

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ

Харківський національний університет радіоелектроніки

Факультет

Центр післядипломної освіти

(повна назва)

Кафедра

Програмної інженерії

(повна назва)

КВАЛІФІКАЦІЙНА РОБОТА

Пояснювальна записка

рівень вищої освіти

другий (магістерський)

Дослідження методів інтелектуального аналізу та керування

наукометричними даними

(тема)

Виконав:

Студент 2 курсу, групи ІПЗдм-20-1

Кулек О.П.

(прізвище, ініціали)

Спеціальніст

121 Інженерія

ь

програмного забезпечення

(код і повна назва спеціальності)

Тип програми

освітньо-наукова

Керівни

проф. Шубін І.Ю.

к

(посада, прізвище)

Допускається до захисту

Зав. кафедри

(підпис)

З.В. Дудар

(прізвище, ініціали)

2022 р.

Харківський національний університет радіоелектроніки

Факультет Центр післядипломної освіти
(повна назва)

Кафедра Програмної інженерії
(повна назва)

Рівень вищої освіти другий (магістерський)

Спеціальність 121 Інженерія програмного забезпечення
(код і повна назва спеціальності)

Тип програми освітньо-наукова
(освітньо-професійна або освітньо-наукова)

Освітня програма Інженерія програмного забезпечення
(повна назва)

ЗАТВЕРДЖУЮ:

Зав.кафедри _____
(підпис)

« ____ » _____ 2022 р.

**ЗАВДАННЯ
НА КВАЛІФІКАЦІЙНУ РОБОТУ**

студента Кулек Олександр Петровичу
(прізвище, ім'я, по батькові)

1. Тема роботи Дослідження методів інтелектуального аналізу та керування наукометричними даними

затверджена наказом університету від 24.03.2022 р. № 31 Стз

2. Термін подання роботи до екзаменаційної комісії 15 05 2022р.

3. Вихідні дані до роботи _____


4. Перелік питань, що потрібно опрацювати в роботі мета роботи, аналіз проблемної галузі і постановка задачі, опис запропонованих варіантів оптимізації, використовувані методи та алгоритми, опис розробленої програмної системи, опис застосованих програмних рішень, аналіз можливих застосувань

КАЛЕНДАРНИЙ ПЛАН

№	Назва етапів роботи	Терміни виконання етапів роботи	Примітка
1.	Аналіз предметної галузі	31 березня 2022	<i>виконано</i>
2.	Огляд існуючих методів	10 квітня 2022.	<i>виконано</i>
3.	Розробка алгоритмів, проектування та розробка ПЗ	15 квітня 2022	<i>виконано</i>
4.	Підготовка пояснювальної записки	20 квітня 2022	<i>виконано</i>
5.	Спецчастина	28 квітня 2022	<i>виконано</i>
6.	Підготовка презентації та доповіді	03 травня 2022	<i>виконано</i>
7.	Попередній захист	05 травня 2022	<i>виконано</i>
8.	Нормоконтроль, рецензування	07 травня 2022.	<i>виконано</i>
9.	Занесення роботи в електронний архів	08 травня 2022.	<i>виконано</i>
10.	Допуск до захисту в зав. кафедри	11 травня 2022	<i>виконано</i>

Дата видачі завдання 28 березня 2022р.

Студент


(підпис)

Керівник роботи

(підпис)

проф. Шубін І.Ю.

(посада, прізвище, ініціали)

РЕФЕРАТ / ABSTRACT

Кваліфікаційна робота магістра містить: 87 с., 24 рис., 5 табл., 38 джерел.

НАУКОМЕТРИЯ, КВАЛИМЕТРИЯ, НЕЙРОННА МЕРЕЖА, ЛИНГВИСТИЧНИЙ АНАЛІЗ.

Об'єктом дослідження є методи онтологічних описів.

Метою роботи є дослідження й розробка математичних методів і алгоритмів аналізу лінгвістичної експертної інформації, отриманої з колекцій текстів заданої предметної області, з метою її формалізації для застосування в інтелектуальних системах обробки наукової інформації.

Методи розробки – методи імовірнісного тематичного моделювання, методи кластерного аналізу, статистичний експеримент.

У результаті реалізовано алгоритм аналізу й обробки наукометричної експертної інформації стосовно до завдання побудови онтології

SCIENTIFICS, QUALIMETRY, NEURAL NETWORK, LINGUISTIC ANALYSIS.

The object of research is the methods of ontological descriptions.

The aim of the work is research and development of mathematical methods and algorithms for analysis of linguistic expert information obtained from collections of texts of a given subject area, in order to formalize it for use in intelligent systems of scientific information processing.

Methods of development - methods of probabilistic thematic modeling, methods of cluster analysis, statistical experiment.

As a result, the algorithm of analysis and processing of scientometric expert information in relation to the task of constructing an ontology is implemented.

Умови публікації пояснювальної записки

Я, _____ Кулек Олександр Петрович _____
(прізвище, ім'я, по батькові)

студент групи _____ ІІЗдм-20-1 здобувач вищої освіти на другому
(магістерському) рівні

кафедра програмної інженерії,
(повна назва кафедри)

заявляю: моя кваліфікаційна робота на тему Дослідження методів
інтелектуального аналізу та керування наукометричними даними ,
(назва роботи)

що буде представлена до ЕК для публічного захисту, виконана самостійно, в ній не містяться елементи плагіату і вона може бути опублікована в електронному архіві відкритого доступу EIArKhNURE. Всі запозичення з друкованих та електронних джерел мають відповідні посилання.

Я ознайомлений з діючим положенням «Про протидію академічному плагіату в ХНУРЕ», згідно з яким виявлення плагіату є підставою для відмови в допуску кваліфікаційної роботи до захисту та застосування дисциплінарних заходів.

ЗМІСТ

Вступ	8
1 Аналіз стану розв'язання проблеми та обґрунтування цілей дослідження	11
1.1 Аналіз джерел наукометричних даних	11
1.2 Аналіз методів розрахунку імпаکت-фактору	15
1.3 Методи ранжування і нормування	17
1.4 Огляд програмних рішень в області наукометрії	20
1.5 Обґрунтування цілей дослідження	22
2 Опис проведених теоретичних досліджень	24
2.1 Огляд підходів до оцінювання публікаційної активності	24
2.2 Методи параметричної теорії систем	27
2.3 Методи інтеграції окремих знань	29
2.4 Методи кваліметричного аналізу наукової діяльності	37
3 Аналіз результатів досліджень	44
3.1 Алгоритми побудови рейтингу публікаційної активності	44
3.2 Алгоритм оцінки публікаційного потенціалу вченого	49
4 Опис розробленого програмного забезпечення	55
4.1 Проектування програмної системи	55
4.2 Проектування аналітичної моделі	57
5 Програмна реалізація паралельного алгоритму аналізу зображень	62
Висновки	65
Перелік джерел посилання	66
Додаток А Перелік джерел посилання за науковими напрямками керівника та науковців кафедри програмної інженерії	70
Додаток Б Звіт результатів перевірки на унікальність тексту	71

Додаток В Слайди презентації	73
Додаток Г Листінг модуля	81
Додаток Д Апробація роботи.....	84
Додаток Е Експертний висновок результатів перевірки кваліфікаційної роботи на відповідність оформлення вимогам ДСТУ	86

ВСТУП

У світовій практиці функціонування інституту науки, як основи оцінки наукової думки й продуктивності наукової діяльності використовуються два підходи – експертний (якісний) і наукометричний (кількісний) [1]. В останні роки в якості інструмента оцінки ефективності діяльності українських учених, дослідницьких організацій, вітчизняної науки в цілому стали активно використовуватися дані про рівень і число публікацій, число і якості цитованих, представлених різноманітними наукометричними показниками. Публікаційні й цитатні наукометричні показники розглядаються як цільові індикатори стану науки.

