

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
Західноукраїнський національний університет
Факультет комп'ютерних інформаційних технологій
Кафедра комп'ютерної інженерії

Куренчук Андрій Сергійович

**Програмний модуль формування рейтингу
викладачів на основі показників Google Scholar та
Web of Science / The software module for forming the
rating of teachers based on Google Scholar and web of
science indicators**

спеціальність: 123 - Комп'ютерна інженерія
освітньо-професійна програма - Комп'ютерна інженерія
Кваліфікаційна робота

Виконав студент групи КІ-41
Куренчук Андрій Сергійович

Науковий керівник
к.т.н. Піцун О.Й.

2023

ЗМІСТ

Перелік умовних скорочень	9
Вступ.....	10
1 Технології та методології оцінювання рівня активності науковців	12
1.1 Статистика – наука про дані, основні поняття та визначення.....	12
1.2 Критерії оцінювання наукових працівників.....	16
1.3 Наукометричні бази даних та системи ідентифікації науковців.....	19
1.4 Висновки до розділу та постановка завдань кваліфікаційної роботи..	23
2 Методи та алгоритми оцінювання активності науковця на основі публікацій	25
2.1 Наукометрична база Web of Science.....	25
2.2 Наукометрична база Google Scholar	29
2.3 Алгоритм обчислення рейтингу наукового працівника.....	33
3 Програмний модуль визначення рейтингу науковця на основі публікацій.....	36
3.1 Структура, моделювання та функціональні можливості модуля.....	36
3.2 Функції взаємодії з серверами наукометричних баз даних	44
3.3 Тестування програмного додатку	47
4 Техніко-економічний розділ	Ошибка! Закладка не определена.48
4.1 Розрахунок витрат на розробку програмного додатку	Ошибка! Закладка не определена.48
4.2 Визначення експлуатаційних витрат.....	Ошибка! Закладка не определена.53
4.3 Визначення економічної ефективності та терміну окупності ..	Ошибка! Закладка не определена.57
Висновки	50
Список використаних джерел.....	51
Додаток А Програмний модуль взаємодії з наукометричною базою Web of Science.....	Ошибка! Закладка не определена.65

					КР.КІ.8351890.00.00.000 ПЗ		
Змн.	Лист	№ докум.	Підпис	Дата			
Розробив	Куренчук А.С.				Літ.	Арк.	Акрушів
Перевір.	Піцун О.Й.				8	49	
Консульт.	Савка Н.Я.				ЗУНУ, ФКІТ, КІ-41		
Н. Контр.	Мельник Г.М.						
Затвердив	Дубчак Л.О.						
ПРОГРАМНИЙ МОДУЛЬ ФОРМУВАННЯ РЕЙТИНГУ ВИКЛАДАЧІВ НА ОСНОВІ ПОКАЗНИКІВ GOOGLE SCHOLAR TA WEB OF SCIENCE							

Додаток Б Довідка про впровадження **Ошибка! Закладка не определена.**69

Додаток В Світлокопія виданої публікації**Ошибка!** **Закладка** **не**
определена.70

					КР.КІ.8351890.00.00.000 ПЗ			
Змн.	Лист	№ докум.	Підпис	Дата				
Розробив	Куренчук А.С.				ПРОГРАМНИЙ МОДУЛЬ ФОРМУВАННЯ РЕЙТИНГУ ВИКЛАДАЧІВ НА ОСНОВІ ПОКАЗНИКІВ GOOGLE SCHOLAR TA WEB OF SCIENCE	Літ.	Арк.	Акрушів
Перевір.	Піцун О.Й.						8	49
Консульт.	Савка Н.Я.					ЗУНУ, ФКІТ, КІ-41		
Н. Контр.	Мельник Г.М.							
Затвердив	Дубчак Л.О.							

ПЕРЕЛІК УМОВНИХ СКОРОЧЕНЬ

НБ	–	Наукометрична база
РЦ		Рейтинг цитування
КНП	–	Кількість наукових публікацій
АСОРАН	–	Автоматизована система оцінки рівня активності науковця
API	–	Application Programming Interface

					КР.КІ.8351890.00.00.000 ПЗ	Арк.
						9
Змн.	Арк.	№ докум.	Підпис	Дата		

ВСТУП

Актуальність роботи. В даний час дослідження ведуться по всьому світу практично з усіх предметів і генерують мільйони наукових статей та інших документів. У деяких випадках дослідження дають дуже хороші результати, в більшості випадків посередні, а в деяких випадках негативні результати. На підставі результатів дослідження приймаються рішення про нагороди, стипендії, просування по службі, відбір на роботу тощо. Для всіх цих випадків оцінка результатів досліджень стає обов'язковою умовою. Століття тому, коли дослідників було менше, дослідження оцінювали колеги. З плином часу збільшувалася кількість дослідників, розширювалися дослідницькі напрями, збільшувалася результативність досліджень. Сьогодні, навіть за помірними оцінками, налічується близько або більше одного мільйона дослідників, і вони випускають понад два мільйони дослідницьких робіт та інших документів на рік. У такій карколомній ситуації оцінка досліджень постійно виявляється важкою роботою. Для будь-якої нагороди та стипендії можуть бути десятки або сотні номінантів. З-поміж них постало велике питання, як вибрати найкращого кандидата. Експертні оцінки в багатьох випадках виявляються суб'єктивними. В результаті рішення стають упередженими. У 1963 році з'явився Індекс наукового цитування (SCI), який охоплює літературу з 1961 року. Через кілька років Юджин Гарфілд, засновник SCI, підготував список із 50 найбільш цитованих вчених на основі першого цитування автора SCI 1967 року. Вже наступного року, тобто 1969, двоє вчених, які фігурують у списку, напр. Дерек Х. Р. Бартон і Мюррей Гелл-Манн отримали Нобелівську премію. Це підтвердило корисність аналізу цитат. Якщо зі списку з 50 нагороду отримують двоє, це не є середнім досягненням для прогнозу. Це передбачення відкрило шлях для аналізу цитування, оскільки воно було вільним від суб'єктивності. Навіть для однолітків аналіз цитувань став корисним інструментом. Однак аналіз цитування не був вільним від недоліків. Для оцінки дослідження були потрібні інші показники. Аналіз цитування разом із рецензуванням забезпечили

					КР.КІ.8351890.00.00.000 ПЗ	Арк.
Змн.	Арк.	№ докум.	Підпис	Дата		10

кращий висновок у незліченних випадках. Необхідно було щось більше, щоб зробити рішення значною мірою надійним. Поява Всесвітньої павутини (WWW) дала наві можливість для розвитку ідей по отриманні універсальної оцінки активності науковців та їх вклад в розвиток наукової думки. Глобальна мережа зривувала кордони і тепер науковці з різних частин світу можуть вільно обмінюватись думками, досліджувати роботи один одного при цьому відпала потреба витрачати час на далекі подорожі або листування.

Проте це породило і низку проблем, оскільки доступ до інформації та генерування нових ідей став більш відкритий, то почало з'являтися безліч критеріїв оцінювання роботи науковці, або посилання на опубліковану статтю змінюються достатньо часто. А отже виникає необхідність створення програми яка могла б моніторити та повідомляти про зміни які відбулись. Тому задача по створенню програмного модуля оцінки активності науковця на основі аналізу його публікацій є актуальною задачею.

Метою кваліфікаційної роботи є розробка програмного модуля формування рейтингу викладачів на основі показників Google Scholar та Web of Science.

Для досягнення поставленої мети необхідно вирішити наступні завдання:

- 1) дослідити методології та технології оцінювання наукових працівників;
- 2) провести дослідження наукометричних баз даних;
- 3) проаналізувати особливості роботи з API Google Scholar;
- 4) проаналізувати особливості роботи з API Web of Science;
- 5) розробити алгоритм формування рейтингу викладачів на основі показників Google Scholar та Web of Science.

б) провести тестування спроектованої структури програмного модуля формування рейтингу викладачів на основі показників Google Scholar та Web of Science.

					КР.КІ.8351890.00.00.000 ПЗ	Арк.
						11
Змн.	Арк.	№ докум.	Підпис	Дата		

1 ТЕХНОЛОГІЇ ТА МЕТОДОЛОГІЇ ОЦІНЮВАННЯ РІВНЯ АКТИВНОСТІ НАУКОВЦІВ

1.1 Статистика – наука про дані, основні поняття та визначення

Статистика – це наука про збір, аналіз, інтерпретацію, представлення та організацію даних. Іншими словами, це математична дисципліна для збору, узагальнення даних. Крім того, можна сказати, що статистика – це розділ прикладної математики. Однак у статистиці є дві важливі та основні ідеї; це невизначеність і варіація. Невизначеність і варіацію в різних областях можна визначити лише за допомогою статистичного аналізу. Ці невизначеності в основному визначаються ймовірністю, яка відіграє важливу роль у статистиці. Поєднайте програмування та машинне навчання, і ви отримаєте гарний опис основних навичок науки про дані. Статистика використовується майже в усіх аспектах науки про дані. Вона використовується для аналізу, перетворення та очищення даних. Оцінює та оптимізує алгоритми машинного навчання. Вона також використовується для представлення думок і висновків. Наука про дані є міждисциплінарною сферою. Одним із будівельних блоків науки про дані є статистика. Без гідного рівня знань статистики було б дуже важко зрозуміти або інтерпретувати дані.

Статистика допомагає нам пояснити дані. Людство використовує статистичні дані, щоб отримати результати щодо сукупності на основі вибірки, зібраної з цієї сукупності. Крім того, машинне навчання та статистика мають багато збігів.

Під час розгляду статистики можна виділити 10 головних її принципів та вихначень (рисунок 1.1):

1. Популяція та вибірка. У статистиці весь набір необроблених даних, які ви можете мати для тестування чи експерименту, називають генеральною сукупністю. З ряду причин неможливо виміряти закономірності та тенденції в усьому населенні. З цієї причини статистика дозволяє нам взяти вибірку, виконати певні обчислення на цьому наборі даних і, використовуючи

					КР.КІ.8351890.00.00.000 ПЗ	Арк.
Змн.	Арк.	№ докум.	Підпис	Дата		12

ймовірність і деякі припущення, ми можемо з певним ступенем достовірності зрозуміти тенденції для всієї сукупності або передбачити майбутні події. Статистика дозволяє нам взяти вибірку або частину генеральної сукупності, виконати певні обчислення на цій сукупності даних і, використовуючи ймовірність і деякі припущення, ми можемо з певним ступенем достовірності зрозуміти тенденції для всієї сукупності

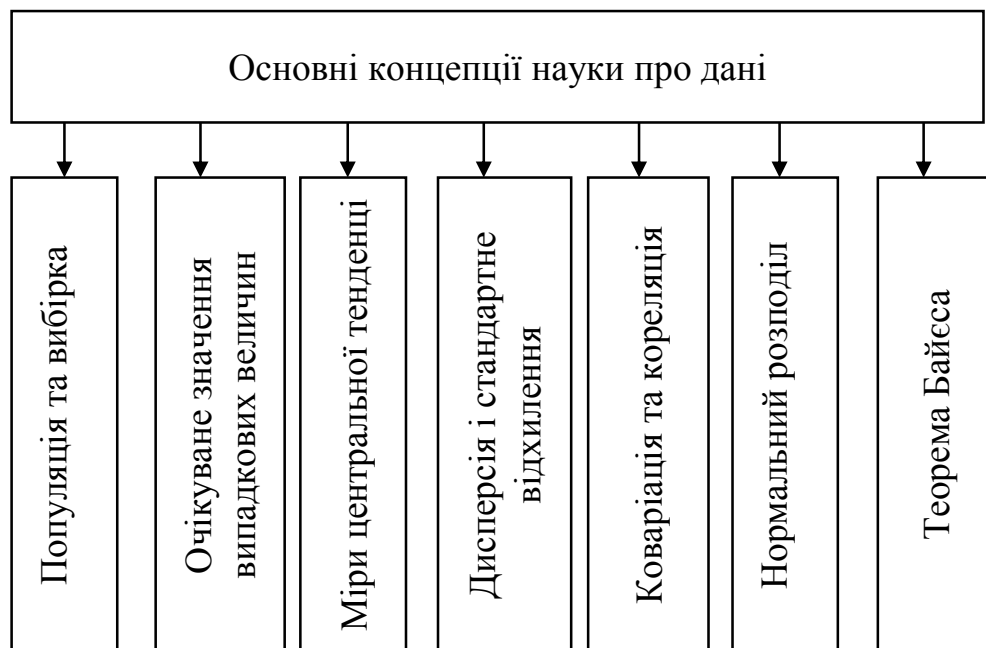


Рисунок 1.1 – Класифікація основних статистичних визначень та концепцій

2. Нормальний розподіл. Розподіл ймовірностей – це функція, яка показує ймовірності результатів події чи експерименту. Розглянемо функцію (тобто стовпець) у кадрі даних. Ця функція є змінною, і її функція розподілу ймовірностей показує ймовірність потрапляння значень, які вона може приймати. Функції розподілу ймовірностей дуже корисні в прогнозній аналітиці або машинному навчанні. Ми можемо робити прогнози щодо сукупності на основі функції розподілу ймовірностей вибірки з цієї сукупності.

