Tikrit Journal of Pure Science Vol. 30 (5) 2025

DOI: https://doi.org/10.25130/tjps.v30i5.1821

IRAQI Academic Scientific Journals

Tikrit Journal of Pure Science

ISSN: 1813 – 1662 (Print) --- E-ISSN: 2415 – 1726 (Online)

Journal Homepage: https://tjpsj.org/

Tikrit Journal of Pure Science

Bibliometric Analysis of Scientific Research in Iraqi Universities Using Google Scholar

Kanaan J. Brakas[1], Mafaz Alanezi[2]

[1]*General Directorate of Kirkuk Education, Kirkuk, Iraq.*

[2]*ICT Research Unit, Computer Center, University of Mosul,Iraq*

Received: 11 Nov. 2024 **Received in revised forum:** 5 Feb. 2025 **Accepted:** 11 Feb. 2025

Final Proof Reading: 13 Oct. 2025 **Available online:** 25 Oct. 2025

ABSTRACT

Google Scholar plays an essential position in evaluating researchers' activities and influencing the scientific status of universities. This look at focuses on Iraqi universities, where improving studies output is essential for fostering medical competition and improving worldwide ratings. We advanced a framework to display and analyze bibliometric information from Google Scholar profiles of Iraqi university faculty, using advanced internet scraping strategies to gather complete statistics. Our framework additionally identifies false articles and citations by staring at creator behavior over an exact period, supplemented by way of administrative evaluations for inconsistencies. Addressing gaps in cutting-edge bibliometric analyses, our studies present treasured insights to decorate the medical recognition of Iraqi universities and sell an aggressive academic environment. This study now not only contributes to knowledge studies efficiency and its impact on university ratings but also offers sensible suggestions for improving instructional requirements and visibility. Despite the essential position of Google Scholar, there's a lack of comprehensive frameworks to efficiently screen and examine bibliometric data, in Iraqi universities. Enhancing research output and identifying false articles and citations are essential to provide accurate tests of studies efficiency and educational requirements, in the end benefiting the medical popularity of Iraqi universities.

Keywords: Bibliometric Analysis, web scraping, Google Scholar, University Rankings, Scientific Metrics

Name: Mafaz Alanezi **E-mail:** mafazmhalanezi@uomosul.edu.iq

Tikrit Journal of Pure Science Vol. 30 (5) 2025

DOI: https://doi.org/10.25130/tjps.v30i5.1821

التحليل الببليومتري للبحث العلمي في الجامعات العراقية باستخدام كوكل سكولار

كنعان جلال براخاس[1]، مفاز العنزي[2]

[1] *مديرية تربية كركوك، كركوك، العراق*

[2] *وحدة بحوث تكنولوجيا المعلومات والاتصالات، مركز الحاسبة الالكترونية، جامعة الموصل، الموصل، العراق*

الملخص

تلعب خدمة Google Scholar دورًا أساسيًا في تقييم أنشطة الباحثين والتأثير على الوضع العلمي للجامعات. تركز هذه الدراسة على الجامعات العراقية، حيث يعد تحسين مخرجات الدراسات أمرًا ضروريًا لتعزيز المنافسة الطبية وتحسين التصنيفات العالمية. لقد طورنا إطارًا لعرض وتحليل المعلومات الببليومترية من ملفات تعريف Google Scholar لأعضاء هيئة التدريس بالجامعات العراقية، باستخدام استراتيجيات web scraping المتقدمة لجمع إحصائيات كاملة. يحدد إطارنا أيضًا المقالات والاقتباسات الزائفة من خلال النظر في سلوك المؤلف على مدى فترة زمنية محددة، مع استكمالها بالتقييمات الإدارية للتناقضات. من خلال معالجة الثغرات في التحليلات الببليومترية الحديثة، تقدم دراساتنا رؤى ثمينة لتعزيز الاعتراف الطبي بالجامعات العراقية وتعزيز بيئة أكاديمية تنافسية. لا تساهم هذه الدراسة الآن في كفاءة الدراسات المعرفية وتأثيرها على تصنيفات الجامعات فحسب، بل تقدم أيضًا اقتراحات عملية لتحسين المتطلبات التعليمية والرؤية. على الرغم من المكانة الأساسية لخدمة Google Scholar، إلا أن هناك نقصًا في الأطر الشاملة لفحص وفحص البيانات الببليومترية بكفاءة، وخاصة في الجامعات العراقية. إن تعزيز مخرجات البحث وتحديد المقالات والاستشهادات المزيفة أمر ضروري لتوفير اختبارات دقيقة لكفاءة الدراسات والمتطلبات التعليمية، مما يعود في النهاية بالفائدة على الشعبية الطبية للجامعات العراقية.

Introduction

More idea has been given to website accessibility because of the big advancements in web sites and ongoing statistics sharing [(1)]. On the opposite side, the quick advancement of statistics generation, mainly cloud computing and open-source software program, gives quite a few possibilities that college libraries can exploit to higher serve purchaser demands [(2)]. Evaluating college faculty studies is critical for international ratings like QS and Shanghai. Bibliometric analysis is important for assessing the legitimacy and impact of scholarly contributions, specifically in Iraqi universities, wherein there is a developing emphasis on academic achievement by using worldwide standards. Our study addresses a vast hole by means of providing an in-depth evaluation of studies output and its effect the usage of Google Scholar metrics [(3)].

Unlike previous research specializing in wide bibliometric critiques, this study examines higher schooling in Iraq. Using advanced web scraping techniques to extract and analyze information from Google Scholar, it affords a comprehensive view of Iraq's academic environment, assessing researchers' and institutions' performances and figuring out possibilities and demanding situations inside the Iraqi educational machine.