Незважаючи на те, що немає об'єктивних підстав вважати, що наукометричні показники, засновані на вимірі публікацій і цитат, здатні прямо відбивати рівень і якість наукових досліджень, як таких [2], багато дослідників відзначають, що розробка наукометричних вимірів і їхнє зіставлення відкривають нові грані навчальної діяльності, дають можливість скорегувати її напрямки, структуру й зміст, дати їй оцінку, зрівняти її результати з результатами інших досліджень.

Сьогодні дослідникові для рішення завдання найбільш ефективного опублікування результатів досліджень, у числі іншого, необхідно відповісти на наступні питання:

- у якому виді й обсязі опублікувати нові наукові результати;
- які видання можуть бути рекомендовані для опублікування наукових результатів;
- як вибрати авторитетний науковий журнал;
- як на ці питання відповідають найбільш успішні колеги.

Вочевидь, не всякий дослідник має час та компетенції, щоб відповідати на

ці питання, при цьому від відповідей на них багато в чому залежить його репутація й доступність результатів досліджень для світового наукового співтовариства. Особливо, дана проблема загострюється на рівні керівництва науковими колективами, лабораторіями й дослідницькими організаціями в цілому, коли керівникові необхідно ухвалювати управлінські рішення із приводу фінансування тих або інших досліджень, забезпечення кадровими й іншими ресурсами. Керівник потребує аналітичної інформації щодо перспективності тих або інших досліджень, дослідницького й публікаційного потенціалу тих або інших дослідників.

Деяку інформацію для керівника надають міжнародні реферативні й бібліографічні бази даних і їхні аналітичні додатки. Що розраховуються в них індивідуальні наукометричні показники відіграють ключову роль при вивченні динаміки наукових досліджень, побудові міжнародних рейтингів дослідницьких і освітніх організацій, проходженні атестації дослідниками й ухваленні рішення про виділення дослідницьких грантів.

На сучасному етапі розвитку інформаційних технологій важливу роль здобувають інтелектуальні рекомендаційні системи. Рекомендаційні системи – це програми, які намагаються передбачити, які об'єкти (фільми, музика, книги, публікації, веб-сайти) будуть цікаві користувачеві, маючи певну інформацію про нього. Наукометричний профіль дослідника або дослідницької організації в цілому в тієї або іншій наукометричній базі даних цілком відповідає необхідним умовам розробки рекомендацій.

Таким чином, в цей час доцільна розробка математичних моделей, алгоритмів та програмної системи, покликаної допомогти керівникам дослідницьких груп, а також індивідуальним дослідникам приймати рішення із приводу поліпшення публікаційної активності, наукометричної результативності, визначати слабкі й сильні сторони, напрямки й конкретні рішення у вигляді автоматизованих рекомендацій по підвищенню величини й репрезентативності наукометричних показників.

Розглянута тема здобуває особливу актуальність для підвищення

конкурентоспроможності університетів серед провідних світових науково-освітніх центрів.

Таким чином, метою роботи є аналіз методів логічного ранжування, методів кваліметрії, що застосовуються в науково-педагогічній діяльності, методів визначення наукометричних показників та розробка системи підтримки прийняття рішень у керуванні науковими дослідженнями, що формує індивідуальні рекомендації з поліпшення публікаційної активності.

1 АНАЛІЗ СТАНУ РОЗВ'ЯЗАННЯ ПРОБЛЕМИ ТА ОБҐРУНТУВАННЯ ЦІЛЕЙ ДОСЛІДЖЕННЯ

1.1 Аналіз джерел наукометричних даних

В області наукометрії й інфометрії в цілому напрацьований великий базис математичних методів для дослідження динаміки наукових результатів і публікаційної активності. Відомі в міжнародній практиці аналітичні методи й показники, що найбільш повно відповідають завданню індивідуальної наукометричної оцінки автора (Дж. Гірш (J. Hirsch)), статті і журналу (Ю. Гарфілд (E. Garfield)) і т.п. Удосконалюється аналітичний інструментарій (Incites, Scival etc.) у програмних системах міжнародних реферативно-бібліографічних баз даних Scopus, Web of Science тощо.

Разом з тим, недостатня увага, як у теорії, так і в практиці приділена підтримці прийняття управлінських рішень у сфері менеджменту наукової діяльності:

- відсутні універсальні методики оцінки публікаційної активності й побудови публікаційного рейтингу залежно від управлінських завдань;
- недостатньо вивчені питання формування й керування публікаційним потенціалом окремих дослідників і дослідницьких колективів;
- не пророблені питання роботи з «більшими» наукометричними даними із застосуванням методів інтелектуального аналізу;
- відсутні програмні розробки в області підтримки ухвалення рішення й розробці індивідуальних рекомендацій з поліпшення публікаційної активності, наукометричної результативності й реалізації дослідницького й публікаційного потенціалу.

Наукометрія визначена як дисципліна, що вивчає еволюцію науки через обчислення виміру й статистичну обробку наукової інформації. Наукометрія виникла зі статистичної бібліографії (даний термін увів бібліограф Е. Хамл в 1923

р., застосувавши його до ранжирування країн по числу журнальних статей у певних областях) і в цей час поряд з бібліометрією і вебометрією, є складовою частиною інфометрії.

Перші спроби кількісного вивчення публікацій були зроблені А. Шторхом і Ф. Аделунгом [3], ними в 1810 г. був складений бібліографічний показчик, що відбиває публікації 1801-1806 рр. по різних параметрах. Дослідник П. Вальден [4] в 1911 р. уперше застосував метод аналізу цитування для вивчення внеску вчених окремих країн у розвиток хімії. Потім, у 1917 р. була опублікована робота Ф. Коула (F. Cole) і Н. Ильса (N. Eales) [5], у якій був проведений статистичний аналіз наукових публікацій по порівняльній анатомії.

Базовий термін «бібліометрія» вперше зустрічається в роботах засновника теорії інформатики й розроблювача універсального десяткового класифікатора (УДК) П. Отле (P. Otlet) в 1934 р. Із цього часу інтенсивно розвивається застосування математичних і статистичних методів до видань і різних публікаціям, а також виявляються базові закономірності, використовувані в інфометрії (розподіл Ципфа [6], закон розсіювання Бредфорда [7], закон зворотного квадрата тощо).

Наприкінці 50-х років ХХ століття у зв'язку з експонентним ростом числа наукових журналів і публікацій була звернена увага на зміну характеру публікаційної активності й значимості наукових досліджень. Однак, наукометрія як окремий напрямок формулюється пізніше – в 60-ті роки, при цьому сам термін «наукометрія» уведений тільки в 1969 р.