Нормальний (гаусівський) розподіл – це функція розподілу ймовірностей, яка виглядає як дзвін.

Типова крива нормального розподілу. Вершина кривої вказує на найбільш імовірне значення, яке може прийняти змінна. У міру віддалення від піку ймовірність значень зменшується.

3. Міри центральної тенденції. Центральна тенденція – центральне (або типове) значення розподілу ймовірностей. Найпоширенішими показниками центральної тенденції є середнє значення, медіана та мода. Середнє – це середнє значення в серії. Медіана – це значення посередині, коли значення сортуються за зростанням або спаданням. Режим – значення, яке з'являється найчастіше.

4. Дисперсія і стандартне відхилення. Дисперсія є мірою варіації між значеннями. Вона розраховується шляхом додавання квадратів різниць кожного значення та середнього, а потім ділення суми на кількість зразків. Стандартне відхилення є мірою того, наскільки розкидані значення.

5. Коваріація та кореляція. Коваріація – це кількісна міра, яка показує, наскільки варіації двох змінних відповідають одна одній. Точніше кажучи, коваріація порівнює дві змінні з точки зору відхилень від їх середнього (або очікуваного) значення. Кореляція – це нормалізація коваріації стандартним відхиленням кожної змінної. Ця нормалізація скасовує одиниці, і значення кореляції завжди знаходиться між 0 і 1. Якщо порівнюємо зв'язок між трьома або більше змінними, краще використовувати кореляцію, оскільки діапазони значень або одиниці можуть викликати хибні припущення.

6. Центральна гранична теорема. У багатьох областях, включаючи природничі та соціальні науки, коли розподіл випадкової величини невідомий, використовується нормальний розподіл. Центральна гранична теорема (CLT) обґрунтовує, чому в таких випадках можна використовувати нормальний розподіл. Відповідно до CLT, коли ми беремо більше зразків із розподілу, середні значення вибірки будуть прагнути до нормального розподілу незалежно від розподілу сукупності. Нормальний розподіл описується в термінах середнього значення та стандартного відхилення, які можна легко обчислити. І якщо знаємо середнє значення та стандартне відхилення нормального розподілу, то можемо обчислити майже все про нього.

					КР.КІ.8351890.00.00.000 ПЗ	Арк.
Змн.	Арк.	№ докум.	Підпис	Дата		14

7. *P*-значення. *P*-value є мірою ймовірності значення, яке приймає випадкова змінна. Припустимо, що у нас є випадкова величина *A* і значення *x*. *P*-значення *x* – це ймовірність того, що обрана змінна *A* буде рівна значенню *x* або будь-яке значення, яке має такий самий або менший шанс спостерігатися.

8. Очікуване значення випадкових величин. Очікуване значення деякої наперед невідомої випадкової величини – це середньозважене значення всіх можливих значень змінної. Вага тут означає ймовірність виникнення події, що випадкова величина набере певного значення. Дискретні випадкові величини приймають скінченно багато або зліченно нескінченно багато значень. Кількість дощових днів у році є дискретною випадковою величиною. Безперервні випадкові величини приймають незліченну нескінченну кількість значень. Наприклад, час, який потрібно від вашого дому до офісу, є постійною випадковою змінною. Залежно від того, як його вимірювати (хвилини, секунди, наносекунди тощо), він приймає незліченну нескінченну кількість значень.

9. Умовна ймовірність. Ймовірність просто означає ймовірність події та завжди приймає значення рівне від 0 до 1 (0 і 1 включно). Ймовірність події *A* позначається як $p(A)$ і обчислюється як кількість бажаних результатів, поділена на кількість усіх результатів. Наприклад, коли кидаєте кубик, ймовірність отримати число менше трьох дорівнює $2/6$. Кількість бажаних результатів дорівнює 2 (1 і 2); загальна кількість результатів – 6. Умовна ймовірність – це ймовірність того, що деяка подія *A* відбудеться, якщо інша подія, пов'язана з подією *A*, вже відбулася.

10. Теорема Байєса. Згідно з теоремою Байєса, ймовірність події *A*, якщо подія *B* вже відбулася, може бути обчислена за допомогою ймовірностей події *A* та *B*, а також ймовірності події *B*, якщо *A* вже відбулася. Теорема Байєса є настільки фундаментальною та повсюдною, що існує поле під назвою «байєсовська статистика». У байєсівській статистиці ймовірність події або гіпотези як доказ вступає в гру. Тому попередні й апостеріорні ймовірності відрізняються залежно від доказів.

					КР.КІ.8351890.00.00.000 ПЗ	Арк.
Змн.	Арк.	№ докум.	Підпис	Дата		15

1.2 Критерії оцінювання наукових працівників

В оцінці наукових досліджень досить очевидно, що «наукова якість» є одним із найважливіших критеріїв, і в нашому випадку дослідницької діяльності вона залишається нарівні з іншими ключовим виміром: відповідністю розвитку. Критерії та процедури для оцінки дуже ідентичні тим, які застосовуються для оцінки досліджень загалом: експертні оцінки, панелі, бібліометричні індекси тощо. Практичне застосування цих інструментів у випадку дослідницької розробки іноді може бути досить складним. Проте, здається, існує загальний консенсус щодо того, що якість має виражатися так само, як і для будь-якого наукового дослідження, з тими ж проблемами та тими ж обмеженнями.

Про наукове ставлення дослідника можна судити лише з його діяльності, тобто з проведених ним досліджень і з його публікацій. Для дуже молодих дослідників нам, можливо, доведеться вдатися до інтерв'ю або до свідчень їхніх наставників. Загальна оціночна комісія може відповісти на такі питання:

Яка кількість публікацій по відношенню до кількості років, присвячених дослідженню, і враховуючи решту обов'язків особи (дидактичні та управлінські завдання, інші послуги громаді,...)?

Зважаючи на фінансові можливості, чи бере він/вона участь у конференціях з науковими доповідями та публікаціями в збірниках?

Який обсяг досліджуваних тем: чи є гарний розкид чи той самий предмет розглядається знову і знову? У будь-якому випадку, чи виправдано це характером досліджуваних проблем?

Також може бути корисним попросити оцінити старшого наукового співробітника, який особисто знає дослідника, якого оцінюють (хоча деякі можуть бути надто схильні давати позитивну оцінку...). Тут можуть бути такі запитання:

Чи цією людиною керує цікавість зрозуміти явища, чи вона задовольняється поверхневим описом?

					КР.КІ.8351890.00.00.000 ПЗ	Арк.
Змн.	Арк.	№ докум.	Підпис	Дата		16

Чи він ретельно опрацьовує всі деталі своєї роботи та усуває всі інші можливості, перш ніж робити поспішні висновки?

На риснку 1.2 наведено класифікацію основних груп критеріїв на основі яких відбувається оцінка наукових працівників.



Рисунок 1.1 – Недоліки ручної передачі показників електроенергії

Дослідники займаються наукою. Тому науковий успіх, як правило, приписують окремим особам (а не установам чи дослідницьким групам). Оскільки ці атрибуції можуть покращити або зруйнувати репутацію дослідника, історія науки відзначена незліченними суперечками щодо пріоритету, який надається значущим результатам дослідження. Найпрестижніша та найвідоміша нагорода, яку сьогодні може отримати вчений — це Нобелівська премія.

Щороку вчені з низки різних дисциплін нагороджуються цією премією за видатні наукові досягнення. Оскільки уряди звернулися до нових інструментів державного управління для забезпечення більшої ефективності, виникла потреба в кількісних показниках ефективності та порівняльному аналізі. Премії (не лише Нобелівська) не можна проаналізувати кількісно, щоб дати оцінку більшості дослідників: вони досить рідкісні події для дослідників і часто присуджуються за досягнення, які лежать у далекому минулому. Тому в природничих науках стало звичним використовувати бібліометричні показники для вимірювання продуктивності. Особливо за останні кілька років бібліометрична оцінка окремих дослідників привернула особливу увагу. У 2005 році Хірш (Hirsch) представив індекс h , який дає інформацію про продуктивність вченого та вплив його чи її публікацій в одному числі (h – це кількість публікацій із принаймні h цитувань). Індекс h відносно швидко став дуже популярним. Однак, індекс h лише обмежено підходить для оцінки продуктивності дослідника. Бібліометричний аналіз результативності досліджень у природничих науках і науках про життя базується на двох фундаментальних припущеннях:

1) результати важливих досліджень публікуються в журнальних статтях. Тому кількість статей, які опублікував дослідник, дещо говорить про продуктивність його чи її досліджень.

2) Кожне нове дослідження має бути тісно пов'язане з поточними або минулими дослідженнями.

Ці близькі посилання позначені цитатами. Оскільки цитування відображають когнітивний вплив цитованої публікації на цитовану публікацію, цитування розглядаються як міра впливу, який публікація має на науку. Неважко знайти кількість публікацій і цитат, указаних для окремих вчених у доступні літературні бази даних. Оскільки обидві цифри (кількість публікацій і цитувань) пов'язані з науковою практикою, а дані легкодоступні, вони стали найважливішими інструментами для кількісної оцінки окремих дослідників. Сьогодні оцінювальні дослідження виходять за рамки простого визначення кількості публікацій і цитувань дослідника; також використовуються численні

					КР.КІ.8351890.00.00.000 ПЗ	Арк.
Змн.	Арк.	№ докум.	Підпис	Дата		18

бібліометричні індикатори, що дозволяє охопити багатовимірну природу наукових досягнень у її складності. Наприклад, пропонується використовувати вісім різних показників (таких як середня кількість публікацій на рік або загальна кількість цитувань). Кожен показник має певні переваги та може компенсувати недоліки іншого. Змістовна картина результативності дослідження виникає лише тоді, коли враховуються кілька показників. Однак слід враховувати, що багато показників, вибраних для дослідження, корелюють з кожним у високій мірі – навіть якщо результати відрізняються на детальному рівні. Таким чином, метрики, які використовуються в оціночному дослідженні, не повинні, наскільки це можливо, призводити до надлишкових результатів.

Хоча бібліометрія протягом багатьох років була окремою галуззю досліджень, все ще немає одноманітності в тому, як бібліометричний аналіз застосовується до окремих дослідників.).

1.3 Наукометричні бази даних та системи ідентифікації науковців

Для ефективного аналізу цитувань потрібні комплексні інструменти, які записують нещодавно створену наукову літературу, надану вченими та дослідниками з усього світу в усіх предметних областях. Комерційно доступні бази даних індексування цитувань стали дуже всебічними інструментами для аналізу цитувань, картографування науки, картографування міжнародних спільних досліджень і аналізу тенденцій у нових галузях науки. Індекс наукового цитування, запроваджений у 1964 році Інститутом наукової інформації (ISI), широко використовувався для аналізу цитування та вимірювання досліджень. На початку двадцять першого століття ми бачимо появу нових інструментів і нових методів вимірювання науки, наукових журналів, установ і окремих людей. Деякі з інструментів також є у вільному доступі для наукових спільнот, щоб допомогти їм зрозуміти експоненційне зростання наукової літератури. Вибух знань став неминучим для наукових

					КР.КІ.8351890.00.00.000 ПЗ	Арк.
Змн.	Арк.	№ докум.	Підпис	Дата		19

спільнот, оскільки бачимо, що деякі наукові онлайн-журнали навіть публікують більше тисячі статей на рік. Таким чином, всю літературу необхідно відстежувати та індексувати на благо суспільства. Інноваційні інструменти та показники тепер вимірюють вплив і вплив кожної рецензованої наукової публікації. Вимірювання тепер знову зосереджено на рівні статей, оскільки вимірювання рівня журналу за допомогою звичайних індикаторів на основі цитування обговорювалося як неадекватне, упереджене або спотворене.

Незважаючи на те, що ScimagoJR став законодавцем моди як комплексний інструмент для оцінки ефективності досліджень і для вимірювання впливу наукових комунікацій за допомогою методів аналізу цитування, до цього є багато інших. Ряд баз даних цитат було запущено різними організаціями. З появою високошвидкісних Інтернет-технологій і запуском багатьох електронних журналів було запущено низку бібліографічних баз даних із цитуванням для надання безперебійного доступу до записаних людських знань онлайн для наукових спільнот у всьому світі. На рисунку 1.3 наведено орієнтовний список наявних на даний момент баз даних цитувань, якими користуються мільйони дослідників у всьому світі.

