

ПРИРОДНИЧІ ТА ТЕХНІЧНІ НАУКИ

УДК 004.4:001.891

ЗАСТОСУВАННЯ МЕТОДУ ПАРСИНГУ ДЛЯ ЕФЕКТИВНОГО ПОШУКУ ІНФОРМАЦІЇ У ДОСЛІДНИЦЬКІЙ ДІЯЛЬНОСТІ

Ю.О. Вакуленко, О.С. Щербіна

Анотація. У статті розкрито сутність парсингу як автоматизованого методу пошуку наукової інформації. Досліджено конкретний приклад пошукової машини Web Scraper, визначено його роль в дослідницькій діяльності. Методологічною основою роботи є системний підхід, а також застосування підходу наукової об'єктивності, критичного та структурно-системного підходу до літературної та джерельної бази роботи. Специфіка досліджуваної теми передбачає використання методу парсингу для підвищення ефективності пошуку інформації у дослідницькій діяльності.

Ключові слова: парсинг; пошук інформації; дослідницька діяльність; пошукова машина; скрапінгові системи.

У процесі наукового дослідження завжди виникає потреба у великому обсязі інформації, для того щоб її опрацювати та використати у власній науковій роботі. Проте науковцеві потрібна не будь-яка інформація, не будь-які відомості, факти, ідеї, а лише ті відомості, які є найбільш цінними й актуальними для його напряму наукового дослідження. Тому, на сьогодні залишається актуальним питання правильного методу пошуку достовірних жерел інформації, який, по-перше – буде забезпечувати швидке отримання бажаних результатів заданим запитом, а, по-друге – надаватиме інформаційну глибину результатів та корисність даних для самого дослідника.

Для проведення наукових досліджень пошук джерел інформації є одним із ключових етапів наукового дослідження, тому можна зазначити низку критеріїв пошуку джерел інформації для проведення такої роботи:

1. Визначення кола питань, що будуть вивчатись;
2. Хронологічні межі пошуку необхідних джерел інформації;
3. Уточнення можливості використання зарубіжних джерел інформації;
4. Уточнення видів джерел інформації (книги, дисертації, статті, видання тощо);
5. Визначення ступеня відбору джерел інформації – усі з одного питання чи тільки окремі матеріали.

З кожним роком ускладнюється пошук потрібної інформації. Тому всі науковці мають знати основні методи пошуку інформації.

Метод пошуку інформації – це сукупність операцій, спрямованих на знаходження джерел інформації, які потрібні для розроблення теми наукової роботи. [1]

Методи пошуку джерел інформації наведені у таблиці 1.

Таблиця 1 – Методи пошуку джерел інформації

Метод пошуку	Характеристика методу
Ручний	Здійснюється за бібліографічними картками, картотеками, друкованими вказівниками
За допомогою Інтернет	Інформацію отримують з Web-сайтів Інтернет
Аналіз рукописних та друкованих джерел	Отримують науково-практичні джерела інформації у галузі дослідження
Автоматизований	Використовують комп'ютерні пошукові програми (наприклад, парсинг) для аналізу електронних бібліотек і баз даних

Розглянемо більш детально автоматизований метод пошуку наукової інформації.

Задля пошуку такої важливої і достовірної інформації застосовують парсинг або метод парсингу

Парсинг – це процес аналізу або розбору певного контенту на складові за допомогою роботів-парсерів (спеціальних програм або скриптів).

Найвідоміші парсери в мережі – це пошукові роботи, які аналізують сторінки, зберігають дані аналізу у себе в базі і потім при пошуку видають релевантні та актуальні документи або ту ж саму достовірну інформацію.

Парсинг сайтів є ефективним рішенням для автоматизації збору інформації, адже ця програма має здатність знаходити серед тисячі сайтів потрібну інформацію, відкидати зайве, упаковувати дані в необхідному вигляді, і в подальшому науковець уже може забирати релевантні дані для їх використання у створенні та розробці своєї наукової роботи.

За допомогою нього можна оцінити як швидко програма може впоратись із поставленим завданням, чи достовірна та чи інша інформація, чи взагалі відчувається релевантність у процесі роботи тощо. [2]

Метод парсингу має кілька етапів для витягнення потрібної інформації із наукових сайтів:

1. Збір контенту.

Пошукова машина парсинг 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, розділи статей, наукових робіт інших авторів в яких він може знайти для себе корисні дані та використати їх у своєму науковому дослідженні. Парсер 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 даних інших наукових робіт.

Запропонована методика дасть можливість дослідити та проаналізувати дану інформацію, чи потрібна вона дослідникові для внесення її у свою роботу.