Our work altogether provides grasping clinical creation and its guidelines for university ratings and notoriety on each neighborhood and international scales. The Iraqi Ministry of Higher Education and Scientific Research commands that researchers and university teachers maintain up with dynamic Google Scholar profiles, making this research basic. Our analysis gives primary insights into the adequacy of this guiding principle and its effect on

instructional visibility and norms. Moreover, the observe gives novel information amassing and analysis strategies, setting a benchmark for destiny bibliometric research in related fields. This research fills an indispensable hole as well as presents viable suggestions for enhancing the worldwide acknowledgment and scholarly results of Iraqi universities. In general, this study gives a more comprehensive comprehension of the condition of the field, filling in as an important device for researchers, teachers, and lawmakers planning to encourage a climate of academic greatness and development.

Google Scholar has many elements, which can be viewed as a critical resource for researchers, academics, and universities. It offers the chance to make a profile for each researcher, master, or instructor. Through this, one can get to the main publications and research papers of academic organizations, offering fantastic abilities and offices for researching huge and present-day academic writing in all areas of science and references. This addresses an extraordinary wealth for researchers and academics to stay aware of all new publications in their field of specialization [4]. It additionally gives the number of citations to articles, permitting universities and mindful authorities to study and gauge the effect of an author's research in the scientific and academic fields. Google Scholar likewise empowers the disclosure of scientific joint efforts around articles published by authors and co-authors, making it more straightforward for universities to find the scientific interests of researchers at universities and other academic foundations with comparative scientific interests. This helps to increase the quality of published articles in terms of type and number and improve fruitful cooperation between specializations and different academic institutions at both local and global levels.

Google Scholar provides essential metrics like the h-index, i10-index, citation counts, publication dates, and the number of research papers, aiding researchers in accurate analysis and evaluation. It is widely used by Iraqi researchers and faculty for its rich, open-source resources, which we utilized in our proposal to analyze data from Iraqi researchers and universities. However, it has drawbacks, such as lacking strict quality control and including unreliable sources, affecting analysis accuracy. Citation counts are less comprehensive than services like Scopus and Web of Science. Additionally, it allows the addition of fake research to profiles due to non-unique identification, violating original researchers' intellectual property rights.

The data and information provided by Google Scholar are essential for conducting an evaluation and analyzing the level of researchers and their scientific works, thus assessing the level and reputation of the university and comparing it with other universities to improve the spirit of competition between universities and researchers who belong to those universities. However, the analysis process directly from the Google Scholar website is very complicated because it does not provide a comprehensive summary of the researchers' data and compile it in an easily analyzable way. Therefore, we had to use special techniques to collect data from Google Scholar using web scraping.

Web scraping relates to the activity of extracting data and information from websites. Various types of web scraping techniques are used, such as static scanning, which involves retrieving data from a website with a consistent HTML structure and is suited to websites that lack dynamic content or undergo unexpected changes [5]. Dynamic scraping entails retrieving data and information from web pages that use JavaScript to load content dynamically. Tools like Selenium may be applied to automate interactions with a internet site and gather the preferred statistics. Finally, API scraping refers to using utility programming interfaces (APIs) supplied by some web sites, permitting builders to retrieve dependent information. This era includes

interacting with those APIs and extracting the favored facts and records [6,7].

The rest of the paper is structured as follows: Section presents an outline of associated works. In Section Three, we provide a detailed method of the proposed model, which incorporates subsections including facts collection, Methodological Justification, Limitations and Implications, and statistics garage. Moving ahead, Section Four gives Data Analysis, Interpretation, and Addressing Research Objectives. Section 5 summarizes the conclusions drawn from the observe and descriptions future avenues of labor. Finally, Section Six accommodates the references stated on this paper."

This paper makes several giant contributions to the sphere of bibliometric analysis and educational assessment:

Novel Framework for Data Collection and Analysis: We advanced an modern framework the use of superior internet scraping techniques to collect and examine bibliometric statistics from Google Scholar profiles of Iraqi university school. This approach presents a comprehensive angle on academic performance and complements the accuracy of research exams.

Identification of False Articles and Citations: Our framework consists of mechanisms to hit upon false articles and citations through monitoring author behavior over the years. This addresses a important gap in present bibliometric analyses, ensuring the reliability and integrity of the facts.

Insights into Iraqi Higher Education: By focusing at the particular context of Iraqi universities, our take a look at offers precious insights into the challenges and opportunities in these educational surroundings. Our findings contribute to understanding a way to enhance research output and visibility in line with worldwide standards.

Practical Recommendations: We gift sensible recommendations for reinforcing the scientific reputation of Iraqi universities. These consist of strategies for enhancing educational requirements, growing visibility, and fostering a competitive educational environment.

Related work

In this scholarly examination, we delved into the evaluation of new research courses that caught the attention of Google Scholar. The look at initiated by[8] Aimed at summarizing the e-book developments of medical papers using web scraping strategies to extract statistics from Google Scholar. This effort efficiently harvested information for 238 researchers and 2523 articles. The retrieved records become systematically saved in an SQL database and utilized to unveil vast insights into article publishing, highlighting info like creator profiles, affiliations, quotation counts, and titles with outcomes exportable in PDF or Excel codecs. Pratiba et al. (9) focused on creating an interface employing web scraping methods and Python to link authors' Google Scholar publications directly to MySQL and Excel, facilitating easier access and management of scholarly works. Martín-Martín et al.[10] conducted an expansive citation analysis, uncovering that Google Scholar accounted for the highest citation percentages compared to other platforms, with particular discrepancies noted across different disciplines and databases. Victoriano et al.[11] examined and tracked the publication outputs of Bulacan State University's faculty, aiming to aggregate and monitor the institution's scientific contributions indexed on Google Scholar. Parikh et al. [12] demonstrated how machine learning could effectively identify bots, enhancing web traffic and user behavior visualization using Kibana and converting unstructured web data into structured formats for better analysis. Han et al.[13] provided guidance on efficiently collecting online hotel data, emphasizing the potential and pitfalls of analyzing data from various online platforms. Schedlbauer et al.[14] highlighted the significance of soft skills in medical informatics education, using web crawling, extraction, and text mining to underscore the

balance between technical knowledge and experiential learning. López-Cózar et al.[15] aimed to elucidate the functioning of Google Scholar for bibliometric analyses, exploring the search engine's attributes and the potential bibliometric applications based on its data. McCoy et al.[16] developed an interactive bibliometric ranking website, leveraging Google Scholar citations to feature over 1300 biomedical informatics researchers worldwide. Lastly, Chertow et al.[17] Applied bibliometric techniques to explore the key elements of commercial symbiosis within business ecology, demonstrating Google Scholar's application in shooting a broader variety of articles in comparison to traditional databases, as a consequence presenting a more comprehensive view of the educational panorama on this discipline.