Name of Citation Database	Launched	Scope	Owned by	Terms of Availability
<i>Science Citation Index (SCI)</i>	1964	Global	Thomson Reuter	Subscription-based with Web of Science
<i>Social Science Citation Index (SSCI)</i>	1972	Global	Thomson Reuter	Subscription-based with Web of Science
<i>Arts & Humanities Citation Index (A&HCI)</i>	1978	Global	Thomson Reuter	Subscription-based with Web of Science
<i>Scopus</i>	2004	Global	Elsevier B.V.	Subscription-based
Google Scholar Citations	2004	Global	Google Inc.	Freely Available Online
Microsoft Academic Search	2003	Global	Microsoft Research	Freely Available Online
CiteSeerX (CiteSeerX.ist.psu.edu)	1997	Global; Subject specific	Pennsylvania State University, USA	Freely Available Online
<i>Indian Citation Index (IndianCitationIndex.com)</i>	2009	India/ South Asia	Knowledge Foundation and Diva Enterprises	Subscription-based

Рисунок 1.3 – Найбільш популярні наукометричні бази даних

Ці бази даних цитат насамперед допомагають у бібліографічному пошуку опублікованої літератури. Бази даних цитувань індексують літературу, опубліковану в рецензованих академічних журналах та інших каналах академічної комунікації, як-от книги, матеріали конференцій, тези та дисертації.

У багатьох країнах деякі імена осіб із поєднанням імені та прізвища дуже поширені або подібні. Під час пошуку в онлайн-базах даних за певним іменем учасника результат пошуку може показати більше, ніж очіувалося, кількість бібліографічних записів статей, написаних різними особами зі схожим іменем. Таким чином, унікальний ідентифікатор автора або учасника наукової публікації дуже важливий для того, щоб відрізнити особу від групи осіб із подібним іменем. Унікальний ідентифікатор допомагає вирішити неясність автора. Зараз спільнотам дослідників широко доступні дві онлайн-системи для отримання унікального ідентифікатора автора чи учасника дослідження. Ці дві системи, а саме ResearcherID і Open Researcher and Contributor ID (ORCID), доступні на веб-сайтах ResearcherID.com і ORCID.org відповідно.

Ці дві системи сумісні та можуть мати пов'язані дані між собою. Ідентифікатори авторів на ResearcherID.com пов'язані з базою даних Web of Science (WoS). Профіль на цьому веб-сайті може отримати бібліографічні записи з WoSdatabase, що відповідають внеску автора. Ідентифікатори авторів на ORCID.org пов'язані з базою даних Scopus. Профіль на цьому веб-сайті може отримати бібліографічні записи з бази даних Scopus, які відповідають внеску автора. Коли ви надсилаєте свої рукописи до журналів WoS або Scopus для публікації, вони зазвичай приймають рукописи через систему онлайн-подання та рецензування. У процесі подання журнали зазвичай запитують будь-який з унікальних ідентифікаторів, ResearcherID або ORCID iD. Журнали, охоплені WoS, зазвичай віддають перевагу ResearcherID, тоді як журнали, охоплені Scopus, віддають перевагу ORCID iD. Деякі системи онлайн-подання рукописів приймають обидва ідентифікатори. База даних WoS доступна для пошуку за обома унікальними ідентифікаторами. Веб-сайти ResearcherID.com і ORCID.org стали цінними ресурсами для авторів та інших дослідників, а також установ.

					КР.КІ.8351890.00.00.000 ПЗ	Арк.
Змн.	Арк.	№ докум.	Підпис	Дата		21

Хоча автор може створити загальнодоступний профіль на цих двох веб-сайтах і отримати відповідно унікальний ідентифікатор, обидва веб-сайти підтримують онлайн-реєстр авторів, який можна шукати за іменем, ключовими словами, приналежністю та країною. Якщо для автора доступний загальнодоступний профіль, його список публікацій і повнотекстовий вміст цих публікацій (через зовнішні гіперпосилання) можуть переглядати інші дослідники, фінансові агенції та установи, які прагнуть до академічної співпраці.

ResearcherID.com, створений Thomson Reuters, є глобальним веб-реєстром авторів і дослідників. Дослідник може вільно створити унікальний ідентифікатор під назвою ResearcherID у цій системі, який є постійним за своєю природою та може бути доданий до баз даних видавців для унікальної ідентифікації його/її як учасника. Крім участі в реєстрі авторів, дослідник може створити публічний профіль і додати свій список публікацій із бази даних WoS або файлу RIS. Хоча список публікацій доступний у профілі дослідника, певні показники цитування, мережа цитування статей і мережа співпраці можуть бути видимі власнику профілю, а також профілям пошуку інших користувачів у цьому онлайн-реєстрі. Профіль ResearcherID допомагає відстежувати кількість цитувань, середнє цитування та h-індекс автора з бази даних WoS. Таким чином, цей веб-сайт стає дуже корисним інструментом для вимірювання рівня авторства, якщо дослідник має велику кількість робіт, які проіндексовані в базі даних WoS.

ORCID (Open Researcher and Contributor ID) Подібно до ResearcherID.com, ORCID.org надає постійний цифровий ідентифікатор, який відрізняє дослідника від усіх інших дослідників. Створення ORCID iD для дослідника дуже просте та безкоштовне. Тут ви повинні надати певні особисті та професійні дані, щоб включити своє ім'я в реєстр унікальних ідентифікаторів дослідника. Після успішної реєстрації на веб-сайті створюється унікальний ідентифікатор ORCID і створюється профіль користувача. Ви можете інтегрувати свої інші профілі або унікальні ідентифікатори авторів, доступні в інших місцях, наприклад ResearcherID, Scopus і LinkedIn. Ваш список публікацій також буде додано до вашого профілю, який містить бібліографічну інформацію опублікованих

					КР.КІ.8351890.00.00.000 ПЗ	Арк.
						22
Змн.	Арк.	№ докум.	Підпис	Дата		

наукових робіт і гіперпосилання на повний текст кожної роботи. Список публікацій можна отримати з бази даних Scopus, яка включатиме бібліографічний запис статей, опублікованих у журналах, що охоплюються Scopus. Інші відповідні роботи також можна додати у свій профіль шляхом імпорту бібліографічних даних із файлу RIS вашого списку публікацій. ORCID.org підтримує пошуковий реєстр дослідників, який допомагає ідентифікувати дослідників у вашій галузі, з установи, співавтора, міста чи країни. Фінансуючі агенції також можуть відстежувати роботи дослідників, які вони фінансують або розглядають фінансування в найближчому майбутньому. Веб-сайт надає API, які підтримують міжсистемний зв'язок і автентифікацію для онлайн-систем спонсорів, видавців та інших, яким потрібні ідентифікатори ORCID.

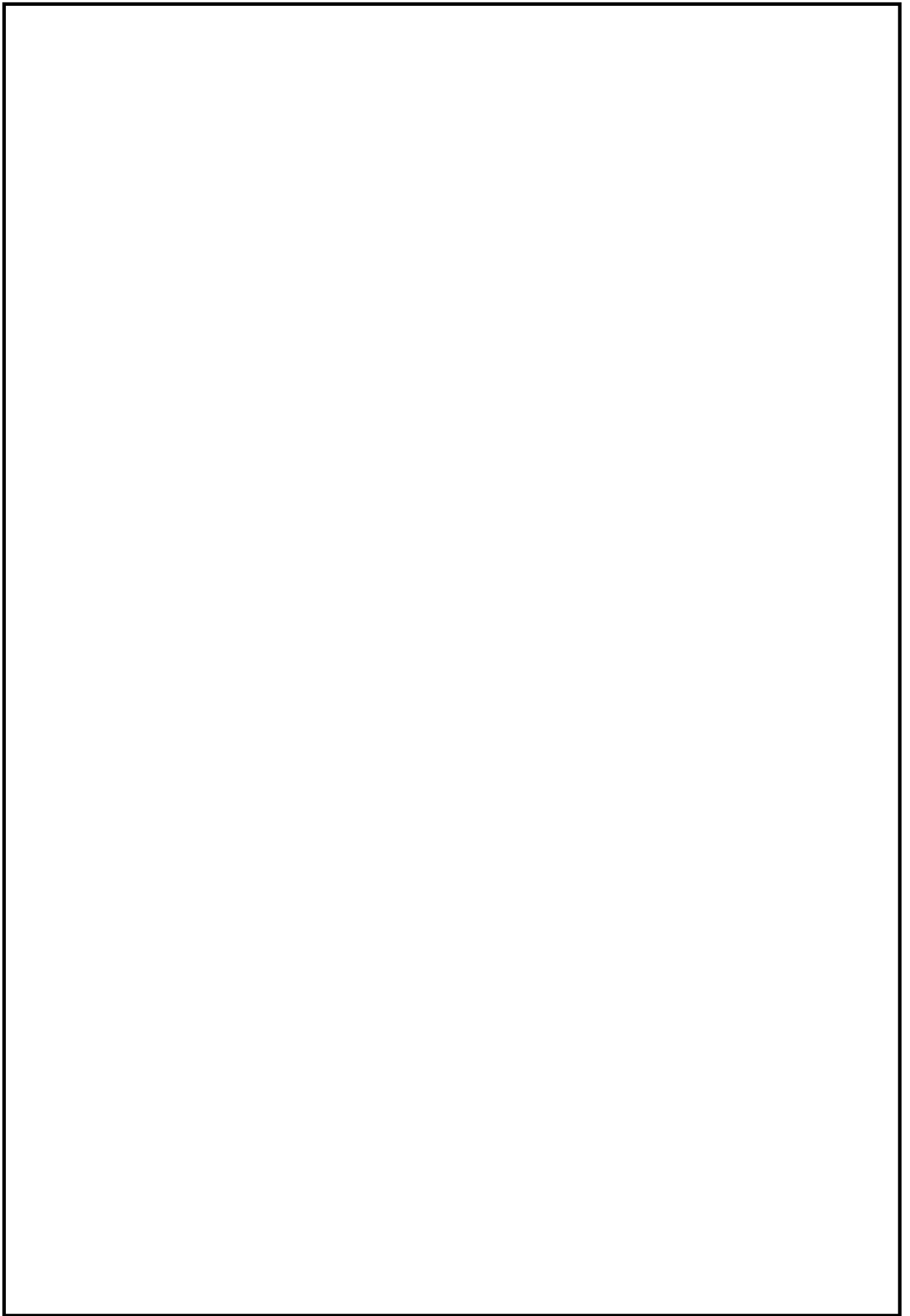
1.4 Висновки до розділу та постановка завдань кваліфікаційної роботи

В розділі було проведено огляд та класифікацію технологій оцінки активності наукового працівника на основі різних критеріїв оцінювання, проведено аналізу оцінки активності науковця на основі цитованості його публікацій, а також проаналізувано сучасні наукометричні бази даних.

Для досягнення поставленої мети необхідно вирішити наступні завдання:

- 1) дослідити методології та технології оцінювання наукових працівників;
- 2) провести дослідження наукометричних баз даних;
- 3) проаналізувати особливості роботи з API Web of Science;
- 4) проаналізувати особливості роботи з API Google Scholar;
- 5) розробити алгоритм формування рейтингу викладачів на основі показників Google Scholar та Web of Science.
- 6) провести тестування спроектованої структури програмного модуля формування рейтингу викладачів на основі показників Google Scholar та Web of Science.

					КР.КІ.8351890.00.00.000 ПЗ	Арк.
Змн.	Арк.	№ докум.	Підпис	Дата		23



					КР.КІ.8351890.00.00.000 ПЗ	Арк.
Змн.	Арк.	№ докум.	Підпис	Дата		24

2 МЕТОДИ ТА АЛГОРИТМИ ОЦІНЮВАННЯ АКТИВНОСТІ НАУКОВЦЯ НА ОСНОВІ ПУБЛІКАЦІЙ

2.1 Наукометрична база Web of Science

Інститут наукової інформації (ISI) створив високоповажні бази даних цитувань у всіх основних предметних галузях, а саме. SCI, SSCI та A&HCI. ISI випускав їх у друкованому форматі, а також у форматі CD-ROM через регулярні проміжки часу як періодичні видання. У друкованому форматі для кожного тому SCI або SSCI існував індекс із трьох частин, а саме: Індекс джерела, Індекс теми та Індекс цитування. У Показчику джерел подано повну бібліографічну інформацію про автора цитування та цитовану роботу. У Індексі цитування роботи, цитовані протягом певного року, перераховуються в алфавітному порядку за іменем цитованого автора, після якого йдуть імена цитованих авторів. У Предметному показчику, який називається Permuterm Subject Index, перераховуються важливі слова в назві статті. У 1992 році компанія ISI була придбана компанією Thomson Scientific & Healthcare і стала новою організацією під назвою Thomson ISI. Пізніше Thomson ISI став частиною підрозділу Healthcare & Science Thomson Reuters. Thomson ISI представила новий веб-продукт під назвою Web of Science (WoS), який пропонував послуги індексування та реферування (I&A) для світових спільнот дослідників. WoS включив усі свої бази даних цитат, а саме SCI, SSCI та A&HCI, щоб зробити WoSmore всеосяжним і конкурентоспроможним порівняно з іншими подібними продуктами в цьому сегменті «Інновації в вимірюванні науки та наукових досліджень». Усі його бази даних цитат SCI, SSCI та A&HCI стали доступними через єдиний шлюз. Для WoS була представлена розширена версія SCI під назвою Science Citation IndexExpanded, щоб охопити більше рецензованих академічних журналів з усього світу. Окрім баз даних цитування SCI, SSCI та A&HCI, WoS також включає Book Citation Index та Conference Proceedings Citation Index. Однак ще кілька баз даних цитат не є частиною WoS, але включені в інший комплексний продукт Thomson

					КР.КІ.8351890.00.00.000 ПЗ	Арк.
Змн.	Арк.	№ докум.	Підпис	Дата		25

Reuters, а саме Web of Knowledge (WoK). Також можна помітити, що з впровадженням WoK WoS став підмножиною комплексного продукту WoK. Нещодавно WoS розширив охоплення журналів, опублікованих у різних регіонах світу, включаючи Латинську Америку, Азію та Африку. Він також розширив охоплення рецензованих журналів відкритого доступу, щоб зробити цей продукт більш конкурентоспроможним.