Підсумовуючи вищесказане, можна зазначити, що метод парсингу істотно прискорює процес роботи особливо, якщо є ключові слова. Налаштувавши роботу, можливо оперативно підібрати необхідні для просування запити, тому що парсер за короткий термін обходить тисячі сторінок, фільтрує представлені дані, відбираючи серед них потрібні, після чого пакує отриманий результат для подальшої обробки. Це дає можливість дослідникові відстежувати появу нових наукових робіт інших авторів, щоб науковець зміг взяти щось корисне для себе і використати це у своїх дослідженнях. [4]

Для більш детального вивчення та ознайомлення з методом парсингу, наведемо конкретний приклад пошукової машини.

Web Scraper (від англ. Scraping – «вишкрібання»), веб-збирання або витягнення веб-даних) являє собою перетворення у структуровані дані інформації з веб-сторінок, які призначені для перегляду людиною за допомогою браузера.

Як правило, виконується за допомогою комп'ютерних програм, що імітують поведінку людини в інтернеті, або з'єднуючись з веб-сервером напряму по протоколу HTTP, або управляючи повноцінним веб-браузером. Але буває і скрапінг за допомогою копіювання даних людиною. Це форма копіювання, в якій конкретні дані збираються та копіюються з інтернету, як правило, в базу даних або електронну таблицю для подальшого пошуку чи аналізу. [3]

Web Scraper включає в себе завантаження та вилучення. Спочатку завантажуються сторінка (що робить браузер, коли ви переглядаєте сторінку), після цього можна

здійснювати добування потрібної інформації. Зміст сторінки може бути проаналізовано, переформатовано, його дані скопійовані в електронну таблицю тощо. Веб-скрапери, як правило, беруть щось із сторінки, щоб використати це для інших цілей. Прикладом може бути пошук і копіювання імен та телефонних номерів або компаній та їх URL-адрес до списку (контактне сканування).

Веб-сторінки побудовані за допомогою текстових мов розмітки (HTML та XHTML) часто містять велику кількість корисних даних у текстовій формі. Однак більшість веб-сторінок призначені для кінцевих користувачів, а не для зручності автоматичного використання. Через це були створені набори інструментів, які «збирають» веб-вміст. Веб-скрапери – це прикладний програмний інтерфейс для вилучення даних з веб-сайту.

Існують методи, які деякі веб-сайти використовують для запобігання веб-скрапінгу. Наприклад, виявлення та заборона ботів від сканування (перегляду) своїх сторінок. У відповідь на це існують веб-скрапінгові системи, які спираються на використання методів аналізу об'єктної моделі документа, комп'ютерного бачення та обробку тексту на природній мові, щоб імітувати пошук людини, щоб дозволити збирати вміст веб-сторінок для автономного синтаксичного аналізу.

Але ця програма не дивлячись на те, що є розповсюджена між дослідниками та допомагає їм у пошуках потрібної для них інформації, має свої переваги та недоліки. Переваги та недоліки пошукової машини Web Scraper наведені в таблиці 2.

Таблиця 2 – Переваги та недоліки пошукової машини Web Scraper

Переваги	Недоліки
Автоматизація роботи сайту	Вимагає подальшого редагування інформації
Працює автономно	Знайдена інформація не унікальна, деколи її публікація без посилання на першоджерело або згоди правовласника порушує авторські і корпоративні права
Миттєве отримання інформації	Налаштування програми проводяться перед запуском і для змін потрібно втручання програміста
Висока точність групування інформації	При зміні структури сайтів-донорів іноді спостерігаються помилки в роботі пошукової машини, що вимагає переналаштування під нові реалії
Скорочує деякий час для пошуку потрібної інформації	Не може повністю замінити роботу людини

Ефективність досліджень за допомогою парсера значною мірою залежить від використання джерел інформації. Зазвичай дослідники використовують офіційні та перевірені сайти, на яких не має фальшивої інформації для їх наукових робіт.

Тому для того щоб уникнути неправдивих даних, вагомо використовувати парсинг, де можна вказувати тему пошуку та ключові слова для полегшення знаходження інформації для науковця. [5]

Отже, підсумовуючи вищесказане, можна зазначити, що використання саме парсингу для підвищення ефективності пошуку інформації для дослідників є найбільш дієвим способом, особливо, якщо є ключові слова. Налаштувавши роботу, можна оперативно підібрати необхідні для просування запити, тому що парсер за короткий термін обходить тисячі сторінок, фільтрує представлені дані, відбираючи серед них потрібні, після чого пакує отриманий результат для подальшої обробки.

