсягів даних / В.М. Ткачов, Ю.А. Кривобоков, К.П. Гвоздецька // Міжнародна наукова інтернет-конференція "Інформаційне суспільство: технологічні, економічні та технічні аспекти становлення (випуск 49)" / Збірник тез доповідей: випуск 19 (м. Тернопіль, 10 червня 2020 р.). – Тернопіль. – 2020. – 31-33 с.
2. Krivoulya G. Implementation of mobile eye tracking systems for preventing emergency situations based on monitoring of driver behavior / Krivoulya G., Tokariiev V., Tkachov V., Hunko M // Проблеми інформатизації : тези доп. 7-ї міжнар. наук.-техн. конф., 13-15 листопада 2019 р., м. Черкаси, м. Харків, м. Баку, м. Бельсько-Бяла : [у 3 т.]. Т. 3 / Черк. держ. технолог. ун-т [та ін.]. – Харків, 2019. – С. 36.
3. Voropaieva K.A. Parametric Model Of The Information System Determination Of Prohibited Products Food / K.A. Voropaieva, V.M. Tkachov // Дев'ята міжнародна науково-технічна конференція «Сучасні напрями розвитку інформаційно-комунікаційних технологій та засобів управління». – Баку-Харків-Жиліна. – 11-12 квітня 2019 р. – С. 30.