Для автотизованої взаємодії за базою даних використовується внутрішня API. Даний набір функцій дозволяє організовувати запити та отримувати відповіді згідно запропонованого шаблону, що дозволяє швидко та з мінімальними часовими затратами проектувати та реалізовувати програмні модулі взаємодії з даною наукометричною базою даних.

Проаналізуємо основні етапи та можливості даного API. Web of Science API підтримує розширений пошук у Web of Science на основі вашої підписки для отримання метаданих на рівні елемента, включно з кількістю цитувань і адресами/філіями учасників. Додаткові операції підтримують можливість виявлення пов'язаних записів, а також цитованих посилань і елементів цитування. API підтримує відповіді JSON і XML, і ця документація підтримує спробу обох типів відповідей; однак моделі та приклади значень наразі дійсні лише для JSON. Регулювання та обмеження швидкості застосовуються відповідно до угоди з Clarivate Analytics, а квоти буде вказано в заголовку відповіді (наприклад, `x-rec-ampmonth-remaining: '15467'` або `x-req-reqpersec-remaining: '3'`).

Для формування запиту доступні декілька варіантів, які передаються за методом GET. Перелік можливих параметрів наведено у таблиці 2.1. У випадку успішного виконання запиту клієнт отримає повну інформацію згідно обраних параметрів та зможе провести парсинг відповіді. У результаті виконання запиту користувач може отримати різні коди відповіді, серед можливих варіантів:

- код 200 – Успіх, доступність полів у відповіді залежатиме від бази даних, у якій здійснюється пошук;
- код 400 – Поганий запит;
- код 404 – Не знайдено;

					КР.КІ.8351890.00.00.000 ПЗ	Арк.
Змн.	Арк.	№ докум.	Підпис	Дата		26

– код 500 – Внутрішня помилка сервера;

Таблиця 2.1 – Параметри запиту до бази даних Web of Science

Назва параметра	Опис параметра
databaseId * string (query)	База даних для пошуку. WOK представляє всі бази даних. Доступні значення : WOS, ARCI, BCI, BIOABS, BIOSIS, CABI, CCC, CSCD, DCI, DIIDW, FSTA, INSPEC, KJD, MEDLINE, PPRN, SCIELO, WOK, ZOOREC
lang string (query)	Мова пошуку. Цей елемент може приймати лише одне значення: en для англійської мови. Якщо мова не вказана, за замовчуванням передається англійська.
uniqueId * string (path)	Ідентифікатор основного елемента(iv), який потрібно шукати, наприклад: WOS:000270372400005. Не може бути нульовим або порожнім рядком. Кілька значень розділяються комою
count * integer	Кількість записів для повернення має бути 0-100
firstRecord * integer	Конкретний запис, якщо такий є в наборі результатів для повернення. Не може бути менше 1 і більше 100 000.
sortField string (query)	Впорядкувати за полями. Ім'я поля та порядок за пунктами, розділені «+» або « », використовуйте A для ASC і D для DESC, наприклад: PY+D. Кілька значень розділяються комою.
viewField string (query)	Параметр для обмеження записів вибраними полями. Оскільки багато полів однозначно ідентифікуються комбінацією елемента й атрибута, можливо, ви не зможете точно обмежити вміст запису. Наприклад, параметр viewField 'titles' завжди повертатиме заголовки документів, назви публікацій (як повні, так і скорочені) і назви книг.
optionOther string (query)	Інший варіант(и) представлення результату. HL і On/Off розділені знаком «+» або « », де HL=HighLight: напр. HL+Uвімк.
links string (query)	Прапор для повернення шлюзових посилань (Джерело та Цитування) для кожного запису, повернутого у відповідь. За замовчуванням false

Арк.

КР.КІ.8351890.00.00.000 ПЗ

27

Змн. Арк. № докум. Підпис Дата

У разі успішного виконання запиту, у відповіді буде побудована за схемою, на наведена на рисунку 2.1.

```

ReferenceResponse {
  Data
    [ReferenceRecord {
      UID string
      citedAuthor string
      timesCited integer
      year integer
      page integer
      citedWork string
      citedTitle string
      doi string
    }]
  QueryResult {
    QueryID string
    RecordsSearched integer
    RecordsFound integer
  }
}

```

Рисунок 2.1 – Схема відповіді від сервера у разі успішного виконання запиту

Використовуйте пошук дослідників, щоб шукати записи про авторів і профілі дослідників у Web of Science Core Collection. Пошук дослідників допомагає ідентифікувати статті, авторами яких, ймовірно, є одна особа, і переглядати профілі цих авторів, які містять показники та іншу наукову діяльність (наприклад, рецензування). Будь-який автор, чия стаття проіндексована в Web of Science Core Collection, має авторський запис.

Пошук може шукати автора за іменем або ідентифікатором автора (Web of Science ResearcherID або ORCID). Щоб знайти автора за іменем, введіть прізвище та ім'я або ініціали. У поле «Прізвище» можна додати дефіс, пробіл або апостроф.

Результати пошуку дослідника:

Записи автора, які були заявлені автором і є профілем дослідника Web of Science, позначаються в результатах пошуку зеленою галочкою.

Уточніть результати пошуку за іменем автора, організацією та тематичною категорією, щоб отримати більш відповідні результати.

Користувач може почати пошук нового автора безпосередньо створивши новий запит.

2.2 Наукометрична база Google Scholar

Google Scholar, запущений у 2004 році компанією Google Inc., здається, є найбільшою у світі базою даних для індексування та цитування рецензованої наукової літератури, яка охоплює більше академічних журналів та інших наукових матеріалів, ніж аналогічні інші бази даних цитат, такі як Scopus, Web of Science і Microsoft Академічний пошук. Він став найбільшою у світі системою пошуку наукової літератури. Це простий спосіб широкого пошуку наукової літератури з багатьох дисциплін і джерел. Його джерела включають статті, тези, книги, реферати та судові висновки від академічних видавництв, професійних товариств, онлайн-сховищ, університетів, предметних шлюзів та інших веб-сайтів. Google Scholar вказує на доступність наступних функцій для кінцевих користувачів:

- пошук у всій науковій літературі з одного зручного місця;
- вивчайте пов'язані роботи, цитати, авторів і публікації;
- знайдіть весь документ у своїй бібліотеці або в Інтернеті;
- бути в курсі останніх подій у будь-якій галузі досліджень;

Перевірте, хто цитує ваші публікації, створіть публічний профіль автора.

GSC – це персоналізоване джерело інформації для авторів, щоб відстежувати цитування своїх опублікованих статей. Як автор ви можете перевірити, хто цитує ваші публікації, побудувати графік цитувань у часі та обчислити кілька показників цитування. Ви також можете зробити свій профіль загальнодоступним, щоб він міг з'являтися в результатах пошуку Google Scholar, коли люди шукатимуть ваше ім'я, наприклад, Річард Фейнман. У загальнодоступному профілі відображена інформація включає: ім'я вченого, поточну приналежність, широкі сфери дослідницьких інтересів, бібліографічні відомості про всі статті, кількість цитувань, отриманих кожною статтею, імена співавторів, кількість підписників цього профілю тощо. Цей профіль також

					КР.КІ.8351890.00.00.000 ПЗ	Арк.
Змн.	Арк.	№ докум.	Підпис	Дата		29

містить кілька показників ефективності та показників цитування, наприклад як загальна кількість цитувань, h-індекс, індекс i-10; і загальна кількість цитувань за останні п'ять років, як показано на малюнку 17. Коли ви реєструєтесь в GSC як автор, ці показники цитувань обчислюються та оновлюються автоматично, коли Google Scholar знаходить нові цитати вашої роботи в Інтернеті. Ви можете вибрати автоматичне оновлення списку статей або переглядати оновлення самостійно, або оновлювати статті вручну в будь-який час.

По аналогії з Web of Science наукометрична база Google Scholar використовує власний набір шаблонів для проведення пошуку та аналізу інформації про авторів, їхні публікації, цитованість тощо. Якщо розглядати запити, то їх можна здійснювати різними підходами, проте завжди необхідно дотримуватись певного стандарту. Запити можна відправляти як методом GET так і виконувати їх за допомогою засобів різних мов програмування.

Google Scholar Author API надає змогу отримувати авторські результати зі сторінки пошуку Google Scholar Author. Під час запитів до даної сторінки необхідно дотримуватись ряду вимог, а саме використовувати параметри які буде розуміти серверна частина Google Scholar Author API. В запит необхідно включати як обов'язкові параметри так і додаткові. Перелік обов'язкових параметрів наведено в таблиці 2.2.

Таблиця 2.2 – Перелік основних параметрів Google Scholar Author API

Група параметрів	Назва параметра	Наявність в запиті	Опис
Загальні параметри	engine	Обов'язковий	База API. Для цього запиту потрібно встановити значення google_scholar_author
Параметри запиту	api_key	Обов'язковий	Ваш ключ AP
Параметри запиту	device	При потребі	Пристрій, використаний для вашого пошуку. Можна встановити значення desktop, mobile або tablet.
Пошуковий	author_id	Обов'язковий	Представляє

запит			ідентифікатор автора.
-------	--	--	-----------------------

Окрім основних параметрів в Google Scholar Author API передбачено ще ряд додаткових які дозволяють отримати розширені відповіді. Перелік додаткових полів для розширення можливостей запиту наведено в таблиці 2.3.

Таблиця 2.3 – Перелік полів для формування розширеного запиту

Група параметрів	Назва параметра	Наявність в запиті	Опис
Розширені параметри	view_or	При потребі	Використовується для перегляду окремих частин сторінки. Його можна встановити на view_citationабо list_colleagues
Розширені параметри	sort	При потребі	Сортує результати. Можна налаштувати title сортування за назвою або pubdate за датою публікації
Розширені параметри	citation_id	При потребі	Він використовується для отримання певної цитати. Це потрібно, коли view_or встановлено view_citation
Відображення	start	При потребі	Зміщення результатів пошуку Google. Позначає кількість результатів, які потрібно пропустити.
Відображення	num	При потребі	Кількість результатів, отриманих на кожній сторінці.
Локалізація	hl	При потребі	Мова, яку ви хочете використовувати для пошуку Google.

На основі висвітлених параметрів користувачі можуть створювати та відправляти запити. Приклад GET наведено нижче:

`https://serp.shifter.io/v1?engine=google_scholar_author&api_key=YOUR_API_KEY&author_id=LSsXyncAAAAJ`

В результаті виконання подібного запиту можна отримати відповідь в вигляді JSON відповіді яка матиме наступну структуру (рисунок 2.2), а інформації після додаткового аналізу може бути використана для подальшої статистичної обробки.

```
{
  "search_metadata": {
    "id": "60802ce5de983402dbelae4c",
    "status": "Success",
    "json_endpoint": "https://serpapi.com/searches/79beba2444307aad/60802ce5de",
    "created_at": "2021-04-21 13:47:17 UTC",
    "processed_at": "2021-04-21 13:47:18 UTC",
    "google_scholar_author_url": "https://scholar.google.com/citations?user=LS",
    "raw_html_file": "https://serpapi.com/searches/79beba2444307aad/60802ce5de",
    "total_time_taken": 1.26
  },
  "search_parameters": {
    "engine": "google_scholar_author",
    "author_id": "LSsXyncAAAAJ",
    "hl": "en"
  },
  "author": {
    "name": "Cliff Meyer",
    "affiliations": "Dana-Farber Cancer Institute and Harvard T.H. Chan School",
    "email": "Adresse email validée de jimmy.harvard.edu",
    "interests": [
```

Рисунок 2.2 – Приклад відповіді на запит до Google Scholar Author

Для створення та відправки запитів необхідно отримати спеціальний ключ, який можна взяти на сервісі Serpapi. Такі пошукові системи, як Google і Bing, постійно вдосконалюють своїх ботів і алгоритмів пошукової оптимізації (SEO). Тому дуже важливо використовувати API SERP для видалення даних, щоб залишатися актуальними для результатів веб-сторінки.

SERP посилається на сторінки, які повертає пошукова система після введення ключового слова на панелі пошуку. API стосуються програмних додатків, які допомагають отримувати результати пошуку з пошукової системи в режимі реального часу, працюючи на певній мові програмування, щоб надсилати запити та відповіді в певному форматі. Ці API дозволяють

аналізувати, переглядати, відстежувати та покращувати видимість вашого веб-сайту в пошуковій системі.