While these studies provide treasured insights into bibliometric evaluation and records extraction from Google Scholar, they disclose top notch gaps and weaknesses. Primarily, these works often lack a comprehensive framework that addresses the precise challenges confronted by way of Iraqi universities, where there is a pressing need to enhance research output and visibility according to global standards. Additionally, the strategies employed may not sufficiently cope with the detection of false articles and citations, that could extensively effect the accuracy of bibliometric critiques. The reliance on Google Scholar's records also offers obstacles, as it includes low-exceptional or unreliable resources and lacks complete citation coverage in comparison to databases like Scopus and Web of Science. These gaps underscore the necessity for an improved and context-specific method to bibliometric evaluation tailor-made to the specific academic landscape of Iraq.

METHODOLOGY

Our study makes use of a complete and methodical approach custom-made to cope with the splendid challenges of bibliometric evaluation in Iraqi universities. Recognizing the requirement for nitty gritty and particular records, we used advanced net scraping equipment to collect data straightforwardly from Google Scholar. This stage is broadly utilized by Iraqi scholars and offers a rich assortment of academic papers alongside metrics such as citation counts, h-index, and i10-index scores.

The choice to utilize web scraping as our essential information assortment method is legitimate by its capacity to effectively obtain exceptional and comprehensive information, fundamental for our broad and variable dataset. We utilized both static and dynamic web scraping techniques, leveraging tools like BeautifulSoup for static substance and Selenium for dynamic substance. This approach mechanizes the assortment of information from various Google Scholar profiles, essentially reducing the time and exertion contrasted with manual methods.

Furthermore, web scraping works with an organized assortment of information from otherwise unstructured web pages, resulting in a more streamlined and precise analysis. In spite of potential challenges such as website updates and information inaccuracies, we executed hearty approval methods and continuously adjusted our scraping techniques to guarantee the unwavering quality and heartiness of our study.

Data Collection

We utilized both static and dynamic web scraping strategies to assemble all fundamental data successfully, see **Figure 1**. For gathering fundamental insights regarding authors and their works, static web scraping was used, exploiting Google Scholar profile pages' stable HTML setup. This technique fits perfectly for non-dynamic content, enabling quick and straightforward data collection without interacting with the web page components.

On the other hand, we applied dynamic web scraping for profiles and pages that load data asynchronously, typically through JavaScript. This method, which often uses tools like Selenium, allows for simulating user interactions on the web

page, accessing data not readily available in the HTML source. By combining these two techniques, we secured a detailed dataset that covers a broad spectrum of metrics, essential for a thorough bibliometric analysis.

The two methods were used in the current study at different levels of data collection. This study was implemented in four Iraqi universities: the University of Mosul, the University of Baghdad, the University of Basra, and the University of Salahaddin, depending on the ID of the university. Static web scraping was used at the first level to obtain general data for the faculty members, authors, and academics affiliated with each university. Data such as the author's name, total number of citations, author's ID, name of the university, field of interest in which the author's articles are published (specialization), URL of the author's page, and other information were obtained. Afterward, the data extracted from the first level of research were relied upon, and the authors who were to have more private data extracted in the second level were determined by relying on the dynamic web scraping method. This method was chosen because the research data are presented using JavaScript, especially for authors who have more than twenty article papers. The data extracted include the number of published articles, the college to which the author belongs, the name of the university, h-Index, i10-Index, and other data.

The third level concerns data related to published articles, such as the titles of the articles, the number of citations for each article, the year of publication, the name of the journal in which they were published, and other data that will be mentioned in the analysis stage.

The fourth level involves extracting data related to the current author and their co-authors, such as their IDs, university names, specialties, co-author page URLs, etc.

Methodological Justification

Our decision to use web scraping as the main method for data gathering is supported by its proficiency in obtaining current and comprehensive data, a necessity for the broad and fluctuating dataset needed for our research. This technique enables the automated accumulation of information from numerous Google Scholar profiles, markedly lowering the time and labor compared to traditional manual methods. Besides, web scraping considers the coordinated collection of data from the generally unstructured web pages, prompting a more smoothed out and exact examination.

Limitations and Implications

Despite the fact that web scraping brings many advantages, we should concede its deficiencies, similar to potential errors coming about because of updates to the format or content of websites. To handle these difficulties, we executed an itemized information approval system and reliably changed our scraping procedures to stay up with modifications in the structure of Google Scholar pages. We likewise resolved the moral issues attached to web scraping, guaranteeing our activities were in accordance with Google Scholar's use terms and appropriate legitimate norms.

By the by, there may as yet be predispositions and irregularities in the information given by Google Scholar, for example, fragmented records of citations or contrasts in the culmination of creator profiles, which could impact our review's outcomes. We attempted to counter these disadvantages by approving our assembled information against other bibliometric data sets when achievable and by straightforwardly tending to what these difficulties could mean for the finishes of our exploration.