Багато хто з читачів знає, що термін наукометрія, який відомий в англійській літературі як scientometrics, запропонував радянський вчений Васілій Налімов, який у 1969 році разом із Зінаїдою Мульченко опублікував монографію «Наукометрия. Изучение развития науки как информационного процесса». Таким чином З. Мульченко є співвинахідницею цього терміну.

Рональд Руссо дослідив це питання і виявив, що наукометристи просто часто цитували не ту працю – Налімов використав цей термін ще в 1966 році у роботі «Количественные методы исследования процесса развития науки» [6].

Р. Руссо додає: «Радянська наукометрія бере свій початок із робіт двох вчених – Геннадія Доброва та Васілія Налімова. Вони працювали приблизно в той самий час, один у Києві, а інший у Москві». Самі Налімов та Мульченко пишуть у своїй монографії, що: «В СРСР наукознавством на професійному рівні займається у Києві Г.М. Добров». Монографія Г.М. Доброва «Наука о науке. Введение в общее науковедение» побачила світ у 1966 році. Elsevier та угорська Академія Кіадо створять один відомий фаховий журнал тільки в 1978 році. До речі, в склад редколегії цього журналу увійде й Г. Добров, разом з Д. Джоном де Соллою Прайсом, Ю. Гарфілдом, М. Моравчіком, Т. Брауном.

Наукометрія, наукознавство, наука про науку.– потрібно зазначити, що це були лише початки сучасної наукометрії, а появою індексів цитувань, з якими у більшості асоціюється наукометрія, ми взагалі завдячуємо юристам та індексу Шепарда, що з'явився ще у далекому 1873 році.

Міхаїл Коковській досліджує, що термін «наукознавство» вперше виник у Польщі в 1910–20-х роках як наслідок інтелектуальної співпраці рівненського бібліотекаря Станіслава Міхальського та представників Львівсько-Варшавської філософської школи. У той час українські освітяни та науковці у Львові потерпали через постійні заборони, переслідування та ув'язнення, однак це не завадило їм навчати, досліджувати та навіть організувати Таємний український університет [8].



Рисунок 1.1 – Сенат Українського таємного університету у Львові, 1921 рік [8]

Тому, дослідники припускають, що подібний термін міг запропонувати,

наприклад, Микола Чайковський, або Володимир Вергановський який цілком міг знати про індекс Шепарда. Можливо, такий невідкритий рукопис українського класика десь лежить та чекає на істориків наукометрії.

У Львові народився й Альфред Лотка – автор одного з класичних наукометричних законів. Прайс помер у 1983 році й щоб вшанувати його видатні успіхи спільнота заснувала професійну нагороду на його честь. Сьогодні Derek de Solla Price Memorial Medal – це найвища нагорода в галузі наукометрії. У 1987 році нею був нагороджений Васілій Налімов.

Один з батьків світової наукометрії – жертва сталінських репресій (провів у в'язницях та на засланні приблизно 18 років) (див. рис. 1.2) [9].



Рисунок 1.2 – Васілій Налімов у таборі на Колимі. 1938 рік [9]

Сам В. Налімов писав у своїх спогадах: «За минулі 75 років ми навчилися хоча й мало чого, але дуже важливого – ми навчилися боятися ідеологізації, яка легко переходить в сувору міфологізацію». Помер В.Налімов у 1997 році.

Також, розпочинаючи з 1966 р. було організовано кілька радянсько-польських симпозіумів з проблем наукознавства (зокрема у Львові та Ужгороді), у яких Добров та Налімов регулярно брали участь. В 1979 р. у статтях Л. Блэкерта (L. Blackert), С. Зигеля (S. Siegel) [3] і О. Наке (O. Nacke) [10] з'явився термін «інфометрія», який був визначений як комплекс математичних методів для

дослідження об'єктів інформаційної науки, опису й аналізу їх властивостей, а також заводити, увести до закономірностей з метою оптимізації цих об'єктів при прийнятті рішень. У наслідку інфометрія стає більш широким поняттям, що включає і наукометрію [10].

В основі наукометричного інструментарію лежать закономірності й підходи, сформульовані в рамках бібліометрії, але одних їх недостатньо для дослідження наукових публікацій [11]. Із цієї причини виникли аналітичні методи й показники, які найбільш повно відповідають завданню індивідуальної наукометричної оцінки: імпакт-фактор, індекс Гірша, тощо.

1.2 Аналіз методів розрахунку імпакт-фактору

У міру глобалізації науки й стрімкого зростання обсягу наукової періодики виникла необхідність диференціювання наукових журналів по значимості і якості. Ідея формування якогось рейтингу наукових журналів і додання йому кількісного вираження належить Ю. Гарфілду (E. Garfield) [11]. З 1964 р. в очолюваному їм Інституті наукової інформації (ISI) став випускатися, так званий Science Citation Index (SCI) [12], що систематизує інформацію про цитування наукових статей по бібліографічній базі даних. SCI став прообразом бібліографічного ресурсу Web of Knowledge компанії Thomson Reuters, що включає, зокрема, базу бібліографічних даних Web of Science (Wos) з додатком Journal Citation Reports (JCR).

За Гарфілдом імпакт-фактор наукового журналу – це середня цитованість опублікованих у ньому статей важливість, що характеризує, і авторитетність даного [13]. Класичний імпакт-фактор визначається як кількість посилань, що отримано у даному році із усіх наявних у бібліографічній базі журналів, на статті, опубліковані в даному журналі протягом двох попередніх років, віднесені до цих статей. У подальших роботах ISI показано, що для певних завдань

(диференціації нових журналів, аналізу передових областей дослідження й т.п.) потрібна модифікація розрахунку імпаکت-фактора, наприклад, збільшення періоду обліку цитат.

Для наукометричної характеристики окремого дослідника (за аналогією з імпакт-фактором для журналів) Дж. Гіршем (J. Hirsch) [14] в 2005 р. був запропонований і, надалі, модифікований [15] спеціальний h -індекс. Визначається h -індекс в такий спосіб: дослідник, що опублікував N статей, має індекс рівний h , якщо h його статей цитуються не менш ніж h раз кожна, а кожна з тих, що залишилися ($N - h$) статей цитується не більш ніж h раз. Даний індекс позиціонується його автором, як такий, що характеризує актуальність, затребуваність і накопичений вплив індивідуальних наукових результатів [16], незважаючи на те, що дані властивості h -індексу неодноразово зазнали й зазнають конструктивної критики [17].

Зокрема, оскільки при досягненні дослідником більших значень індексу проявляється його інерційність (він може роками залишатися постійним), для відстеження й прогнозування публікаційної динаміки багатьма авторами пропонуються так звані раціональні модифікації, наприклад h -індекс [18]. Ціла частина цього показника еквівалентна звичайному індексу Гірша, а дробова частина показує, наскільки дослідник наблизився до наступного значення індексу Хірша.

