

Автор:**Джурляк Андрій Володимирович**

студент 5 курсу

Спеціальності 122 Комп'ютерні науки

Науковий керівник:

кандидат педагогічних наук,

старший викладач кафедри комп'ютерної та програмної інженерії

Галицький Олександр Вадимович**ПРОГРАМНИЙ МОДУЛЬ ВЕБСАЙТУ ЕЛЕКТРОННОГО ВИДАННЯ ДЛЯ ЗБОРУ СТАТИСТИЧНИХ ВІДОМОСТЕЙ ПРО ПУБЛІКАЦІЇ**

Анотація. У сучасному світі відомості є найціннішим ресурсом, а її найбільшим джерелом є глобальна мережа. У таких випадках мова йде про величезні обсяги даних, опрацювати, які вручну – неможливо. Тому автоматизований збір та опрацювання даних є надзвичайно корисним і важливим процесом.

Ключові слова: вебсайт, парсинг, дані.

Вступ. Одною із основних потреб людського життя є дані. Ми з кожним днем дізнаємось щось нове, читаєм, дивимось телебачення, слухаємо радіо, спілкуємося з людьми. На сьогоднішній день пошукові системи збирають і видають величезні обсяги відомостей, які можуть бути важливими для певних груп користувачів. Інколи людина, користуючись глобальною мережею, втрачає багато свого дорогоцінного часу, щоб знайти необхідні відомості, адже потрібно переглянути не один вебсайт, наштовхуючись на рекламу, інтерфейс деяких вебсайтів є доволі таки складним і знайти там щось потрібне потребує багато часу. Інколи людина заходить в «глухий кут», потрібні дані не знайдені, залишається обмаль часу на їх пошук. Виникає потреба в пришвидшеному процесі збору відомостей по різних вебсайтах.

Мета дослідження: є розробка програмного модуля вебсайту електронного видання для збору статистичних відомостей про публікації.

Виклад основного матеріалу. Для швидкого опрацювання відомостей використовуються програми парсери. А процес цей називається Вебскрапінгом або парсингом.

Простими словами парсинг – це автоматизований збір відомостей з будь-якого вебсайту, їх аналіз, перетворення та видача у структурованому вигляді, найчастіше у вигляді таблиці з набором даних. Проте суспільство ставиться до цього досить своєрідно – багато хто цим займався, але ніхто про це не говорить публічно. Більше того, парсинг часто засуджується і вважається чимось ганебним. Причина, як і в більшості подібних випадків, у неправильному сприйнятті.

Парсер – це програма, яка призначена для автоматизації постопрацювання отриманих з вебсайту даних. Іншими словами, парсер - це програма з аналізу та перетворення тексту з метою виділити з нього певні фрагменти або видалити зайве. На відміну від людини, парсер швидко і безпомилково відокремлює потрібні і відкидає зайві відомості з HTML документу, і ефективно упакує результат у певному форматі.

Виникає запитання: “Чи законно збирати дані з веб-сайту?”. Так, це законно, тому що відомості в інтернеті загальнодоступні.

Деякі люди сприймають парсинг як DDOS-атаку та ставляться до нього із сумнівом. Однак, це зовсім різні речі, і при парсингу ми, навпаки, намагаємося якнайменше

навантажувати цільовий вебсайт і не нашкодити бізнесу. Як у випадку зі здоровим паразитизмом – ми не хочемо, щоб бізнес «відкинув копита», інакше нам не буде на чому паразитувати.

У нашій роботі буде застосовуватися парсинг вебсайту з доменом e-journals.npu.edu.ua. І будуть збиратися такі дані:

- Поточний номер
- Кількість спеціальностей
- Загальну кількість тез
- Кількість тез по окремій спеціальності
- Отримано Подань
- Прийнято Подань
- Відхилено Подань
- Опубліковано Подань

Висновок. У ході написання роботи було досліджено шляхи парсингу вебданих у глобальній мережі, проблеми які виникають. Визначено актуальність задачі, роль програми-парсера. Досліджено інструментарій за допомогою якого можна реалізувати дане завдання, а також проаналізовано варіанти синтаксичного розбору вебсторінок, імпорту та експорту даних. Глибоко ознайомлено зі способами захисту від парсингу та методами їх обходу.

Список використаних джерел:

1. Lapinskyi, V. V., Mykytenko, P. V., Halytskyi, O. V. (2021). DESIGN OF MEDICAL INFORMATION SYSTEMS USER INTERFACE. Information Technologies and Learning Tools, 85(5), 1-13. <https://doi.org/10.33407/itlt.v85i5.44072>.
2. Охлопков Д. Как спарсить любой сайт?. Habr. URL: <https://habr.com/en/post/579336/> (дата звернення: 28.01.2023).
3. Парсинг сайтів: що це і навіщо він потрібен?. Webpromo. URL: <https://web-promo.ua/ua/blog/parsing-sajtov-hto-eto-i-zachem-nuzhen/> (дата звернення: 28.01.2023).
4. Парсинг: що це таке, для чого потрібен, як захиститись. Dalistrategies. URL: <https://dalistrategies.com/ua/shho-take-parsing-i-dlya-chogo-vikoristovuietsya/> (дата звернення: 28.01.2023).
5. Учасники проектів Вікімедіа. Web scraping – Вікіпедія. Вікіпедія. URL: https://uk.wikipedia.org/wiki/Web_scraping (дата звернення: 28.01.2023).