DOI: https://doi.org/10.25130/tjps.v30i5.1821

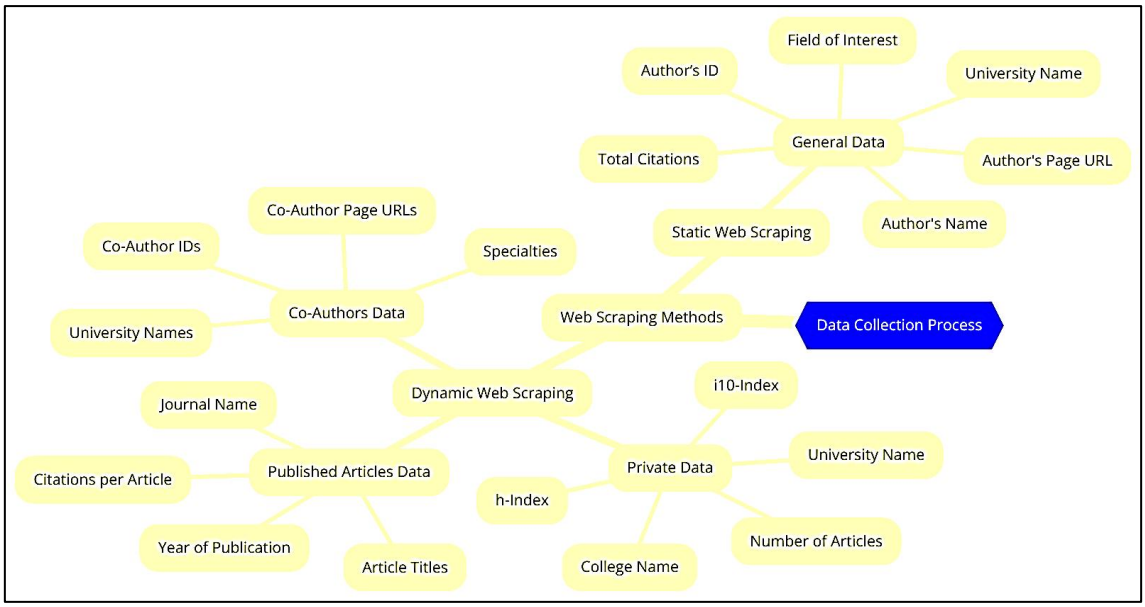


Fig. 1: Data Collection Process

Data Storing

The data collected from the first level of data scraping are saved in the primary table in the database, named 'university-staff'. This table contains 4,368 authors from the University of Baghdad, 3,305 authors from the University of Mosul, 1,566 authors from the University of Basra, and 996 authors from Salahaddin University, totaling 10,235 authors in this study. The main input relied upon is the author's ID, which is used to access the author's page in the second level of data scraping.

At this stage, general information about each author, such as the h-index, i10-index, total number of citations, and others, as well as information related to the first level of web scraping, is collected and stored in a table named 'Author-info', see figure2.

Then, relying on the university's ID and the author's ID, the third level of data extraction begins. Through this, we collect information about each author's published research papers, such as the titles of the research, the number of published research papers, the number of citations for each research paper, the year of publication, and other information. These data are stored in a table called 'Research-info.'

Following this, the fourth level of data extraction begins, through which we collect information related to the collaborations of authors (co-authors) and their information. This is stored in a table called 'Co-Author-info', which contains details about them. Using the database facilitates the processes of data input, modification, and accelerates the analysis and evaluation of the universities and their authors.

DOI: https://doi.org/10.25130/tjps.v30i5.1821

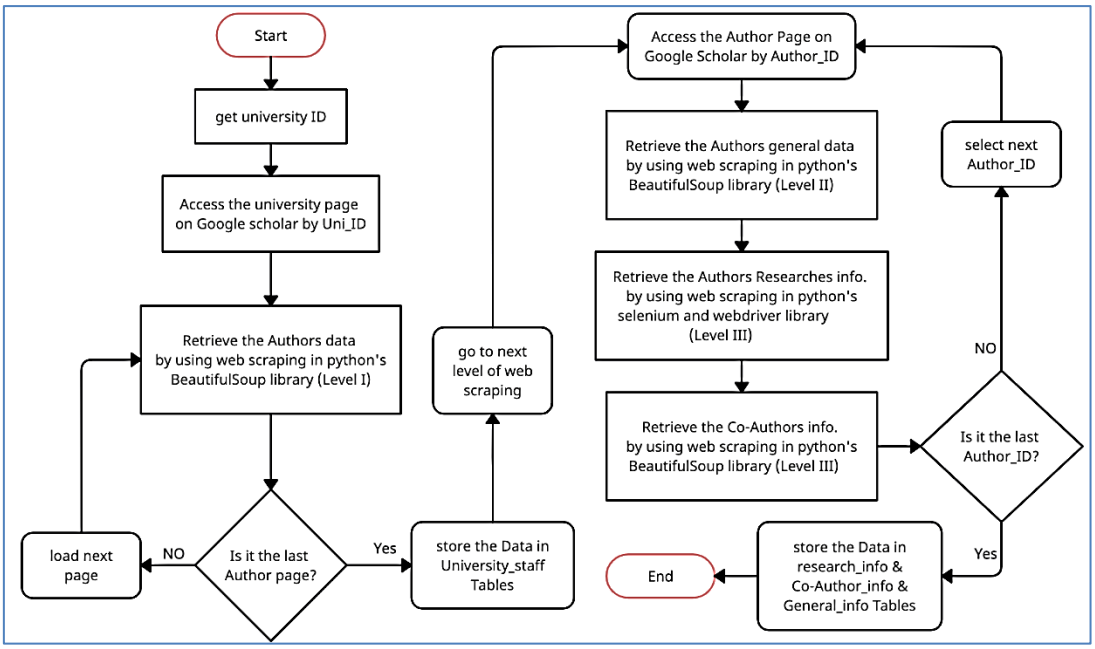


Fig. 2: Data collection and storing diagram.

Data Analysis and Interpretation

The analytical stage of our investigation was meticulously designed to extract significant insights from the extensive dataset collected via web scraping. Once obtained, the data were analyzed through multiple layers, encompassing both quantitative and qualitative dimensions. Our primary goal was to evaluate the scholarly productivity and influence of Iraqi academic institutions as reflected on Google Scholar.

