

Технічні науки

Прасолов Андрій Павлович

студент,

Національний технічний університет України

«Київський політехнічний інститут»

Прасолов Андрей Павлович

студент,

Национальный технический университет Украины

«Киевский политехнический институт»

Prasolov A.

student

National Technical University of Ukraine

«Kyiv Polytechnic Institute»

**СИСТЕМА АНАЛІЗУ НАУКОВЦІВ УНІВЕРСИТЕТУ ПО
ПАРАМЕТРАМ GOOGLE SCHOLAR
СИСТЕМА АНАЛИЗА УЧЕНЫХ УНИВЕРСИТЕТА ПО ПАРАМЕТРАМ
GOOGLE SCHOLAR
THE SYSTEM FOR THE ANALYSIS OF PARAMETERS OF THE
UNIVERSITY SCIENTINST USING THEIR PARAMETERS IN GOOGLE
SCHOLAR**

Анотація: Розроблено систему для аналізу параметрів науковців університету за їх параметрами у Google Scholar.

Ключові слова: Google Scholar, аналіз, парсинг.

Аннотация: Разработана система для анализа параметров ученых университета по их параметрам в Google Scholar.

Ключевые слова: Google Scholar, анализ, парсинг.

Summary: The system was developed for the analysis of parameters of the university scientists using their parameters in Google Scholar.

Keywords: Google Scholar, analysis, parsing.

Вступ

Google Scholar – це величезна база статей науковців з різних університетів, міст, країн та континентів, але ця система має обмежений функціонал. Виникла необхідність перегляду статистики параметрів науковців різних наукових підрозділів, таких як кафедри та факультети та можливість їх швидкого порівняння, тому було прийнято рішення розробити систему для аналізу даних науковців з бази Google Scholar.

Опис системи

«Система аналізу науковців Університету по параметрам GOOGLE SCHOLAR» аналізує науковців по 3 параметрам: присутності у GOOGLE SCHOLAR, цитованості та Н-індексу.

- **Присутність у Scholar** показує кількість статей науковця у базі Google Scholar.
- **Цитованість** показує кількість бібліографічних посилань на статті науковця у базі Google Scholar.
- **Індекс Гірша** - це показник впливовості науковця, заснований на кількості його статей та їх цитування.

Цитованість та Н-індекс підраховує GOOGLE SCHOLAR, а присутність визначається кількістю робіт автора, що входять у GOOGLE SCHOLAR.

Система парсить базу GS університету за адресою: https://scholar.google.com.ua/citations?view_op=view_org&hl=uk&org=5596117057032671997 і збирає дані по кожному автору, id якого знаходиться в базі системи. Також система може шукати потрібних науковців та парсити їхні дані двома доступними користувачам методами (пошук по id та пошук по імені).

Дані перевіряються і заносяться у базу системи. З точки зору актуальності – парсинг дає актуальну інформацію на час парсингу, а дані у базі на час перезапису чи зміни бази, але, враховуючи повільну динаміку зміни даних в базі GOOGLE SCHOLAR, різниця між даними в базі системи та актуальними може бути дуже малою, або може бути взагалі відсутньою, отож дані з бази можуть «реально» демонструвати параметри науковців на поточний момент.

Слід відзначити, що парсинг достатньо об'ємна обчислювальна процедура і на неї витрачається багато часу, тому пошук по ід та пошук по імені обмежені всього 10 полями, адже чим більша кількість науковців перевіряється, тим більше часу це займе.

Методи пошуку:

Першим є метод, коли користувач знає персональні ід науковців, дані про яких необхідно зібрати. Сторінка для пошуку по ід показана на скріншоті нижче (рис.1).

В ліве поле вводиться ім'я, воно буде записане у таблицю навпроти параметрів, отриманих з сторінки, яка належить науковцю, ід якого записане у поле праворуч.

Name1	id1
Name2	id2
Name3	id3
Name4	id4
Name5	id5
Name6	id6
Name7	id7
Name8	id8
Name9	id9
Name10	id10

Перевірити

Якщо не заповнювати поля, то будуть застосовані дані за замовчуванням.
Перевірка займе певний час.

Рис.1 – Перевірка параметрів науковців за їх ID

Другий метод здійснює пошук науковця по імені. Пошук по імені є досить не точним і це не проблема даного сервісу. Проблема полягає в тому, що ім'я науковця може мати декілька транскрипцій та може існувати декілька

людей з однаковими/подібними прізвищами. Пошук обмежений першою десяткою знайдених за цим запитом науковців. Якщо пошук не дав результатів, то ви отримаєте помилку. Сторінка для пошуку по імені показана на наступному скріншоті (рис.2).