Крім того, h -індекс сильно варіює в різних галузях науки (з різною нормою цитування й співавторства), тому для забезпечення порівнянності наукометричних результатів [19], запропонований доповнений індекс Гірша (h'), що обчислюється за формулою :

$$h' = h_2/N,$$

де N – число співавторів для даної публікації.

Різноманітні по цілям і властивостям модифікації індексу Гірша представлені у вітчизняних роботах, наприклад [20]. Умова рівності числа статей

і числа, що отримано кожної із цих статей цитат (посилань) є у визначенні індексу Гірша досить довільним – у [20] «узагальнили» визначення Гірша, назвавши його h_α -індексом. Індекс рівний h_α , якщо кожна з h_α статей масиву публікацій одержала не менш αh цитувань, а кожна з інших – не більш αh цитувань. Основний зміст даного індексу полягає в тому, що, змінюючи α можна міняти акцент оцінки: при збільшенні α на перше місце виходить цитованість, наявність висока цитованих робіт, а при зменшенні важливіше стає продуктивність, велика кількість публікацій.

1.3 Методи ранжування і нормування

При використанні наукометричних показників для оцінки необхідно забезпечувати порівнянність. При підрахунку необхідно чітко позначення як публікаційного вікна, так і вікна цитування, у протилежному випадку коректна стає неможливою. Крім того, відомо, що значення абсолютних наукометричних показників сильно варіюють залежно від області дослідження, періоду часу, географії й ін.

Ще одним підходом при проведенні крос дисциплінарних сопоставлень є відмову від використання власних значень показників і зіставлення їх рангів, тобто місць або позицій у рейтингу. Наприклад, після сортування списків досліджуваних об'єктів (дослідників, журналів, статей) у двох різних дисциплінах у порядку убутання якого-небудь обраного наукометричного показника проводиться аналіз тільки зайнятих порядкових місць.

Суть рангового методу – розбивка отриманого впорядкованого списку рейтингу на n рівних частин і визначення, у яку із цих частин попадають досліджувані об'єкти. Якщо n приймається рівним 4, тоді говорять про кuartилі. Наприклад, у ранговій системі вважається, що журнали, що потрапили в перший

квартиль, вище журналів, що потрапили в другий, не тільки в цієї, але й у будь-якій іншій дисципліні. При цьому важлива коректність і повнота складання списку для ранжирування. Перевага рангових методів – захист від сильної асиметричності розподілів по даному наукометричному показнику.

Даний метод широко розповсюджений формуються, як світові наукометричні рейтинги університетів (World University Rankings, Academic Ranking of World Universities, Leiden Ranking), так і окремих дослідників (наприклад по базі даних Web of Science Core Collection компанія Clarivate Analytics публікує рейтинг 250 самих цитованих вчених по 22 предметних областях). При цьому наукометричні показники (кількість статей, їх цитованість, імпаکت-фактори й ін.) виявляються важливою складовою зваженої оцінки, а в деяких рейтингах – єдиною.

Бібліографічною називається база даних, що містить сукупність бібліографічних описів видань і окремих публікацій у структурованому (посилальному) виді. Реферативні бази даних, що містять крім бібліографічних даних також реферативні дані (мета-дані) про видання й окремих публікаціях, зокрема анотації.

Як правило, бібліографічні бази даних є реферативними, але можуть містити в тієї або іншій формі первинні тексти (наприклад, у повні тексти зберігаються для відображення контексту цитування й контекстного пошуку, але при цьому можуть бути не доступні користувачам за прямим призначенням).

На сьогоднішній день існує близько сотні міжнародних і національних бібліографічних і реферативних баз даних, диференційних як по галузях знання, так і за обсягом і сфері використання. Серед них у якості основних як по загальній думці, так і на основі критеріїв Міністерства освіти й науки України, визначено НМБ Scopus і Web of Science – це бази даних по науковим публікаціям, у яких обробляються бібліографії публікацій, анотації до публікацій і списки використаної літератури в публікаціях.

На підставі інформації із цих баз даних формуються основні наукометричні показники (індекс Гірша, показники цитованості, імпакт-фактори) наукової

ефективності організацій, учених і наукових періодичних видань. Саме виникнення й розвиток реферативно-бібліографічних баз даних на основі інформаційних технологій обумовив інтенсивний розвиток методів наукометрії і їх практичну значимість.

НМБ Scopus (Sciverse Scopus) – найбільша у світі мультидисциплінарна бібліографічна й реферативна база даних, створена видавницькою корпорацією «Elsevier». Scopus охоплює понад 18 тис. наукових журналів (52% з них європейські) від 5 000 наукових видавництв миру, включаючи близько 13 млн патентів США, Європи і Японії, а також матеріали наукових конференцій. Дані про публікації з Scopus використовуються при складанні міжнародних і вітчизняних рейтингів університетів: QS World/BRICS/EECA, Times Higher Education.

НМБ Web of Science (Wos) – мультидисциплінарна реферативно-бібліографічна база даних Інституту наукової інформації США (Institute for Scientific Information, ISI), представлена на платформі Web of Knowledge компанії «Thompson Reuters». Web of Science охоплює понад 12 000 журналів й 148 000 матеріалів конференцій в області природних, суспільних, гуманітарних наук і мистецтва. Дані про публікації з Web of Science використовуються при складанні міжнародних і вітчизняних рейтингів університетів: U-Multirank, Round Ranking, US News, Global Institutional Profiles Project.

Наразі, наукометричні бази даних і розраховані по них індивідуальні наукометричні показники відіграють ключову роль при вивченні динаміки наукових досліджень, побудові міжнародних рейтингів дослідницьких і освітніх організацій, проходженні атестації дослідниками й ухваленні рішення про виділення дослідницьких грантів.

1.4 Огляд програмних рішень в області наукометрії

У наш час у розглянутій області переважно застосовуються вбудовані аналітичні додатки до реферативних баз даних. Кожний такий додаток є пропрієтарним (Incites, Scival і ін.) і працює тільки зі своєї наукометричною базою, у той час як докладний аналіз вимагає зіставлення показників з Scopus, Web of Science і внутрішньо-університетської інформаційної системи, у тому числі бази портфоліо й т.п.

Досягнення в області інформаційних технологій орієнтують на відкритість наукометричних аналітичних додатків (наприклад, Google Scholar, Microsoft Academic Search) і дані сервіси нарощують потенціал [5]. Спроба інтеграції декількох джерел даних на принципах відкритості почата, але її не можна визнати самостійною, оскільки даний проект припускає одночасну роботу тільки з одним джерелом даних (у режимі перемикання між джерелами), а також не надає серйозних аналітичних можливостей.

Зпропонований алгоритм автоматизації процесу аналізу цитування співавторів і структура відповідної бази даних. Вихідні дані попередньо розташовуються в спеціальній базі даних Webcitation. В основі розробленого програмного інструмента лежать методи кластерного аналізу, за допомогою яких у результаті формується карта кластерів зв'язаних авторів по тій або іншій тематиці.

У роботі [20] представлений програмний інструмент аналізу бібліометричних і наукометричних даних, який на основі розрахунку ряду стандартних показників надає нормовані результати для дослідників, науково-дослідницьких груп, дослідних центрів, факультетів або університету в цілому. Інструмент розроблений на Java і використовує централізовану університетську MySQL базу даних, у яку адміністратором довантажуються необхідні відомості про авторів, статті й журналах.