Data Analysis Process

At first, we quantitatively surveyed the gathered data, looking at citation counts, h-index, and i10-index figures for each researcher, as well as the complete distribution count per organization. This approach gave an itemized perspective on the research scene across Iraqi universities, highlighting predominant patterns and distribution habits. We utilized factual methods such as regression analysis and ANOVA to analyze bibliometric indicators across various universities, fields, and time spans. This similar analysis not only uncovered the general standings of different organizations but additionally shed light on the advancement of research yields over the long haul. Visual guides, including bar charts and dispersed plots, were utilized to outline key discoveries. All insightful methods were thoroughly lined up with our review targets, meaning to recognize factors impacting the scholarly glory of Iraqi universities both locally and globally.

The system was designed to display basic information from the 'university-staff' table on the main page, allowing users to identify or choose one or a specific group of authors. The search function enables users to display authors based on any information present within that table, such as specialty, college, university, or number of citations for authors. This facilitates the process of analysis and evaluation. After determining the specific choice of authors, several types of analyses are prepared, as shown in Figure 3:

A. General Information: Analyze the general information by using the database tables to obtain or calculate this information:

1. Total number of articles already published.
2. Total number of citations in the last five years.
3. The number of h-index and i10-index.
4. The number of h-index and i10-index in the last five years.
5. Number of articles published each year.

DOI: https://doi.org/10.25130/tjps.v30i5.1821

Author Name: [Jasim M. Mahdi]		University ID: [University of Baghdad]	
University: [University of Baghdad]		Author ID: [tBmWtucAAAAJ]	h_index: [25]
Fields: [(Thermal Energy Storage) (Thermal Management) (Renewable Energy Systems)		Total Researches: [77]	Last 5 Years h_index: [25]
V_Email: [coeng.uobaghdad.edu.iq]		Total Citations: [3027]	i10_index: [38]
College: [Assistant Professor, Department of Energy E]		Last 5 Years Citations: [2982]	Last 5 Years i10_index: [38]
In year 2018 : Number of Researches 5		In year 2017 : Number of Researches 5	In year 2020 : Number of Researches 6
In year 2019 : Number of Researches 3		In year 2021 : Number of Researches 18	In year 2022 : Number of Researches 20
In year 2016 : Number of Researches 1		In year 2023 : Number of Researches 11	In year 2024 : Number of Researches 4
In year 2013 : Number of Researches 1		In year 2012 : Number of Researches 1	In year 2010 : Number of Researches 1
Global Level Co_Author 15		Local Level Co_Author 5	University Level Co_Author 0

Fig. 3: Analyzing the general author's information.

B. Author Collaboration at different levels: locally, at the level of Iraqi universities, and globally, at the international university level.

C. Article Information: Analyze the article's information by using the database's tables to obtain or calculate this information, as shown in Figure 4:

1. The article title.
2. The year of publication.
3. The number of citations for each article.
4. The URL of each article.
5. The name of the journal in which the article was published.
6. The names of the first, second, …, and sixth authors.

University	Author ID	URL	Title	Publish Year	Journal	Cite By	1st Author	2nd Author	3rd Author	4th Auth	5th Author	6th
University of Baghdad	tBmWtucAAAAJ	https://schol	Solidification enhancement of PCM in a triplex-	2018	Applied Energy	251	JM Mahi	EC Nsofc				
University of Baghdad	tBmWtucAAAAJ	https://schol	Melting enhancement in triplex-tube latent hea	2017	Applied energy	246	JM Mahi	EC Nsofc				
University of Baghdad	tBmWtucAAAAJ	https://schol	Accelerated melting of PCM in energy storage	2018	international Jc	241	JM Mahi	S Lohras	DD Gan	EC Ns		
University of Baghdad	tBmWtucAAAAJ	https://schol	Solidification enhancement with multiple PCMs	2020	Applied Energy	231	JM Mahi	HI Mohar	ET Hasl	P Talc	EC Nso	
University of Baghdad	tBmWtucAAAAJ	https://schol	Hybrid heat transfer enhancement for latent-h	2019	International Jc	221	JM Mahi	S Lohras	EC Nso			
University of Baghdad	tBmWtucAAAAJ	https://schol	Simultaneous energy storage and recovery in t	2019	Energy convers	206	JM Mahi	S Lohras	DD Gan	EC Ns		
University of Baghdad	tBmWtucAAAAJ	https://schol	Melting enhancement in triplex-tube latent the	2017	International Jc	182	JM Mahi	EC Nsofc				
University of Baghdad	tBmWtucAAAAJ	https://schol	Solidification enhancement in a triplex-tube lat	2017	Energy 126, 5(	162	JM Mahi	EC Nsofc				
University of Baghdad	tBmWtucAAAAJ	https://schol	Solidification of a PCM with nanoparticles in tri	2017	Applied Therma	152	JM Mahi	EC Nsofc				
University of Baghdad	tBmWtucAAAAJ	https://schol	Multiple-segment metal foam application in the	2018	Journal of Ener	139	JM Mahi	EC Nsofc				

Fig. 4: Analyzing the Article information.

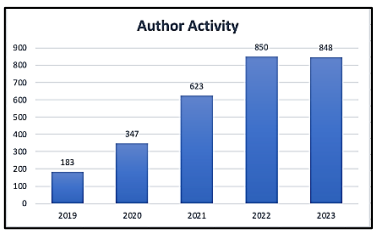
D. Collaboration Author Information: Analyze the co-author's information by using the database's tables to obtain or calculate this information, as shown in Figure 5:

1. Co-author ID.
2. Co-author name.
3. Co-author's validation email.
4. URL of the co-author.

University	Author ID	Co_Author ID	Name	University	Email	URL
University of Baghdad	tBmWtucAAAAJ	C0DL55kAAAAJ	Dr. Hayder I. Mohammed	Physics department / Collage of Edu	email at garmian.edu.k	https://scholar.google.co
University of Baghdad	tBmWtucAAAAJ	WbTuZdcAAAAJ	Emmanuel C. Nsofor	Professor, Mechanical Engineering D	email at siu.ed	https://scholar.google.co
University of Baghdad	tBmWtucAAAAJ	zDW_NYMAAAAJ	Nidhal Ben Khedher	Mechanical engineering, College of E	email at uoh.edu.s	https://scholar.google.co
University of Baghdad	tBmWtucAAAAJ	4waThoEAAAAJ	Mohammad Ghalambaz	PhDVerif	email at tdtu.edu.vn	https://scholar.google.co
University of Baghdad	tBmWtucAAAAJ	2YjiDekAAAAJ	Sina Lohrasbi	EBNER GROUP, Upper Austria (Lohra	email at alumni.tugraz.a	https://scholar.google.co
University of Baghdad	tBmWtucAAAAJ	PMsRCKuNZNMC	Mohammadreza Ebrahimnataj Tiji	Qom University of TechnologyVerifie	email at qut.ac.	https://scholar.google.co
University of Baghdad	tBmWtucAAAAJ	rHeDOikAAAAJ	Dr Azher M Abed	r conditioning and Refrigeration Engi	email at uomus.edu.iq	https://scholar.google.co
University of Baghdad	tBmWtucAAAAJ	ERaylKIAAAAJ	Assist. Prof Dr. Anmar Dulaimi	University of Warith Al-Anbiyaa, Iraq	email at uowa.edu.iq	https://scholar.google.co
University of Baghdad	tBmWtucAAAAJ	idF7cfwAAAAJ	Amir Hossein Eisapour	PhD Candidate, Toronto Metropolitar	email at torontomu.c	https://scholar.google.co
University of Baghdad	tBmWtucAAAAJ	mPZrDuAAAAAJ	Prof. Davood Domiri Ganji(D.D.Ganj	Mecha. Engin.,Babol Noshirvani univ	email at nit.ac	https://scholar.google.co

Fig. 5: Analyzing the Article information.

E. Last Five-Year Activity Chart: Analyze the activity of an author over the last five years by using the database's tables to obtain or calculate this information, as shown in Figure 6.



Author ID: [tBmWtucAAAAJ]	year1 [2019]	year2 [2020]	year3 [2021]	year4 [2022]	year5 [2023]
University: [University of Baghdad]	value1 [183]	value2 [347]	value3 [623]	value4 [850]	value5 [848]

Fig. 6: Analyzing the Author's Activity.

Interpretation of Results

In this part of our work, we made sure that each analysis and interpretation directly supported our preset research goals. For every major discovery, we offered an in-depth explanation of its connection to the broader objectives of our investigation, aiming to boost scientific production and elevate the global standing of Iraqi universities.

We also pointed out any anomalies or unforeseen findings, providing possible reasons and pondering their consequences for future studies and policy development. This method ensured our conclusions were not just data-based but also significant and practical within the realm of Iraqi higher education and scholarly research.

Before beginning the analysis, it is crucial to select the researchers for the study. The selection is based on various criteria. In some cases, the focus might be on a single researcher, relying exclusively on their data. On the other hand, a gathering of researchers might be chosen in light of normal characteristics such as their academic affiliation. On the off chance that particular researchers need total data, such as subtleties of their university affiliation, the determination interaction should then rely upon accessible institutional data. Accordingly, researchers lacking comprehensive data are avoided from the analysis, influencing the study's accuracy. The essential justification behind these errors is the absence of a uniform data passage standard among Iraqi universities and the researchers' conflicting data accommodation, which is further exacerbated by Google Scholar's casual data section strategies, influencing data completeness and accuracy.

Furthermore, this study recognizes a few challenges related to Google Scholar, including the gamble of remembering made-up articles for a researcher's page without repercussions. The simplicity with which researchers can add deceptive works, whether purposefully or accidentally, entangles the approval of every published material, particularly across huge datasets, demanding significant investment and exertion. This issue emerges from Google Scholar's shortfall of rigid measures to safeguard protected innovation rights. To counter this, we suggest implementing a system to flag questionable or fraudulent articles for subsequent manual verification, thus streamlining the authentication process. This system would monitor an author's publication activity and raise alerts for unusual patterns, with investigations conducted to ascertain legitimacy. Furthermore, analyzing the rapid accumulation of citations could help identify whether the author has engaged in unethical citation practices, necessitating manual effort and additional resources like Scopus for a more accurate citation analysis, beyond what Google Scholar can provide.

The information in the previous section provided analyzes the data of an individual author. However, all of this analysis can also be applied to a group of authors who share certain characteristics, such as the university or college to which they belong, among other factors. The following is a presentation of the results obtained from analyzing the data of a group of authors, according to the name of the university, to facilitate comparisons between the four universities accredited in this study, see **Figure (7).** These comparisons will help in discussing the results, identifying the problems and difficulties faced, and assessing the accuracy of the results.