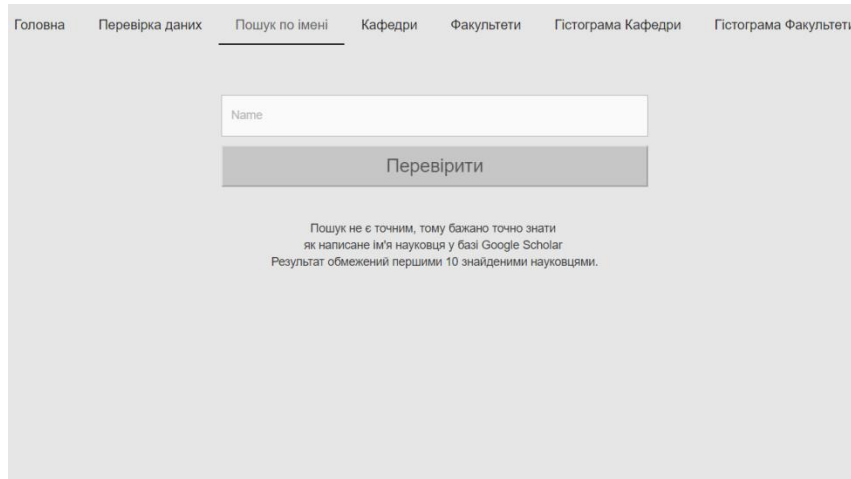


Рис.2 – Пошук параметрів науковця по його імені

Розроблена система має зручний, інтуїтивно зрозумілий Веб інтерфейс як для користувача (науковця), так і для адміністратора сайту.

Система дозволяє через парсинг шукати науковців по прізвищу, імені та ІД у GOOGLE SCHOLAR, а також знаходити суму даних для співробітників підрозділів. Також таблицю з даними можна вивантажити собі на комп'ютер у файлі формату csv, що легко відкривається за допомогою програми Microsoft Excel. Присутня можливість сортувати дані в таблиці за будь-яким із стовпчиків, для цього необхідно просто натиснути на назву стовпчика, за яким необхідно провести сортування.

На наступному скріншоті (рис.3) надана таблиця для кафедри СП.

Головна Перевірка даних Пошук по імені Кафедри Факультети Гістограма Кафедри Гістограма Факультети

Ім'я	Присутність у Scholar	Цитованість	Індекс Грша	Посилання
Анатолій Петренко	101	351	11	Посилання
Олексій Бритов	6	131	1	Посилання
Валерій Рогоза	23	22	3	Посилання
Олександр Безносик	21	18	2	Посилання
Валерій Стіканов	4	18	3	Посилання
Анна Кисельова	17	17	3	Посилання
Геннадій Кисельов	15	15	3	Посилання
Олексій Сергєєв-Горчинський	11	15	2	Посилання
Олег Цурін	21	10	2	Посилання
Костянтин Харченко	24	4	1	Посилання
Олексій Чкалов	12	4	1	Посилання
Богдан Булах	7	2	1	Посилання
Вікторія Гюргізова-Гай	6	2	1	Посилання
Валерій Романов	1	0	0	Посилання
Віталій Артюхов	4	0	0	Посилання
Олег Капшук	1	0	0	Посилання
Павло Свірін	7	0	0	Посилання
Богдан Кирюша	11	0	0	Посилання

Загальна кількість присутності у GS: 292
Загальна кількість цитованостей: 609

Всього

Вивантажити таблицю

Copyright © 2016 Прасолов, Цурін КПІ-Телеком

Рис.3 – Приклад виведеної таблиці для кафедри

Для порівняння структурних підрозділів будуються відповідні гістограми, як це показано на скріншоті нижче (рис.4).

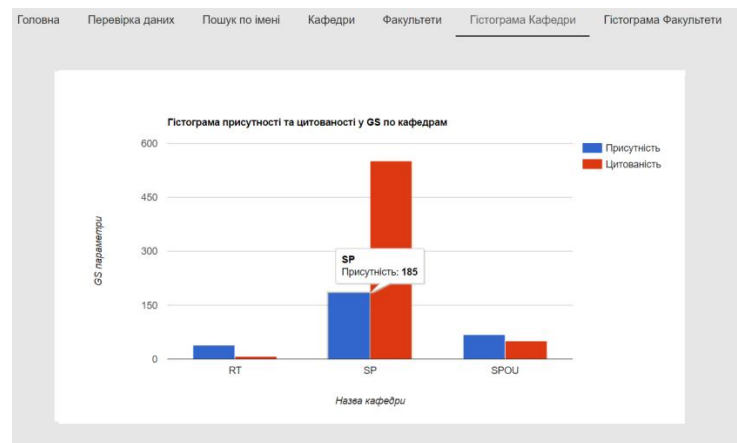


Рис.4 – Приклад гістограми для кафедр

Гістограма на вищезазначеному скріншоті побудована на тестовому наборі даних. Вона подвійна, тобто для кожного структурного підрозділу ми отримуємо по два стовпчики: синім показана присутність (кількість статей у GS), червоним цитованість (кількість бібліографічних посилань на ці статті). При збільшенні кількості підрозділів стовпчики будуть звужуватися. Також при наведенні на стовпчик, можна побачити цифру, що відповідає його значенню.

Висновки

Було розроблено систему для аналізу даних науковців університету по їх параметрам (присутність, цитованість, Індекс Гірша) з бази Google Scholar. Розроблена система відповідає поставленим до неї вимогам (вивід даних у вигляді таблиць для кожного підрозділу та можливість вивантаження цих таблиць; побудова гістограм по підрозділам). Також було описано принцип роботи цієї системи.

Література:

1. About Google Scholar. – Режим доступу:

<https://scholar.google.com/intl/en/scholar/about.html> – Дата доступу:
22.07.2016.