На основі запиту кінцевої точки API Google SERP дозволяють отримувати інформацію з SERP, таку як звичайні результати, платні результати, рекомендовані фрагменти та географічне розташування. Ви також можете використовувати API Google SERP, щоб за частки секунди перевірити статистику щодо кількості пошуків, пошукових архівів або місць розташування, які підтримує API.

2.3 Алгоритм обчислення рейтингу наукового працівника

Проаналізувавши існуючі критерії проведення оцінки наукових та науково-педагогічних працівників можна зробити висновки, що на сьогоднішній день не існує єдиного критерію для отримання цілісної оцінки. Наукові сфери розвиваються не однаково, популярність теж є різною як і кількість науковців які займають дослідженнями у певному напрямку. Проте оцінка активності науковців та наукових груп є важливим критерієм, оскільки це є чи не одним критерієм для об'єктивного розподілу ресурсів які необхідні для проведення досліджень. За основу запропонованого алгоритму було запропоновано взяти кількість публікацій які були зареєстровані в наукометричних базах Google Scholar та Web of Science. Дані наукометричні бази є популярними та відслідковують більшість публікацій які з'являються у світі, тому їхні дані можуть бути максимально об'єктивними. Для отримання більш нормалізованого коефіцієнта проводити аналіз будемо протягом останніх 5 років та знаходити середнє значення між двома значеннями. Середнє значення дозволить прослідкувати динаміку активності науковця та рівень охоплення аудиторії яка ознайомена з науковими працями за рахунок кореляції між показниками двох наукометричних баз.

Запропонований алгоритм містить наступні кроки:

					КР.КІ.8351890.00.00.000 ПЗ	Арк.
Змн.	Арк.	№ докум.	Підпис	Дата		33

- 1) Ввід імені або науковця.
- 2) Пошук інформації про заданого науковця в наукометричній базі Google Scholar. Якщо інформація відсутня переходимо на крок 8.
- 3) Пошук інформації про заданого науковця в наукометричній базі Web of Science. Якщо інформація відсутня переходимо на крок 8.
- 4) За формулою знаходимо п'ять коефіцієнтів активності наукового працівника КАНПі:

$$КАНПі = \frac{|Gi + Wi|}{2}.$$

- 5) Знаходимо максимальне та мінімальне значення.
- 6) Визначаємо рівень зміни коефіцієнту КАНП, якщо різниця пари сусідніх значень додатне число, то науковень активно розвивається, якщо ні то спостерігається регрес в його роботі.
- 7) Виводимо інформацію про отриманий висновок про рівень активності науковця та переходмо на крок 9.
- 8) Повідомляємо, що з деяких причин не вдалось отримати інформацію з наукометричних баз даних, а отже неможливо провести аналіз активності науковця.
- 9) Завершення роботи алгоритму.

Перевагами даного алгоритму слід вважати:

- можливість отримання об'єктивної оцінки результатів роботи наукового працівника на основі аналізу його публікацій;
- отримання інформації про кількість виданих публікацій можна отримати і з інших НБД, що дозволить збільшити об'єктивність отриманих даних;
- оцінка проводиться в автоматичному режимі без втручання людського фактору, тому є більш об'єктивною;
- алгоритм відрізняється високою швидкістю та може бути реалізований як самостійно так і в межах системи оцінки роботи наукових працівників.

					КР.КІ.8351890.00.00.000 ПЗ	Арк.
						34
Змн.	Арк.	№ докум.	Підпис	Дата		

Серед недоліків є:

- необхідність використовувати спеціалізоване АРІ для доступу до відповідних НБД;
- залежність від зовнішніх НБД;
- кількість запитів може бути обмежена серверною частиною.

З блок-схемою розробленого алгоритму можна більш детально ознайомитись на рисунку 2.3.

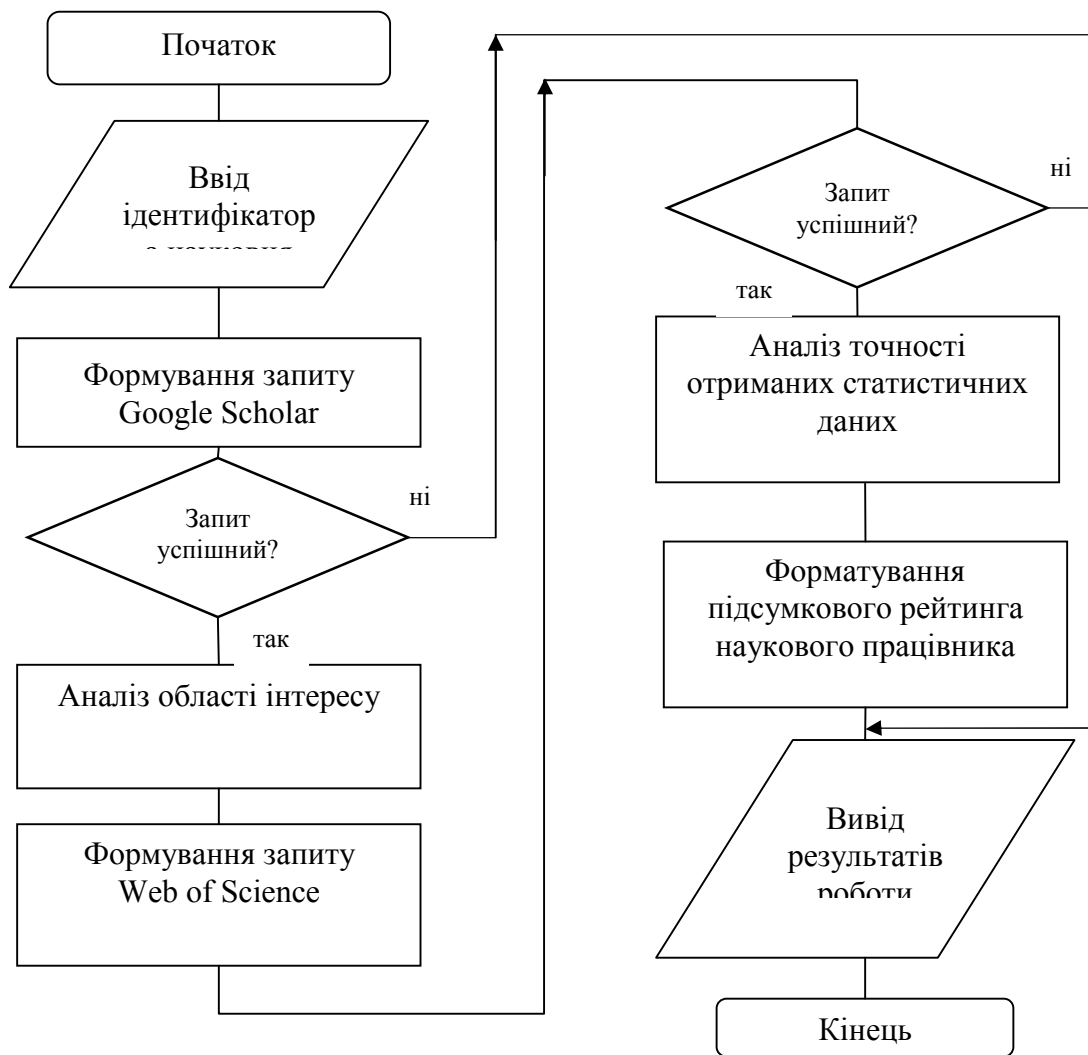


Рисунок 2.3 – Блок-схема алгоритму оцінки наукового працівника на основі аналізу його публікацій

Змн.	Арк.	№ докум.	Підпис	Дата

3 ПРОГРАМНИЙ МОДУЛЬ ВИЗНАЧЕННЯ РЕЙТИНГУ НАУКОВЦЯ НА ОСНОВІ ПУБЛІКАЦІЙ

3.1 Структура, моделювання та функціональні можливості модуля

Для реалізації програмного модуля було обрано мову програмування python. Серед переваг даної мови можна виділити легкість написання коду, велика кількість готових пакетів які легко можна інтегрувати будь-яке програмне рішення з подальшою оптимізацією в залежності від поставлених завдань. Окрім того розроблене API наукометричних баз даних дозволяє створювати та опрацьовувати запити на цій мові. Тому вибір саме цієї мови є очевидним. Для проведення додаткових операцій над отриманими статистичними даним було використано пакет Pandas. Pandas – це пакет Python, який забезпечує швидкі, гнучкі та виразні структури даних, розроблені для того, щоб зробити роботу з «реляційними» або «міченими» даними одночасно легкою та інтуїтивно зрозумілою. Він має на меті стати основним будівельним блоком високого рівня для практичного аналізу реальних даних у Python. Крім того, він має ширшу мету – стати найпотужнішим і найгнучкішим інструментом для аналізу та маніпулювання даними з відкритим кодом, доступним будь-якою мовою.

Pandas добре підходить для багатьох різних типів даних:

Табличні дані з неоднорідно типізованими стовпцями, як у таблиці SQL або електронній таблиці Excel

Упорядковані та неупорядковані (не обов'язково з фіксованою частотою) дані часових рядів.

Довільні матричні дані (однорідно типізовані або неоднорідні) з мітками рядків і стовпців

Будь-яка інша форма наборів спостережень/статистичних даних. Для розміщення даних у структурі даних pandas зовсім не потрібно маркувати дані

Дві первинні структури даних pandas, Series(1-вимірні) і DataFrame(2-вимірні), обробляють переважну більшість типових випадків використання у

					КР.КІ.8351890.00.00.000 ПЗ	Арк.
Змн.	Арк.	№ докум.	Підпис	Дата		36

фінансах, статистиці, соціальних науках і багатьох галузях техніки. Для користувачів R DataFrame надає все, що надає R, data.frame і багато іншого. pandas побудований на основі NumPy і призначений для інтеграції в наукове обчислювальне середовище з багатьма іншими бібліотеками сторонніх розробників.

Функціональні можливості пакету Panda:

Легка обробка відсутніх даних (представлених як NaN) як у даних із плаваючою комою, так і без неї

Змінність розміру: стовпці можна вставляти та видаляти з DataFrame та об'єктів більшої розмірності

Автоматичне та явне вирівнювання даних : об'єкти можуть бути явно вирівняні за набором міток, або користувач може просто ігнорувати мітки та дозволити Series, DataFrame, тощо автоматично вирівнювати дані для вас під час обчислень.

Потужне, гнучке групування за функціями для виконання операцій розділення, застосування та об'єднання над наборами даних, як для агрегування, так і для перетворення даних.

Спростіть перетворення нерівних, по-різному індексованих даних в інших структурах даних Python і NumPy в об'єкти DataFrame.

Інтелектуальне нарізання на основі міток, фантастичне індексування та піднабір великих наборів даних.

Інтуїтивно зрозуміле об'єднання та об'єднання наборів даних.

Гнучка зміна форми та поворот наборів даних.

Ієрархічне маркування осей (можливість мати кілька міток на один крок).

Надійні інструменти вводу-виводу для завантаження даних із плоских файлів (CSV і файлів із роздільниками), файлів Excel, баз даних і збереження/завантаження даних із надшвидкого формату HDF5.

Спеціальна функціональність часових рядів: генерація діапазону дат і перетворення частоти, статистика рухомого вікна, зсув дати та відставання.

Багато з цих принципів спрямовані на усунення недоліків, які часто виникають при використанні інших мов/середовищ наукових досліджень. Для

					КР.КІ.8351890.00.00.000 ПЗ	Арк.
						37
Змн.	Арк.	№ докум.	Підпис	Дата		

дослідників даних робота з даними зазвичай поділяється на кілька етапів: аналіз і очищення даних, їх аналіз/моделювання, а потім упорядкування результатів аналізу у формі, придатній для побудови або табличного відображення. pandas є ідеальним інструментом для всіх цих завдань.

Для уникнення проблем на етапі програмної реалізації модуля оцінювання рейтингу наукових працівників доцільно спроектувати структуру програмного модуля та провести її моделювання. При проектуванні структури за основу було обрано розподілену модель побудов програмних додатків, при якій функції будуть групуватись в окремі логічні блоки які в свою чергу будуть логічно незалежними один від одного. Для об'єднання даних блоків в структуру програмного додатку внесено блок керування. Даний блок дозволить організувати інтерфейси між окремими частинами та проводити перекодування у форма який необхідний для коректного функціонування алгоритмів. Такий підхід має ряд переваг:

- Можливість реалізації та тестування окремих блоків незалежно один від одного;
- Можливість заміни окремого блока на вже готові патеки для зменшення часових затрат на розробку програмного модуля;
- Модифікація та нарощування функціональних можливостей програмного модуля проводиться просто та залежить тільки від поставлених завдань;
- При відсутності реалізації окремого модуля на етапі тестування його можна замінити набором готових тестових вибірок.

Недоліком даного підходу є необхідність узгодження структур даних для обміну інформацією між різними елементами програмного модуля.

Окрім очевидних переваг блокової архітектури програмного додатку при проектуванні структури були враховані висновки які були зроблені під час аналізу програм аналогів. Тому що, користувачі не завжди будуть використовувати усі функціональні можливості програмного додатку, наприклад необхідна інформація тільки по одній наукометричній базі, отже функції для роботи з іншою на жуть бути якщо не видалені то деактивовані для зменшення

					КР.КІ.8351890.00.00.000 ПЗ	Арк.
						38
Змн.	Арк.	№ докум.	Підпис	Дата		

кількості використовуваних програмою. На рисунку 3.1 проілюстровано спрощену схему внутрішніх взаємозв'язків між окремими структурними одиницями програмного модуля оцінювання рейтингу наукових працівників.