Програмний продукт «Publish or Perish Bibliometric Data Tools», [13], інтегрує набір утиліт, що дозволяють передробляти вихідні дані Google Scholar і Microsoft Academic Search, а потім на основі їх аналізу розрахувати ряд показників (див. рис. 1.3):

- загальна кількість публікацій, а також загальна кількість цитованих;
- середнє число цитат на публікацію, на автора, число цитат у рік і ін.;
- базисний h -індекс і його основні модифікації (у т.ч. середньорічний приріст індексу).

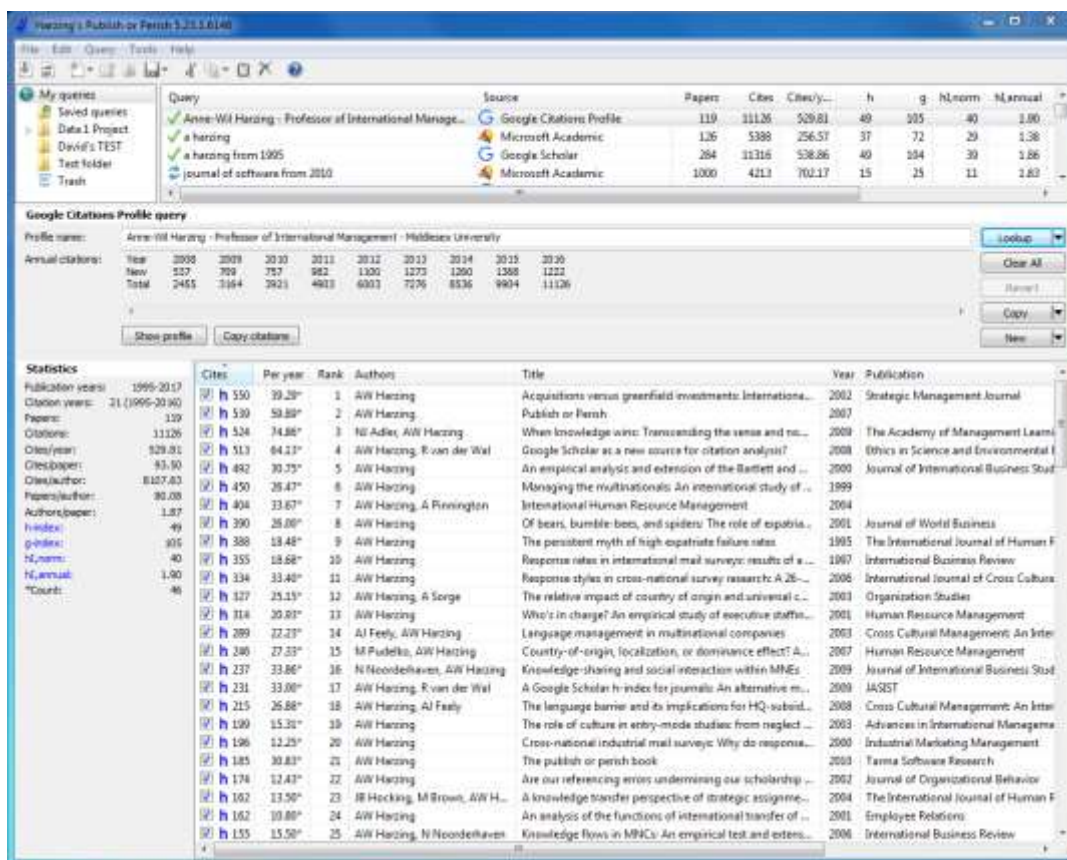


Рисунок 1.3 – Робоча форма Pop Tools [13]

Серед вітчизняних наукових і прикладних розробок у даній області виділяється інтелектуальна вимірвальна система «Ейдос», адаптована до аналізу даних [26]. Система є вітчизняним ліцензійним програмним продуктом і призначена для поглибленого аналізу результатів тестування, що включає ранговий (інформаційний) і кластерно-конструктивний аналіз еталонних описаних класів розпізнавання й ознак, а також аналіз вірогідності заповнення

вихідних даних.

У рамках завдань обчислення наукометричних показників, система «Ейдос» на основі узгодження думок експертів і статистичного аналізу даних формує класифікатор (див. рис. 1.4) і визначає статус окремого автора (вченого).

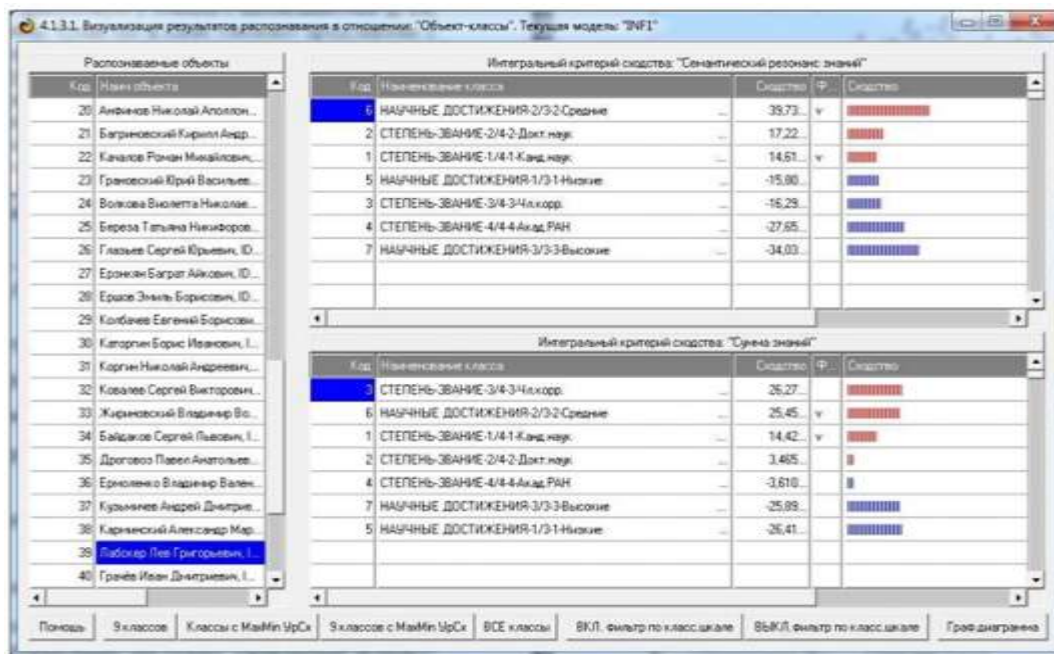


Рисунок 1.4 – Робоча форма системи Ейдос [14]

Представлена модульна архітектура системи аналізу публікаційної активності співробітників науково-освітньої організації, що включає модуль для керування простором співробітників науково-освітньої організації, модуль для керування областю публікаційних об'єктів і модуль для обчислень і аналізу.