DOI: https://doi.org/10.25130/tjps.v30i5.1821

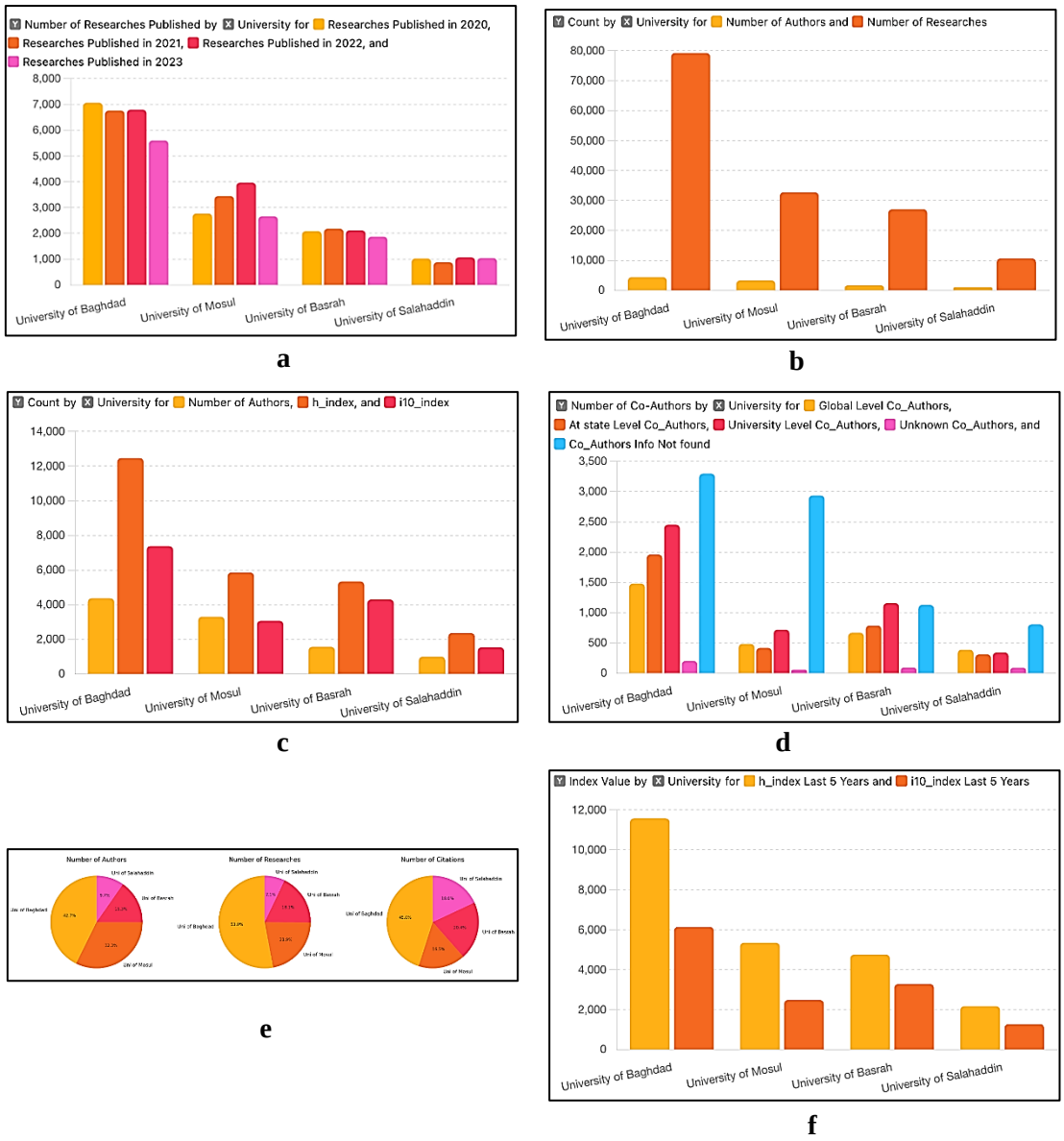


Fig. 7: Results Analysis, (a): Researches Published Over Years by Universities. (b): Number of Authors and Researches by Universities. (c): Authors and Index Metrics by Universities. (d): Co-Authors by Levels and Universities. (e): Number of Citations by Universities. (f): Recent h-index and i10-index (Last 5 Years) by Universities.

The charts in **Figure 7** provided showcase various key trends:

A. Researches Published Over Years by Universities, Figure 7(a):

Universities have shown varying trends in the number of researches published over the years from 2020 to 2023. Some universities have a consistent increase, while others show fluctuations.

B. Number of Authors and Researches by Universities, Figure 7(b):

There is a clear variation in the number of authors and researches across different universities. Some universities have a higher number of authors contributing to research, leading to a higher count of publications.

C. Authors and Index Metrics by Universities, Figure 7(c):

The number of authors, h-index, and i10-index vary significantly between universities. Universities with more authors generally have higher h-index and i10-

index values, indicating the impact and quality of their research output.

D. Co-Authors by Levels and Universities, Figure 7(d):

The levels of co-authors (University level, Global level) differ among universities. Some universities have a higher engagement at the global level, indicating international collaborations. Notably, it was observed that a significant number of researchers did not provide proper information regarding their collaborative partners. Specifically, of the 10,235 authors from the four universities, an astonishing 8,160 failed to acknowledge their co-authors, neglecting to supply any related information. This represents a significant 79% of the total, thus heavily influencing the results of the analysis.

E. Number of Citations by Universities, Figure 7(e):

The number of citations received by universities varies, showing the influence and reach of their research. Some universities have a significantly higher citation count, reflecting the impact of their work.

F. Recent h-index and i10-index (Last 5 Years) by Universities, Figure 7(f):

The h-index and i10-index over the last five years show the recent performance and impact of universities. Universities with higher recent indices are leading in research impact and productivity.

University Ranking

The normalization process was used to rank the universities based on the previous results. Normalization in this context involves scaling the values of different research metrics (such as research published, number of authors, h-index, i10-index, citations, and recent indices) to a common scale, typically from 0 to 1. This is done by dividing each metric's value by the maximum value observed for that metric across all universities, see the results of normalization to rank the universities in Table 1. This process considers fair comparison across various metrics by guaranteeing each metric contributes similarly to the all-out score, no matter what its unique scale. The standardized characteristics are then delivered to work out a whole score for every university, which is applied to rank them. The ranking chart in parent eight gives a visible representation of the colleges in mild of their standardized all-out scores across exceptional studies metrics.