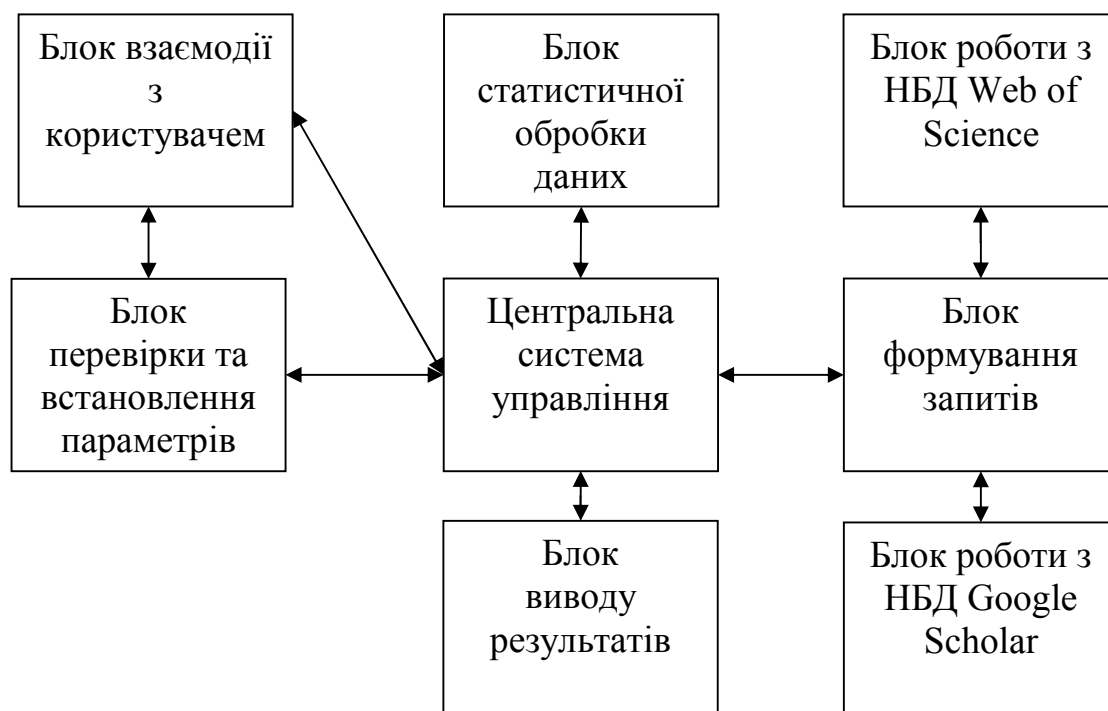


Рисунок 3.1 – Структурна схема програмного модуля формування рейтингу науковців на основі аналізу їх публікацій

Рисунок 3.1 дозволяє оцінити загальну структуру та рівень взаємозв'язку між окремими структурними одиницями програмного модуля. Розглянемо більш детально окремі структурні вузли програмного модуля:

– “ Блок перевірки та встановлення параметрів ” – для коректного функціонування програмного модуля необхідно встановлення ряду параметрів по замовчуванню. Саме на основі них буде проводитись створення запитів та підключення до мережі інтернета та наукометричних ба даних. Дані параметри можуть змінюватись так само як і формати запитів, тому функції даного блоку дозволяють корегувати та модифікувати значення параметрів роботи програмного модуля для надання можливості повноцінного функціонування.

– “ Блок взаємодії з користувачем ” – набір функцій які дозволяють реалізувати зручний для користувача процес управління програмним модулем

та зручність при перегляді отриманих результатів. Серед даних які необхідні програмному модулю для коректного функціонування є особиста інформація про науковця рейтинг якого необхідно дослідити, а саме його ім'я та прізвище або унікальний ідентифікатор в системах НБД Google Scholar та Web of Science. Окрім того користувач повинен обрати який саме тип відображення буде використовуватись під час виводу результатів роботи програми (текстовий звіт чи графік). Серед допоміжних можливостей даний модуль містить систему мікроповідомлень яка була розроблена з метою допомоги користувачів невеликим досвідом.

– “ Блок статистичної обробки даних ” – даний блок на рівні з блоком обміну інформацією з наукометричними базами даних є одним з головним в програмному модулі. Функції даного функціонального вузла проводять аналіз отриманої на основі запитів інформації, проводять первинний аналіз та дозволяють обчислити коефіцієнти активності науковця протягом п'яти років. Окрім того в даному модулі зібрані всі структури даних для зберігання поточних та обчислених даних про обраного науковця.

– “ Блок формування запитів ” – для коректної роботи з різними наукометричними базами даних реалізовано ряд допоміжних функцій, що надають можливість автоматичного формування запитів до відповідних баз даних. В даному блоці проводиться перевірка правильності заповнення полів запиту на корекцію формування самих запитів.

– “Блок роботи з НБД Web of Science” – допоміжний блок для перевірки можливостей відправки запитів до наукометричної бази Web of Science, а також аналіз відповідей отриманих від серверів. При неуспішній відповіді від сервера, даний блок зможе повторно відправити запит та повідомити центральну систему керування модулем про виникнення критичної ситуації.

– “Блок роботи з НБД Google Scholar” – допоміжний блок для перевірки можливостей відправки запитів до наукометричної бази Google Scholar, а також аналіз відповідей отриманих від серверів. При неуспішній відповіді від сервера, даний блок зможе повторно відправити запит та повідомити центральну систему керування модулем про виникнення критичної ситуації.

					КР.КІ.8351890.00.00.000 ПЗ	Арк.
						40
Змн.	Арк.	№ докум.	Підпис	Дата		

– “ Центральна система управління ” – призначення для організації та підтримки роботи програмного модуля, перевірки коректності обміну даними між окремими структурними одиницями, тестування програмного модуля на наявність некоректно працюючих частин або помилок які виникають під час експлуатації програмного модуля. При виявленні помилкової ситуації програмний блок проводить спробу автоматичного виправлення ситуації, а якщо ситуація не підлягає корекції повідомляє про це користувача..

– “ Блок виводу результатів ” – формат виводу інформації є одним з важливих елементів взаємодії користувача з програмним модулем. Оскільки, зручність читання, доступність викладення та формат звіту є основним для повторного використання програмного модуля та можливості його впровадження в більшій системі.

Запропонована структура дозволяє чітко розділити набори функцій які будуть реалізовані під час програмного кодування модуля. Такий підхід дозволяє проводити модифікацію чи заміну окремих функцій без втрати цілісності програмної розробки. Функції можуть замінювати одна одну та забезпечувати стабільну роботу програмного застосунку навіть при виникненні непередбачуваних критичних ситуаціях.

Для додаткової оцінки запропонованої структури програмного додатку було проведено моделювання. Метою моделювання було отримати дані про можливість взаємодії програмного модуля та користувачів під час експлуатації програмного додатку. Для моделювання було обрано моделі які входять до набору універсальних мов моделювання UML. Даний пакет дозволяє проводити дослідження структуру програмних додатків, їх інтерфейсів, проаналізувати дії програмних розробок під час виникнення критичних збоїв тощо. Для проведення дослідження були обрані наступні параметри роботи майбутнього програмного модуля, а саме – можливості доступу користувачів до окремих функціональних блоків. Дане моделювання дозволить дослідити чи користувачі зможуть за допомогою запропонованого програмного модуля виконати поставлені пред ними завдання та оцінити рівень активності наукового

					КР.КІ.8351890.00.00.000 ПЗ	Арк.
						41
Змн.	Арк.	№ докум.	Підпис	Дата		

працівника на основі аналізу його публікацій. Результати моделювання наведені на рисунку 3.2.

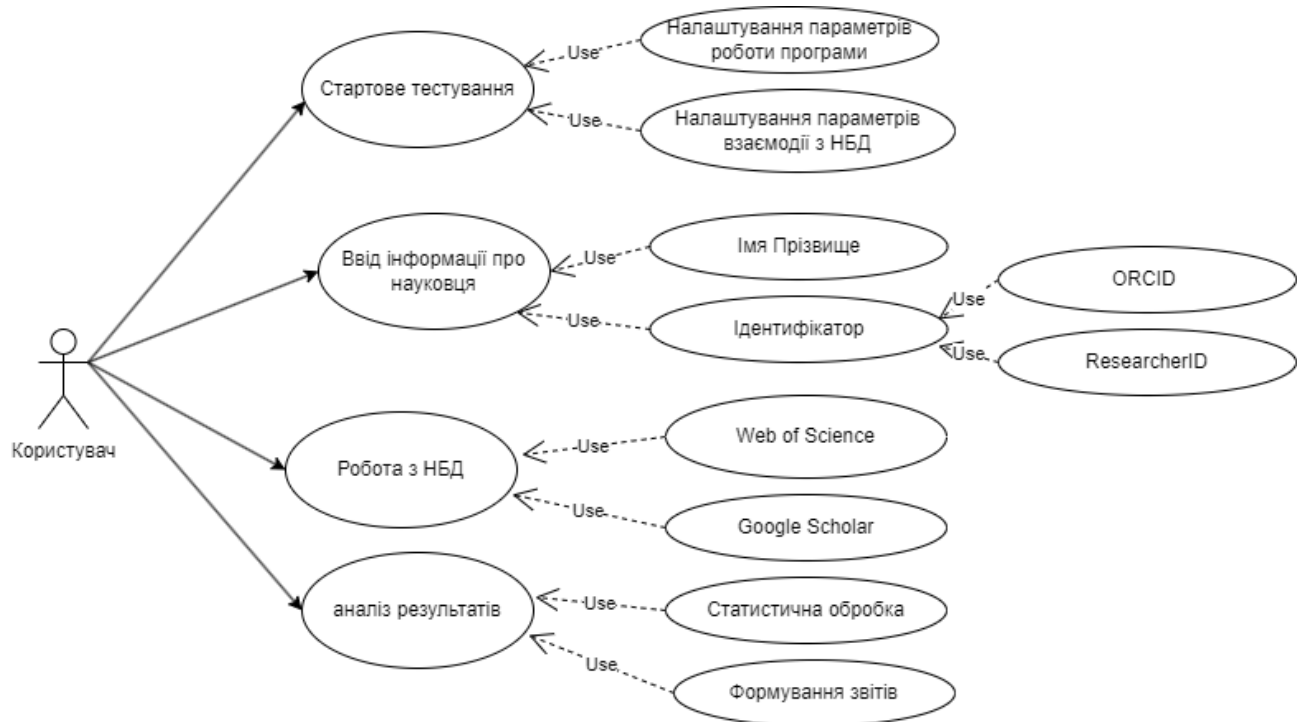


Рисунок 3.2 – Діаграма прецедентів для ілюстрації доступу користувачів до функціональних блоків програмного модуля

Як наведено на рисунку 3.2 в процесі моделювання було виділено одну групу акторів «Користувач». Даний вибір базується на тому припущенні, що користуватись даним програмним додатком будуть користувачі для власних потреба або для вирішення поставлених завдань. Завдання які вирішуватиме програмний модуль не потребують сторонньої допомоги, а параметри роботи просто налаштовуються навіть користувачами з невеликим досвідом роботи. Актори групи «Користувачі» мають повний доступ до усіх функціонально важливих елементів програмного модуля та дозволяють у повній мірі використати можливості програмної розробки.

Окрім оцінки рівня доступності програмного модуля додатково було проведено оцінку складності роботи з ним. Для цього було проведено моделювання послідовності дій користувачів при роботі з програмою прототипом. Для оцінки даного аспекту було використано діаграму послідовності. Даний тип діаграм дозволяє проілюструвати та оцінити необхідні дії та часові затримки які виникають під час виконання певного етапу

загальної задачі. Результати моделювання проілюстровані у вигляді діаграми послідовностей (рисунок 3.3).