1.5 Обґрунтування цілей дослідження

Аналіз показує, що коректним рішенням в описаних умовах може стати комплексна методика, що включає нормування, виправлення на період наукової і публікаційної активності, а також порогову агрегацію й ранжирування. Представлений лише опис концептуальних блоків і їх загальна взаємодія в рамках системи, не орієнтоване на яку-небудь конкретну реалізацію або варіанти

використання.

Питання розробки інструментів автоматизації й інтеграції наукової інформації в просторі різнорідних інформаційних систем, для цієї мети розроблений спосіб автоматизованого збору метаданих про публікації в репозиторій і деякі шаблони обробки даних.

Дана мета визначає наступне коло завдань:

- огляд робіт в області оцінки публікаційної активності вчених;
- розробка методики побудови публікаційного рейтингу на основі профілів учених в Scopus;
- розробка програмної системи для формування рекомендацій на основі аналізу наукометричних даних;
- проведення програмних експериментів.

Теоретична складова роботи полягає в тому, що в ній сформульовано визначення публікаційного потенціалу вченого; даний формальний опис методів інтелектуального аналізу наукометричних даних для рішення завдання оцінки публікаційного потенціалу вченого й розробки індивідуальних рекомендацій з поліпшення публікаційної активності.

Практична складова роботи полягає в тому, що на базі запропонованих підходів і алгоритмів розроблена програмна система, покликана допомагати керівникам дослідницьких груп, а також дослідникам ухвалювати рішення щодо приводу підвищення публікаційної активності, наукометричної результативності, визначати слабкі й сильні сторони, напрямку й конкретні рішення у вигляді автоматизованих рекомендацій з підвищення величини й репрезентативності наукометричних показників.

Методологічною основою дослідження є реляційна алгебра і методи інтелектуального аналізу даних (datamining).

2 ОПИС ПРОВЕДЕНИХ ТЕОРЕТИЧНИХ ДОСЛІДЖЕНЬ

2.1 Огляд підходів до оцінювання публікаційної активності

Наукометричні показники дозволяють тією чи іншою мірою вирішувати два основні завдання:

- оцінити внесок в обмін науковими результатами, здійснити моніторинг публікаційної активності й виконати її зіставлення по окремих дослідниках, наукових колективах, дослідницьким організаціям;
- управляти публікаційною активністю, встановлювати тактичні й стратегічні наукометричні орієнтири, підтримувати баланс публікаційної активності в науковій діяльності.

Виходячи із цих, як правило, послідовно розв'язуваних завдань можна узагальнено сформулювати ряд принципів, що дозволяють, з однієї сторони – позначити їхню таксономію, а з іншої сторони – визначити критерії їх застосування в контексті кожного завдання. У числі основних принципів формування системи наукометричних показників і інтегральної оцінки публікаційної активності, можна назвати наступні.

Некомпенсаторність [21] оцінки – неможливість компенсації низького значення одного з використовуваних критеріїв високими значеннями- мі за іншими критеріями з метою скорочення дисбалансу й можливості штучного маніпулювання.

Диференційованість оцінки – облік галузі науки або предметної області, життєвого циклу наукового колективу й наукових результатів, при агрегуванні й зіставленні, оскільки різні публікаційні кластери мають, що суттєво різняться статистикою і нормою числа співавторів, інтенсивності цитувань публікацій і т.п. [22]

Консервативність і збалансованість оцінки – ігнорування флуктуацій і нівелювання помилок оцінювання, орієнтація на збалансованість і стійкість публікаційної активності.

Інтегративність (адитивність) оцінки – можливість об'єднання індивідуальних інтегральних оцінок в узагальнену оцінку для характеристики наукових колективів різного масштабу.

Інституціональна обґрунтованість – змістовна відповідність застосовуваних критеріїв і граничних значень нормам права, державним стратегічним орієнтирам, ліцензійним, акредитаційним критеріям і параметрам моніторингу ефективності закладів вищої освіти МОНУ [22].

Комплексність оцінки – вичерпне охоплення наукометричних показників і джерел наукометричних даних для усесторонньої оцінки. Зокрема, МОНУ вважає визнаними міжнародними реферативними системами Web of Science, Scopus, Springer, Agris, Georef, Mathscinet, Bioone, Compendex.

Використання наукометричної оцінки для завдань керування виходить на перший план ще один стратегічно важливий пип – забезпечення гнучкого використання для системи мотивації наукових співробітників. Усяка оцінка повинна задовольняти інформаційним потребам основних користувачів, а в цьому випадку основний користувач – керівництво наукового колективу або індивідуальний дослідник.

У рамках даної роботи доцільно зупинитися на розгляді найбільш актуальних і загальних методик наукометричної оцінки й приділити увагу їх короткій характеристиці в розрізі нижченаведених груп.

До першої групи слід віднести різноманітні модифікації канонічних індивідуальних показників, зокрема індексу Гірша [16] і імпаکت-фактора [11]. Їхній найбільш повний огляд представлений [24]. В указаній групі найбільшій уваги заслуговує модифікація індексу Гірша, яка на основі коригувальних коефіцієнтів враховує галузь науки й вік дослідника. Інша цікава методика [25] пропонує коректувати індекс Гірша з допомогою зважування числа цитат кожної публікації на її вік.

Також уваги заслуговує альтернативна дробова модифікація індексу Гірша, щ враховує кількість авторів цитованих статей, що й ілюструє в дробовій частині близькість автора до наступного щабля індексу Гірша. Аналогічний зміст

закладений у методику П. Батісти (Batista P.D.) і ін. [2] – пропонується розраховуватися доповнений індекс Гірша (h') за формулою

$$h' = h_2/N,$$

де N – число співавторів для даної публікації.

Показане, що такий індекс стійкий при порівнянні наукометричних результатів у різних галузях науки (з різною нормою цитування й співавторства).

В окрему підгрупу слід виділити ряд методик, що призначені не стільки для безпосередньої індивідуальної наукометричної оцінки, скільки для визначення її заможності. Яскравим прикладом є модифікація, індексу Гірша, що стійка до маніпулювання, тому що ігнорує викривлення в околиці базового значення Гірша (оскільки опирається на його теоретичне значення, отримане по кращому тренду).

До другої групи нами віднесені комплексні або групові показники зведені індекси, що представляють собою, показники на основі усереднення або згортки вихідних індивідуальних показників. У їхньому різноманітті можна виділити методику, у якій виділені групи показників продуктивності, авторитетності й результативності розраховуючи на одного автора.

Подібні методики носять широку застосовність і в оцінці ефективності публікаційної активності й у використанні для формування фонду стимулювання праці. Наприклад, показане застосування зважених наукометричних індексів для оцінки ефективності підготовки наукових кадрів в аспірантурі. Зпропонована інтеграція такого роду оцінки в інформаційну систему оцінки ефективності наукової школи.

Методика [18] для розрахунку показника публікаційної активності, а також віку вченого й середне-зваженого імпаکت-фактору журналів, у яких опубліковані й цитуються статті. Разом з тим, аналіз подібних методик і показників вказує, що, як правило, застосування того або іншого способу агрегування диктується лише перевагами дослідника й недостатньо обґрунтовується. Зокрема, широко застосовувана зважена сума значень критеріїв вимагає теоретичного

обґрунтування можливості підсумовування й вибору ваг. Для розглянутого завдання такого обґрунтування ні, отже, не можна бути впевненим, що підсумовування зважених значень наукометричних показників може дати задовільний результат.