Аннотація. В статті раскрыто сутність парсинга как автоматизированного метода поиска научной информации. Исследовано конкретный пример поисковой машины Web Scraper, определено роль данной поисковой машины в исследовательской деятельности. Методологической основой работы является системный подход, а также применение подхода научной объективности, критического и структурно-системного подхода к литературной базе работы и источникам. Специфика исследуемой темы предусматривает использование метода парсинга для повышения эффективности поиска информации в исследовательской деятельности.

Ключевые слова: парсинг; поиск информации; исследовательская деятельность; поисковая машина; скрапинговые системы.

Abstract. The article reveals the essence of parsing as an automated method of searching for scientific information. A specific example of the Web Scraper search engine is investigated, the role of this search engine in research is determined. The methodological basis of the work is a systematic approach, as well as the application of the scientific objectivity approach, a critical and structural-systematic approach to the literary base of work and sources. The specifics of the topic under study provides for the use of the parsing method to increase the efficiency of information retrieval in research activities.

Keywords: parsing; search for information; research activities; search engine; scraping systems.

СПИСОК ЛІТЕРАТУРИ

1. Семантика. Парсинг сайту. Етапи парсингу даних. URL : semantica.in/blog/chto-takoe-parsing.
2. Frost R., Hafiz R. and Callaghan P. (2008) «Parser Combinators for Ambiguous Left-Recursive Grammars» 10th International Symposium on Practical Aspects of Declarative Languages (PADL), ACM-SIGPLAN, Volume 4902/2008, P. 167–181, January 2008, San Francisco.
3. Парсинг html-сайтів для споживача. Парсинг даних для використання підприємством. URL : parsing.valemak.com/ru/what-why-how/stages-of-parsing.
4. «Баст» веб-студія. Парсинг сайту. Послуги парсинга сайтів. Переваги парсингу. URL : webstudiobast.com/uk/poslugu-parsinga-sajtiv-dlya-opencart.
5. Парсинг. Метод парсингу. Де використовується парсинг. Особливості його застосування. URL : alexsmokinof.lviv.ua.

УДК 582.284:543.061

ВПЛИВ КОНЦЕНТРАЦІЇ ВОДНО-ЕТАНОЛЬНИХ РОЗЧИНІВ НА ВИДІЛЕННЯ ФЕНОЛЬНИХ СПОЛУК З ПОРОШКІВ ВИЩИХ ГРИБІВ

Т. С. Корман, І. В. Столяр, Ю. О. Лесишина

Анотація. В роботі наведені результати дослідження впливу концентрації водно-етанольних розчинів на виділення фенольних сполук із порошків вищих грибів *Flammulina velutipes* і *Lentinus edodes*, культивованих в Україні. Встановлено, що у складі водно-етанольних екстрактів порошків вищих грибів містяться фенольні сполуки різної природи. Показано, що максимальний вихід фенольних сполук має місце при використанні як екстрагенту 50 %-ого розчину етанолу. Загальний вміст фенольних сполук у складі екстрактів із порошку *Flammulina velutipes*, одержаних за оптимальних умов, майже в 3 рази перевищує вміст фенольних сполук у складі екстрактів із порошку *Lentinus edodes*.

Ключові слова: вищі гриби, *Flammulina velutipes*, *Lentinus edodes*, фенольні сполуки, екстрагування.

Вступ.

Пошук нових джерел біологічно активних речовин з метою одержання ефективних і безпечних продуктів є важливою задачею сучасної біотехнології [1]. Перспективною сировиною для одержання різноманітних дієтичних, лікувально-профілактичних і лікарських засобів є вищі гриби – базидіоміцети. Хімічний склад базидіоміцетів є оптимальним для забезпечення фізіологічних функцій організму людини речовинами і енергією, а наявність унікальних біологічно активних речовин – полісахаридів, меланінів, антибіотиків, стероїдів, низькомолекулярних фенольних сполук – зумовлює їх комплексну імунорегулюючу, цитопротекторну, антиоксидантну та інші дії [2].

Безперечними лідерами серед лікарських грибів-базидіоміцетів є *Lentinus edodes* (японський гриб, шиїтаке), *Ganoderma lucidium* (рейши або лінчжи), *Flammulina velutipes* (зимовий гриб, зимовий опеньок). В медичній практиці їх застосовують як продукти харчування, у вигляді чаїв, водних і спиртових екстрактів; використовують для одержання індивідуальних речовин з фармакологічною активністю. Обсяг виробництва цих грибів у країнах світу постійно зростає, в тому числі і в Україні [3].

Мета даної роботи полягала у дослідженні якісного і кількісного складу водно-етанольних екстрактів із порошків грибів *Flammulina velutipes* і *Lentinus edodes*, культивованих в Україні, і визначенні впливу концентрації екстрагенту на виділення