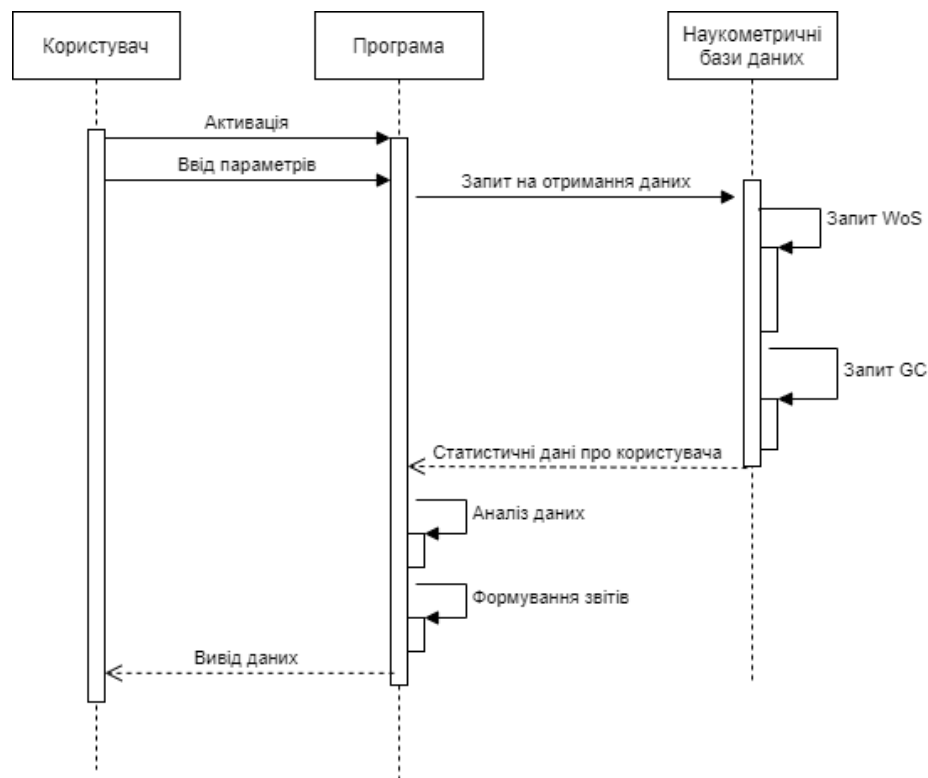


Рисунок 3.3 – Діаграма послідовності кроків взаємодії користувача та модулів програмної системи

Результати моделювання продемонстрували, що для успішного виконання завдання по визначенню рівня активності наукового працівника користувач має здійснити мінімальну кількість операцій, а саме ввести інформацію про особу аналіз якої необхідно провести. Додатково може виникнути необхідність налаштування програмного модуля під індивідуальні потреби користувача, проте даний фактор є незначним, оскільки процедура налаштування може здійснюватись один раз на етапі встановлення програмного модуля. Основні часові затримки виникають на етапі отримання статистичної інформації про особу з НБД. На час виконання даного етапу значно впливають зовнішні чинники, а саме швидкість інтернету та завантаженість серверів.

В загальному, проведене моделювання проілюструвало те що запропонована структура та підхід до принципів функціонування програмного модуля є правильним, відповідає поставленим завданням та може бути

використанам для подальшої програмної реалізації модуля оцінки рівня активності наукового працівника на основі його.

3.2 Функції взаємодії з серверами наукометричних баз даних

Для можливості функціонування розроблюваного модуля необхідно реалізувати функції створення запитів та обробки відповідей від серверів наукометричних баз даних. Для роботи з серверами Google Scholar необхідно сформулювати запит за заданим шаблоном. Приклад встановлення параметрів запиту наведено тижче:

```
paramsQuery = {  
    "api_key": os.getenv("API_KEY"),  
    "engine": " google_scholar_author ",  
    "q": "Oleh Pitsun",  
    "hl": "en",  
    "as_ylo": "2018",  
    "as_yhi": "2022",  
    "start": "0"  
}  
searchResult = GoogleSearch(paramsQuery)
```

Запит даного типу дозволить отримати інформацію по обрану особу, після чого можна буде провести парсинг отриманої відповіді з метою отримання статистичної інформації та подальшої її опрацювання. Параметри які були задані під час даного запиту дозволяють додатково регулювати структуру відповіді яка буде отримання у випадку успішного виконання запиту. В результаті запиту можна отримати відповідь наступного типу (рисунок 3.4):

					КР.КІ.8351890.00.00.000 ПЗ	Арк.
Змн.	Арк.	№ докум.	Підпис	Дата		44

```

{
  "total_citations": 34,
  "citations_per_year": {
    "2018 ": 73 ,
    "2019 ": 28 ,
    "2020 ": 30 ,
    "2021 ": 41 ,
    "2022 ": 36 ,
  },
  "publications": [
    {
      "title": " Evaluation methods of image segmentation quality ",
      "authors": " Pitsun O. Yo. Berezsky O. M. ",
      "venue": " Радіоелектроніка, інформатика, управління, 2018 ",
      "citations": 14,
      "year": 2018
    },
  ],
}

```

Рисунок 3.4 – Приклад відповіді на запит до сервера Google Scholar

Зокрема, в прикладі вказано, що інформація повинна бути висвітлена за чіткий часовий проміжок від 2018 по 2022. Це по-перше зменшить кількість інформації яку буде повертати сервер, що прискорить час роботи програми в цілому. По-друге одразу відсіче неінформативні дані в розрізі проведеного дослідження. Якщо необхідно отримати загальну інформацію про особу, то можна задати запит ось такого типу:

```

import http.client
autorInfo = http.client.HTTPSConnection(
"serpapi.webscrapingapi.com")
autorInfo.request("GET", "/v1?engine=google_scholar_author&
api_key=YOUR_API_KEY&author_id=LSfgHDeAA")
resultInfo = autorInfo.getresponse()
dataInfo = resultInfo.read()
print(dataInfo.decode("utf-8"))

```

В результаті сервер поверне структуровану інформацію, яка зберігається на ньому, а користувачі вже самостійно зможуть вибрати ті дані які необхідні їм для проведення відповідних досліджень.

Для роботи з наукометричною базою Web of Science. Необхідно скористатись іншим підходом до побудови запиту. Приклад побудови запину наведено нижче:

```

from pywos.crawler import WosQuery, construct_search
from pywos.analysis import Papers
import asyncio
qdWOS = construct_search(AI="D-2135-2015", PY="2018-2022")
wqWOS = WosQuery(querydict= qdWOS)

```

					КР.КІ.8351890.00.00.000 ПЗ	Арк.
						45
Змн.	Арк.	№ докум.	Підпис	Дата		


```
loopWOS = asyncio.get_event_loop()
taskWOS = asyncio.ensure_future(wqWOS.main(path="data"))
loopWOS.run_until_complete(taskWOS)
```

По-перше, ми повинні ініціалізувати об'єкт сканування, надавши dict запиту та dict заголовків для всіх http-з'єднань (необов'язково, для заголовків є агент користувача за замовчуванням). `WosQuery(querydict = {}, headers= {})`

Завдання збору даних викликається `WosQuery.main(path=)`. Усі параметри є необов'язковими `path`, крім , який є шляхом для збереження вихідних даних (не потрібно записувати `.json`). `citedchecke` логічним значенням, якщо встановлено значення `true`, також збираються всі цитати в статті запиту. І це є основою для детального аналізу цитувань, наприклад цитат за роками та цитат за іншими. В іншому випадку значенням за замовчуванням `citedchecke false`, у цьому випадку можна отримати лише загальну кількість цитат для кожного запиту. `limit` параметр дає максимальну кількість підключень у пулі підключень `http`. За замовчуванням значення дорівнює 20. Більше число передбачає більшу швидкість, але також передбачає більший ризик збою з'єднання через обмеження наукової мережі. `limit=30` успішно протестовано без збою з'єднання, і такої швидкості достатньо для обробки 1000 документів приблизно за 1 хвилину. Якщо завдання запиту занадто велике, найкраще ввімкнути параметр `savebyeach=True`, щоб кожен документ у запиті зберігався одразу після завантаження. Таким чином, у разі збою з'єднання ми можемо відновити завдання без повторного отримання всіх даних. Це визначається параметром `masklist` основної функції. Якщо `masklist` надано, для всіх `int`-номерів у цьому списку відповідний документ пропускається, щоб уникнути повторення роботи. У підсумку для великого завдання маємо наступні параметри.

Використання внутрішніх API серверів наукометричних баз даних дозволяє проводити параметризовані запити які в подальшому можуть використовуватись як основа реалізації програмних модулів для роботи з наукометричними базами даних та проводити дослідження по активності написання статей за різними напрямками, аналізувати кількість статей окремо взятого автора або групи співавторів, проводити аналіз рівня цитованості тощо.

					КР.КІ.8351890.00.00.000 ПЗ	Арк.
Змн.	Арк.	№ докум.	Підпис	Дата		46

3.3 Тестування програмного додатку

На етапі тестування необхідно перевірити функціональні моливості запропонованого програмного модуля, проаналізувати його стійкість до винекнення помилкових ситуацій як зі сторони клієнта так і зі сторони серверів наукометричних баз даних. Для цього було обрано робочу станцію з середніми тенічними характеристиками, оскільки даний програмний модуль буде опрацьовувати текстову інформацію, а для її обробки немає необхідності використовувати значні технічні ресурси. В даному випадку більші вимоги ставляться до стабільності та швидкомті зеднання між програмою та серверами Google Scholar та Web of Science. Проте, як показали результати тестування навіть при низькій пропускній здатності програмний модуль підтримує стабільну роботу, запити обробляються з допустимими часовими затратами, а загальне враження від користування розробленою програмою залишається позитивним.

Під час проведення тестових випробувань програма генерувала запити до серверів Google Scholar та Web of Science на основі заданих користувачем параметрів, зокрема на рисунку 3.5 проілюстровано порівняння результатів кількості піблікацій автором протягом заданого періоду часу.

					КР.КІ.8351890.00.00.000 ПЗ	Арк.
Змн.	Арк.	№ докум.	Підпис	Дата		47

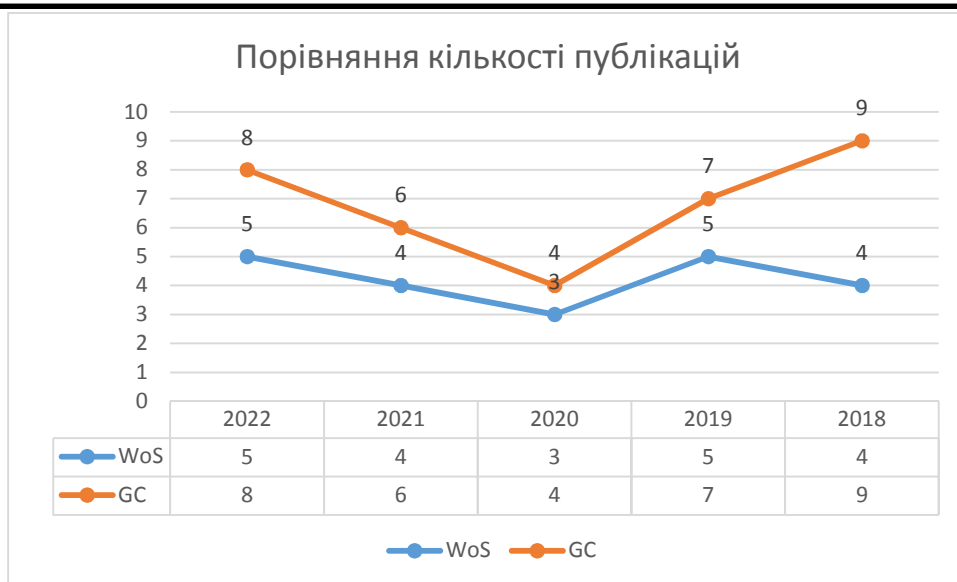


Рисунок 3.5 – Результати опрацювання запиту до серверів наукометричних баз даних для отримання кількості публікацій автора за заданий період часу

Після успішного виконання запиту до сервера, програмний модуль проводить парсинг отриманої інформації з метою отримання стурктуруваних даних про аналізовану особу. Дана інформація ілюструється у вигляді таблиці, де користувач може візуально оцінити зміни активності роботи наукового працівника протягом останніх п'яти років, та зробити первинні висновки. Для отримання кількісних показників оцінки роботи наукових працівників проводиться додаткова статистична обробка отриманих даних, зокрема обчислюється коефіцієнт активності наукового працівника (КАНП). Результати обчислення запропонованого показника наведено на рисунку 3.6.

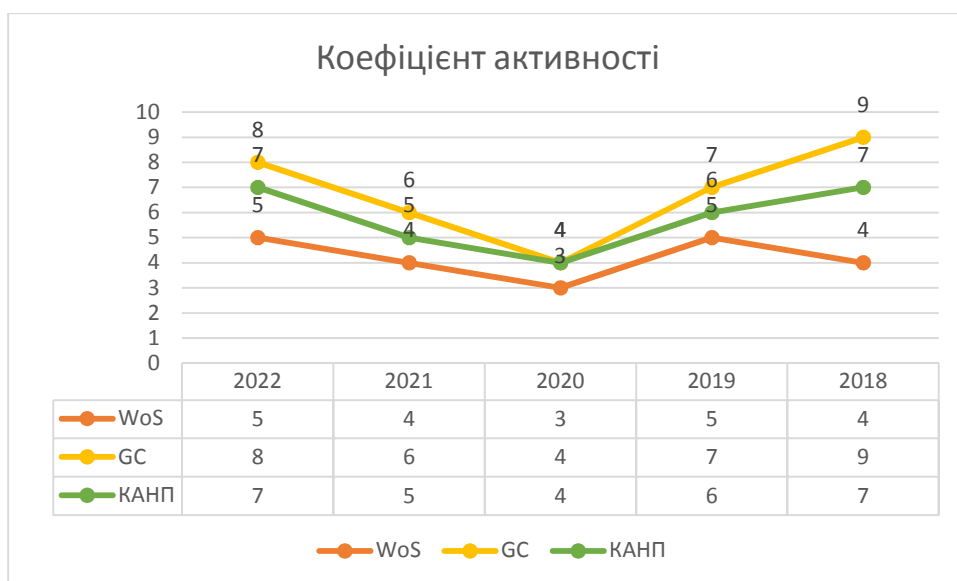


Рисунок 3.5 – Приклад обчислення коефіцієнт активності наукового працівника

Як проілюстровано на рисунку 3.6 запропонований параметр оцінки активності наукового працівника дозволяє прослідкувати динаміку в написанні наукових публікацій на основі даних наукометричних баз даних. Дані інформація може бути використана для побудови систем складання рейтингів науковців незалежно від сфери наукових інтересів, оскільки буде аналізуватись саме вид кривої. При зрозумінні якої можна говорити, що науковець активно займається дослідженнями та має хороші результати. В іншому випадку результати не є задовільними.