Третя група поєднує кваліметричні методики на основі граничної агрегації, бальної й рангової оцінки, а також різноманітні методики ранжирування наборів індивідуальних показників для оцінки публікаційної активності й наукової продуктивності.

Використання кваліметричних (інтервальних, порядкових і абсолютних) шкал для інтерпретації наукометричних показників досить поширене, особливо в області виміру наукової праці і його стимулювання. Подібні методики бальної оцінки й ранжирування систем показників порівняно недавно [25] і ін. Наприклад, індивідуальний рейтинг оцінюється по ряду сукупних показників, у першу чергу, по ступені публікаційної активності й цитованості, а також по відношенню цитованості публікацій до всіх написаних. Найбільш цікавої із усіх методик у даній групі є методика граничного агрегування, доповнена багатомірним ранжуванням, у даній методиці пропонується побудова агрегованого рейтингу, за різними наукометричними критеріями. Показане, що при багатокритеріальним ранжируванні неможлива ситуація, коли низькі ранги за окремими показниками можуть бути «заретушовані» високими оцінками по іншим. Чим більше критеріїв використовується, тим більше статистично значимі відмінності між привласненими рангами.

2.2 Методи параметричної теорії систем

Для виявлення й забезпечення характеристик цілісності освіти нами також використовуються базові ідеї параметричної загальної теорії систем (ПТС). У якості вихідного емпіричного матеріалу при проектуванні цілісних характеристик

освітніх підсистем беруться не тільки дані про існування закономірності, що цікавить нас, в тій чи іншій конкретній підсистемі, а й дані, що відносяться до можливо більшого масиву систем, навіть у тому випадку, якщо в цих даних безпосередньо не розглядаються закономірності, що цікавлять нас, пов'язані із цілісністю. Такі закономірності будуть мати загальносистемний характер, якщо вихідна інформація буде виражатися за допомогою особливого типу властивостей — системних параметрів. Справа в тому, що існують такі властивості, які, співставні з будь-якою системою (однорідність або неоднорідність елементів системи, сталість числа елементів або їх мінливість тощо.). Властивості такого роду є системними параметрами. Наведені приклади системних параметрів ілюструють групу бінарних системних параметрів (параметр може приймати тільки два значення). При здійсненні кваліметрії освітніх систем, зокрема їх цілісності, важливого значення набувають наступні бінарні системні параметри: здатність системи самостійно відновлювати втрачені елементи (субстратна авторегенеративність і, відповідно, не-авторегенеративність) втрачені елементи відновлюються зовнішнім середовищем (субстратна зовнішня регенеративність) і, відповідно, зовнішня нерегенеративність, тощо. [17].

В наукових і освітніх системах з урахуванням їх відкритості й варіативності шляхів саморозвитку особливе значення приймають багатозначні й лінійні системні параметри. Якщо уявити собі значення того або іншого системного параметра у вигляді точок на прямої лінії, то бінарний системний параметр характеризується двома точками на цій лінії, багатозначний системний параметр — кінцевою безліччю точок, а лінійний системний параметр символізує вся лінія з її нескінченною безліччю точок. У якості лінійних системних параметрів у педагогіці можуть бути використані "простота" і, відповідно, "складність" системи. У зв'язку з цим ми зіткнулись з необхідністю проаналізувати інформаційну значимість системних і несистемних параметрів. Основою для цього є формула К. Шеннона, відповідно до якої інформація, що полягає в повідомленні, чисельно дорівнює знятої ентропії:

$$E = - \sum_{i=1}^n p_i \log_2 p_i ,$$

де p_i — імовірність кожного з результатів експерименту, виміру, спостереження тощо.

Якщо піти за відомому прикладу Аристотеля, наприклад, запитувати про будь-яку систему, є вона видимою або невидимою, то на це питання можливі три відповіді; вона видима (ймовірність p_1); вона невидима (ймовірність p_2); до неї взагалі не має сенсу застосовувати поняття (ймовірність p_3). Оскільки число систем нескінченно, а властивість невидимості-видимості можна застосовувати лише до кінцевого, порівняно невеликому класу живих істот, то для будь-якої навмання взятої системи ймовірність того, що їй властива перша властивість p_1 а 0, друге p_2 , а третє $p_3 \ll 1$. Оскільки $\log_j 0 = 0$, то

$$E = - P_1 \log p_1 - P_2 \log p_2 - P_3 \log p_3 \quad \% .$$

Звідси сумарне значення інформації про те, чи мають всі системи зазначені властивості, буде порівняно невелике. Максимальне інформаційне значення будуть мати відомості про характеристики, які має сенс відносити до будь-якої системи, тобто про такі характеристики, які притаманні системі з ймовірностями, що значно відрізняються від нуля.

2.3 Методи інтеграції окремих знань

Поняття «картина світу» досить виразне й говорить саме за себе: воно означає як би зримий портрет світобудови, образно-понятійну копію Всесвіту. З поняттям "картина світу" пов'язане поняття "світогляд", яке підкреслює суб'єктивний ракурс розгляду світу: цілі, цінності, переваги. До світогляду веде не

стільки шлях індуктивного узагальнення експериментальних і природничо-наукових теоретичних даних, скільки характер самих світоглядних проблем – їх загальність (для індивіда, соціальної групи). Особисте знання поставляє лише матеріал, що переосмислюється з метою вирішення інтегральних проблем.

Під світом найчастіше розуміється дійсність, узята як ціле, схоплена в деякій її якісній єдності. «Будь-який дослідник, навіть якщо він не віддає в цьому звіті, прагне створити якусь системну конструкцію з логічно зв'язаними ланками, яку я називаю "картиною світу"» [17]. Однією з таких гіпотетичних системних конструкцій є припущення про фундаментальність законів складання цілісних систем: існують закони кооперативної (резонансної) взаємодії поряд із законами взаємодії частин. До них належить принцип "мінімуму дисипації": якщо, наприклад, закони збереження допускають кілька рівноважних станів (рішень), то реалізується те, якому відповідає мінімальне зростання ентропії, тобто більший порядок у структурі. В 1974 р. І.Д'ярмати показав, що принцип Онзагера і його модифікація, є відомим варіаційним принципом або «принципом різноманітного розвитку».

Коли проблему наукової картини світу обговорюють натуралісти, мова може йти насамперед про фізичну реальність, систему фундаментальних фізичних конструктів, що характеризують основні властивості універсуму: простір, час, речовина, поле. До надстійких елементів будь-якої картини світу відносять принцип збереження енергії, принцип постійного зростання ентропії, фундаментальні фізичні константи.