Table 1: University Ranking

University	Published Researches	Number of Authors	h-index	i10-index	Citations	Recent h-index	Recent i10-index	Total Score	Rank
Uni of Baghdad	1	1	1	1	1	1	1	7	1
Uni of Mosul	0.75	0.8	0.8	0.8	0.6	0.8	0.8	5.35	2
Uni of Basrah	0.5	0.6	0.6	0.6	0.4	0.6	0.6	3.9	3
Uni of Salahaddin	0.25	0.2	0.4	0.4	0.2	0.4	0.4	2.25	4

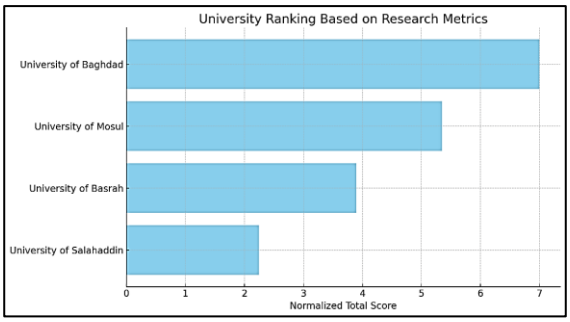


Fig. 8: University Ranking Based on Research Metrics

CONCLUSION

In conclusion, the usage of advanced web scraping strategies to research facts from Google Scholar profiles of Iraqi college students offers valuable insights into research performance, citation styles, and collaboration developments. The look at highlights the importance of up to date Google Scholar profiles in enhancing the visibility and effect of lecturers and their institutions. It additionally identifies areas for development, which include imposing strict statistics entry protocols to prevent instructional fraud. Despite relying solely on Google Scholar, the take a look at suggests integrating additional bibliometric resources like Scopus or Web of Science for a more complete evaluation. The findings provide sizeable pointers for improving the studies excellent and visibility of Iraqi universities and function a basis for destiny research to decorate educational excellence and cooperation.

Acknowledgments: We are grateful for all of the support supplied by the ICT Research Unit, Computer Center /University of Mosul/ Iraq.

Conflict of interests: The authors declared no conflicting pastimes.

Sources of funding: These studies did no longer acquire any precise provide from funding groups within the public, business, or not-for-profit sectors.

Author contribution: The authors contributed similarly to the have a look at.

REFERENCE

1. Hammood M. Testing the agreement among three free accessibility evaluation tools: a case study of Iraqi public universities. Tikrit Journal of Pure Science. 2019;24(3):133-40. https://doi.org/10.25130/tjps.v24i3.383

2. Abdallah SJ, Ouda GK. Using integrated library management systems for the improvement of information services based on cloud computing. Tikrit Journal of Pure Science. 2020;25(4):101-2. http://dx.doi.org/10.25130/tjps.25.2020.076

3. Nama GF, Lukmanul H, Junaidi J. Implementation of K-Means Technique in Data Mining to Cluster Researchers Google Scholar Profile. International Journal of Engineering and Advanced Technology (IJEAT). 2019;9(1). https://doi.org/10.35940/ijeat.a2708.109119

4. Triono T, Darmayanti R, Saputra ND, Afifah A, Makwana G. Open Journal System: Assistance and training in submitting scientific journals to be well-indexed in Google Scholar. Jurnal Inovasi dan Pengembangan Hasil Pengabdian Masyarakat. 2023;1(2):106-14). https://doi.org/10.61650/jip-dimas.v1i2.225

5. Morina V, Sejdiu S. Evaluating and comparing web scraping tools and techniques for data collection. UBT Knowl Cent. 2022.

6. Varshini KS, Uthra RA. Extraction of Meaningful Information from Unstructured Clinical Notes Using Web Scraping. Journal of Circuits, Systems and Computers. 2023;32(03):2350041. https://doi.org/10.1142/s021812662350041x

7. Krotov V, Johnson L. Big web data: Challenges related to data, technology, legality, and ethics. Business Horizons. 2023;66(4):481-91. https://doi.org/10.1016/j.bushor.2022.10.001

8. Rahmatulloh A, Gunawan R. Web scraping with HTML DOM method for data collection of scientific articles from Google Scholar. Indonesian Journal of Information Systems. 2020;2(2):95-104. https://doi.org/10.24002/ijis.v2i2.3029

9. Pratiba D, Abhay M, Dua A, Shanbhag GK, Bhandari N, SINGH U, editors. Web scraping and data acquisition using Google scholar. 2018 3rd International Conference on Computational Systems and Information Technology for

Sustainable Solutions (CSITSS); 2018: IEEE. https://doi.org/10.1109/csitss.2018.8768777

10. Martín-Martín A, Thelwall M, Orduna-Malea E, Delgado López-Cózar E. Google Scholar, Microsoft Academic, Scopus, Dimensions, Web of Science, and OpenCitations' COCI: a multidisciplinary comparison of coverage via citations. Scientometrics. 2021;126(1):871-906. https://doi.org/10.1007/s11192-020-03792-z

11. Victoriano JM, Pulumbarit JP, Lacatan LL, Salivio RAS, Barawid RLA. Data analysis of Bulacan State University faculty scientific publication based on Google Scholar using web data scraping technique. arXiv preprint arXiv:220206063. 2022. https://doi.org/10.25147/ijcsr.2017.001.1.85

12. Parikh K, Singh D, Yadav D, Rathod M. Detection of web scraping using machine learning. Open access international journal of Science and Engineering. 2018;3(1):114-8.

13. Han S, Anderson CK. Web scraping for hospitality research: Overview, opportunities, and implications. Cornell Hospitality Quarterly. 2021;62(1):89-104. https://doi.org/10.1177/1938965520973587

14. Schedlbauer J, Raptis G, Ludwig B. Medical informatics labor market analysis using web crawling, web scraping, and text mining. International Journal of Medical Informatics. 2021;150:104453. https://doi.org/10.1016/j.ijmedinf.2021.104453

15. López-Cózar ED, Orduña-Malea E, Martín-Martín A, Ayllón JM. Google Scholar: the big data bibliographic tool. Research Analytics: Auerbach Publications; 2017. p. 59-80. https://doi.org/10.48550/arxiv.1806.06351

16. McCoy AB, Sittig DF, Lin J, Wright A, editors. Identification and ranking of biomedical informatics researcher citation statistics through a Google Scholar Scraper. AMIA Annual Symposium Proceedings; 2020. https://pmc.ncbi.nlm.nih.gov/articles/PMC7153158

17. Chertow MR, Kanaoka KS, Park J. Tracking the diffusion of industrial symbiosis scholarship using bibliometrics: Comparing across Web of Science, Scopus, and Google Scholar. Journal of Industrial Ecology. 2021;25(4):913-31. https://doi.org/10.1111/jiec.13099

.