До переваг запропонованого алгоритму можна віднести:

- Можливість оцінки науковців з різних сфер діяльності.
- Запропонований критеріє базується на статистичних даних та містить високі об'єктивність під час процесу створення рейтингу.
- Моливість простої інтеграції в складні системи для підсилення їх функціональні можливостей.

Серед недоліків:

- Залежність роботи алгоритму від даних з наукометрични баз даних.

					КР.КІ.8351890.00.00.000 ПЗ	Арк.
						49
Змн.	Арк.	№ докум.	Підпис	Дата		

ВИСНОВКИ

На основі аналізу існуючих методологій та критеріїв оцінки активності наукових працівників та формування їх рейтингів, дослідження основних методів та алгоритмів під час статистичних досліджень, а також огляд програмних засобів по обробці великих масивів даних можна зробити наступні висновки:

1) Досліджено методології та технології оцінювання наукових працівників на основі різних активностей, що дозволило виділити публікації як окрему та проводити порівняння лише за нею;

2) Проведено дослідження наукометричних баз даних для визначення набору даних, які вони зберігають та надають для подальшого аналізу, що дозволило виділити наукометричні бази Web of Science та Google Scholar як одні з найпопулярніших та найвпливовіших на сьогодні;

3) Проведено аналіз особливостей роботи з API Web of Science, ознайомленось їхніми схемами зберігання даних та формування запитів;

4) Проаналізовано особливості роботи з API Google Scholar, що дозволило сформувати набір ключів які необхідні для створення запиту до даної наукометричної бази;

5) Розроблено алгоритм формування рейтингу викладачів на основі показників Google Scholar та Web of Science, що дозволило спроектувати програмний модуль формування рейтингу науковців.

6) Проведено програмну реалізацію та тестування спроектованої структури програмного модуля формування рейтингу викладачів на основі показників Google Scholar та Web of Science, дозволило провести порівняння розробленого програмного прототипу з програмами-аналогами.

					КР.КІ.8351890.00.00.000 ПЗ	Арк.
						50
Змн.	Арк.	№ докум.	Підпис	Дата		

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Glaser, J., Oltersdorf, J. Persistent Problems for a Bibliometrics of Social Sciences and Humanities and How to Overcome Them. 17th International Conference On Scientometrics & Informetrics. 2019.

2. Bornmann, L. Bibliometrics-based decision trees (BBDTs) based on bibliometricsbased heuristics (BBHs): Visualized guidelines for the use of bibliometrics in research evaluation. 2019.

3. Horovyi, V. M. (2015). Quality criteria for research in the context of national interest. Bulletin of the National Academy of Sciences of Ukraine, 6, 74–80. Retrieved from 18 Наукові праці НДФІ. 2019. № 4

4. Mryhlod, O. I., Kenna, R., Holovach, Yu. V., & Berche, B.. On the problem of science evaluation. Bulletin of the National Academy of Sciences of Ukraine, 10, 76–85.

5. Medvedieva, A. Analysis of the system of evaluation of activity of scientific institutions in Ukraine. Scientific works of the National Library of Ukraine named after V. I. Vernadskyi, 43, 270–284.

6. Karpenko, A. V., Budytska, Yu. O. Theoretical and methodological basics of evaluating scientific practice effectiveness in Ukraine. The Collection of Scientific Works of Kirovohrad National Technical University. Economic Sciences, 31, 51–57.

7. Kotliarevskyi, Ya. V., Radchenko, A. I., Melnykov, O. V., & Semeniuk, E. P. Strategic priorities in measuring the publication and publishing works in scholarly research activity. Science and innovation, Vol. 14, No. 5, 5–18.

8. Yehorov, I. Yu. System of integrated indicators for the assessment of scientific technical and innovation activities in the context of European integration. Science and innovation, Vol. 12, No. 4, 19–21.

9. Verkhovna Rada of Ukraine. About scientific and scientific-technical activity (Act No. 848-VIII, November 26). 2016.

					КР.КІ.8351890.00.00.000 ПЗ	Арк.
Змн.	Арк.	№ докум.	Підпис	Дата		51

10. Petrushyna, T. O. The state of science in Ukraine (based on the assessments of Ukrainian and foreign experts). Bulletin of the National Academy of Sciences of Ukraine, 11, 2017. 66–80.

11. Horovyi, V. (2015). Quality criteria for research in the context of national interest. Наукові праці НДФІ. 2019. № 4 (89).

12. Medvedieva, A. Foreign experience in evaluating the effectiveness of scientific activity. Bulletin of the Book Chamber, 6, 49–51.

13. Kopanieva, Ye. National indexes of scientific citation. Library Bulletin, 4, 2018. 29–35.

14. Didenko, Yu. V., Radchenko, A. I. Publication activity as a way of scientific communication and ratings race. Bulletin of the National Academy of Sciences of Ukraine, 9, 82–98.

15. Shapoval, Yu. I.. Scientific publication as a criterion for the evaluation of scientific activity: the experience of Poland. 2017.

16. ОDOTЮК І. Оцінка результатів наукової діяльності в Україні: нормативно-правовий аспект. Проблеми інноваційно-інвестиційного розвитку. 2012. № 3. . 38–42.

17. Main indicators of activity of higher educational institutions of Ukraine at the beginning of 2016/17 academic year: Statistical bulletin. State Statistics Service of Ukraine. Kyiv, 2017. 208 p. – URL: <http://www.ukrstat.gov.ua> [in Ukrainian]

18. About higher education: Law of Ukraine №1556VII dated July 1, 2014. URL: <http://zakon2.rada.gov.ua/laws/show/1556-18/print>. [in Ukrainian]

19. About scientific and scientific-technical activity: Law of Ukraine № 848-VIII of November 26, 2015. URL: <http://zakon2.rada.gov.ua/laws/show/848-19>

20. On the publication of the results of theses for obtaining scientific degrees of a doctor and a candidate of sciences: Order of the Ministry of Education and Science of Ukraine No. 1112 dated October 17, 2012 URL: <http://zakon2.rada.gov.ua/laws/show/z1851-12> [in Ukrainian]

21. Education : Expenditure on education as % of GDP (from government sources): UNESCO Institute for Statistics. Tertiary Education. – URL: <http://data.uis.unesco.org/>.

					КР.КІ.8351890.00.00.000 ПЗ	Арк.
Змн.	Арк.	№ докум.	Підпис	Дата		52

22. Education: Tertiary graduates by level of education: UNESCO Institute for Statistics. – URL: <http://data.uis.unesco.org/> .

23. UNESCO Science Report: Towards 2030. Paris: UNESCO Publishing, 2015.

24. Vse ukrainskie zhurnaly v Scopus i Web of Science [All Ukrainian magazines in Scopus and Web of Science]. URL <https://openscience.in.ua/uajournals>.

25. Katerynchuk, I. Kulyk, V. & Kravchuk, V. Pryntsypy otsiniuvannia rezultativ naukovotekhnichnoi diialnosti [Principles of evaluation of scientific and technological activities]. Visnyk TNTU - Bulletin of TNTU, 65, 154-161

26. Krush, P. V. & Zaborovets, O. P. Analiz teoretyko-metodolohichnykh pidkhodiv do otsinky efektyvnosti ta rezultatyvnosti innovatsiinoi diialnosti naukovykh ustanov Ukrainy [Analysis of theoretical and methodological approaches to evaluate the effectiveness and efficiency innovative activity of scientific institutions in Ukraine]. Suchasni problemy ekonomiky i pidpriemnytstvo - Modern Problems of Economics and Business, 5, 65-71

27. Medvedieva, A. Zarubizhnyi dosvid otsiniuvannia efektyvnosti naukovo diialnosti [The foreign experience of evaluating the effectiveness of research activities]. Visnyk Knyzhkovoï palaty - Bulletin of the Book Chamber, 6, 49-52.

28. Musina, L.A. & Kvasha, T.K. Pidkhody. Approaches, indicators and methods of evaluating the impact of scientific and technological activities in economic development. Kyiv: UkrINTEI, 252

29. Revak, I. O. Methodological approaches to the evaluation of scientific activity in Ukraine. Naukovi visnyk Lvivskoho derzhavnoho universytetu vnutrishnikh sprav. seriia ekonomichna - Scientific Bulletin of the Lviv National State University of Internal Affairs. Series: economic, 2, 343-353.

30. Iamchuk, A. V., Kuranda, T. K. & Yevtushenko, V. M. An evaluation of the effectiveness of scientific and technological activities. Naukovo-tekhnichna informatsiia - Scientific and technical information, 3, 3-6.

31. Спірін О. М., Іванова С. М., Кільченко А. В., Новицька Т. Л. Використання наукометричних баз даних і систем вебаналітики для

					КР.КІ.8351890.00.00.000 ПЗ	Арк.
Змн.	Арк.	№ докум.	Підпис	Дата		53

моніторингу електронних наукових фахових видань. Інформаційні технології в освіті. Херсон, 2020. № 45. URL: <https://lib.iitta.gov.ua/723135>.

32. Іванова С. М., Кільченко А. В. Використання рейтингового оцінювання системи Google Scholar у науковій діяльності. Інформаційні технології в освіті, науці й техніці”(ІТОНТ-2020): матеріали V Міжнар. наук.-практ. конф., м. Черкаси, 21-23 трав. 2020 р. Черкаси: ЧДТУ, 2020. С. 96-97. URL: <https://lib.iitta.gov.ua/720730>.

33. Інформаційні технології і засоби навчання: вебсайт. URL: <https://journal.iitta.gov.ua/index.php/itlt/indexing>.

34. Влох Р. О. Система оцінки українських фахових видань. Наука України у світовому інформаційному просторі. К.: Академперіодика, 2008. Вип. 1. С. 57- 94.

35. Білощицький А.О., Гогунський В.Д. Наукометричні бази та індикатори цитування наукових публікацій. Інформаційні технології в освіті, науці та виробництві. Вип. 4 (5). О.: АО Бахва, 2013. 198-203.

36. Бушуєв С.Д., Білощицький А., Гогунський В.Д. Наукометричні бази: характеристика, можливості і завдання. Інформатизація вищої освіти. Вип. 18. 2014. С. 146-152.

37. Мазаракі А., Притульська Н., Мельниченко С. Інтеграція вітчизняної науки до світової через наукометричні бази даних. Вісник КНТЕУ. № 6. 2001. С. 5-13.

38. Гальчевська О.А. Використання міжнародних наукометричних баз даних відкритого доступу в наукових дослідженнях. Інформаційні технології в освіті. № 23. 2015. С. 115-126.

39. Костенко Л., Жабін О. та ін. Бібліометрика науки: інформаційноаналітична система. Бібліотечний вісник. № 4. 2014. С. 8-12.

40. Alonso, S., Cabrerizo, F., Herrera-Viedma, E., & Herrera, F. h-Index: A review focused in its variants, computation and standardization for different scientific fields. Journal of Informetrics, 3, 273-289.

41. Ball, P. Index aims for fair ranking of scientists. Nature, 436, 900.

					КР.КІ.8351890.00.00.000 ПЗ	Арк.
Змн.	Арк.	№ докум.	Підпис	Дата		54

42. Batista, P.D., Campiteli, M.G., & Kinouchi, O. Is it possible to compare researchers with different scientific interests? *Scientometrics*, 68, 179-189.

43. Bornmann, L., & Daniel, H. What do we know about the h index? *Journal of the American Society for Information Science and Technology*, 58, 1381-1385.

44. Bornmann, L., Mutz, R., & Daniel, H. Are there better indices for evaluation purposes than the h index? A comparison of nine different variants of the h index using data from biomedicine. *Journal of the American Society for Information Science and Technology*, 59, 830-837.

45. Costas, R., & Bordons, M. The h-index: Advantages, limitations and its relation with other bibliometric indicators at the micro level. *Journal of Informetrics*, 1, 193-203. j.joi.2007.02.001

46. Cronin, B., & Meho, L. Using the h-index to rank influential information scientists. *Journal of the American Society for Information Science and Technology*, 57, 1275-1278.

					КР.КІ.8351890.00.00.000 ПЗ	Арк.
Змн.	Арк.	№ докум.	Підпис	Дата		55