Еволюція сучасної картини світу передбачає рух від класичної до некласичної й постнекласичної картини світу, до зміцнення парадигми цілісності. Класична картина світу заснована на досягненнях Галілея й Ньютонів. Їй відповідає графічний образ прогресивно спрямованого лінійного розвитку з жорстко зазначеною детермінацією. Класична картина світу здійснювала опис об'єктів як якби вони існували самі по собі в суворо заданій системі координат. Основною умовою ставала вимога елімінації позбавлених значимості характеристик об'єкта пізнання й усього того, що ставалося до атомарних

(окремих) подій. Класична картина світу зміцнювала претензії на виявлення якогось загального правила або єдиного вірного методу, що гарантує побудову дійсної теорії (такий підхід називають «лівим інтеграційним монізмом»). Некласична картина світу, яка прийшла на зміну класичної, народилася під впливом перших теорій термодинаміки, що заперечують універсальність законів класичної механіки. Графічна модель неklasичної картини світу спирається на образ синусоїди, що описує магістральну пряму розвитку. Чим більше відхилення, тем менше воно ймовірніше, тому що щораз реальне явище наближається до генеральної лінії – "закону середнього". Відсутність детермінованості на рівні індивідів поєднується з детермінованістю на рівні системи (статистична закономірність).

Постнеklasична картина світу у вигляді свого графічного образу має "дерево, що гілкується" (концепція брюссельської школи І.Пригожина). У сучасній постнеklasичній картині світу аналіз суспільних структур передбачає дослідження відкритих систем, у яких велика роль умов, що входять у системи індивідів, локальних змін і випадкових факторів. Припис сучасної постнеklasичної картини світу містять у собі можливість коригування, розлому, перетворення, диференціації, інтеграції, породження "іншого", новацій. До методологічних новацій постнеklasичної картини світу можуть бути віднесені наступні факти:

– зміна характеру об'єкта дослідження - вони все частіше саморозвивають відкриті складні «людино розмірні» системи, які подтребують кваліметричних підходів в визначенні інтегративного вектора;

– посиленна роль міждисциплінарних, комплексних програм у вивченні відкритих систем, що саморозвиваються, це підкреслює актуальність кваліметрії цілісності в сучасному;

– трансформація "ідеалу" ціннісно-нейтрального дослідження: об'єктивно дійсний опис і пояснення стосовно до людинорозмірних об'єктів не тільки допускають, але й припускають включення й аксіологічних, і кваліметричних факторів до складу пояснювальних положень;

– зміцнення "парадигми цілісності", тобто усвідомлення необхідності своєчасної "закладки" фундаменту цілісності в меті, цінності, змісті і оціночній технології освіти;

– конвергенція (зближення) природних і соціальних наук, раціональних і ірраціональних підходів, когнітивних і ціннісних параметрів знання;

– твердження методологічного плюралізму, поліпарадигмального підходу;

– підвищення ролі «розуміючих кваліметричних методик» (із включенням апарата герменевтики), «особистісних оцінних методів» (наприклад, біографічного й автобіографічного), ціннісного та інформаційного підходів, методу соціально-гуманітарних експертиз, рольових та імітаційних ігор, емпатії при оцінці й самооцінці в освіті й ін.;

– впровадження фактора часу в усі наукові дисципліни; "імпортація" і "діалектизація" науки, що дозволяють з єдиних позицій історизму й самоорганізації матерії охопити як єдине ціле макро- і мікросвіти (в зв'язку з цим в сучасній науці знову зростає роль діалектичного методу, який був незаслужено скомпрометований у роки тоталітарного режиму; говорячи про користь діалектичної точки зору, не слід приписувати їй зайвих переваг, перетворювати цей, усього лише один з можливих способів мислення, у довільну догматичну умоглядну схему, і тоді в рамках своєї компетенції, особливо при аналізі розвитку цілісних систем, діалектичний метод може працювати досить ефективно в єдності з багатьма іншими способами й прийомами наукового пізнання);

– з'єднання об'єктивного світу й світу людини, руйнування твердої дихотомії природних і соціальних наук, зближення й взаємодія їх методів, зростання антропоекологічного принципу;

– адекватна оцінка значимості методологічних знань в області міждисциплінарної інтеграції й кваліметрії цілісності навчального знання;

– пошук законів еволюції відкритих нерівноважних систем будь-якої природи, включаючи педагогічні, що передбачають наявність "м'якого" (відповідно до потенційних можливостей самої системи) управління процесом саморозвитку. Практично всі методологічні новації постнекласичного етапу

розвитку науки тією чи іншою мірою містять ідеї інтеграції, які, при включенні їх у механізми кваліметрії, можуть стати стабілізуючим фактором, протидією дезінтегруючих процесів диференціації й варіатизації наукового й навчального знання.

Нове постнекласичне світобачення значною мірою пов'язане із впровадженням в усі приватні науки й наукові дисципліни ідей і методів синергетики – теорії самоорганізації. Принципи самоорганізації, досліджувані цією наукою, поширюються "від морфогенезу в біології деяких аспектів функціонування мозку до флатера крила літака, від молекулярної фізики до космічних масштабів еволюції зірок, від м'язового скорочення до спучування конструкції" [27]. Виникнення синергетики, можливо, знаменує початок нової наукової революції. Вона не просто вводить нову систему понять, але змінює стратегію наукового пізнання, сприяє виробленню принципово нової наукової картини світу. Не випадково передмову до англійського видання книги "Порядок з хаосу" І.Пригожин публікує під заголовком "Новий діалог людини з природою". Важливі філософсько-методологічне, світоглядне, інтегративне значення мають ключові ідеї синергетики. Сьогодні в наукових дослідженнях домінує таке поняття, як "дисипативні структури" (що виникають в "порушених" нерівноважних системах і чудові тим, що між їхніми елементами спостерігаються кореляції, когерентна взаємодія). Останнє пояснює й назва дисципліни. Термін "синергетика" утворений від грецького "sunergos" – разом діючий. Саме співробітництво або когерентна поведінка елементів дисипативних структур і є тим феноменом, який характеризує процеси самоорганізації, зокрема у педагогічних системах.

Феномени самоорганізації, нелінійності, глобальної еволюції неодноразово виступали в якості предмета обговорень [23]. Широке поширення отримали уявлення про становлення порядку через хаос, біфуркаційних змінах, незворотності часу, нестійкості як фундаментальної характеристики еволюційних процесів завдяки опублікованим з Брюссельського Вільного Університету, Шпрінгерівська серія книг з синергетики, у рамках якої за загальною редакцією

Г.Хакена починаючи з 1979 р. видано вже більше 50 томів по самим різним аспектам динаміки самоорганізації в природних, соціальних і когнітивних системах.

Синергетика руйнує багато наших звичних уявлень. Картина світу, малюється класичним розумом, – що світ з жорсткими причинно – наслідковими зв'язками. При цьому причинні ланцюги мають лінійний характер, а наслідок якщо не тотожні причині, то принаймні пропорційно їй. Класичний, традиційний підхід до управління складними педагогічними системами, оцінка ефективності їх функціонування ґрунтувалась на уявленні, згідно з яким результат зовнішнього керуючого впливу є однозначний і лінійний, передбачуваний результат прикладених зусиль відповідає схемі: керуючий вплив - бажаний результат (чим більше вкладаєш енергії, тим більше нібито й віддача). Однак практика, особливо інноваційних освітніх установ Європи показує, що багато зусиль виявляються пошуками, "що йдуть у пісок" або навіть шкодять, якщо вони протистоять власним тенденціям саморозвитку складно організованих науково-педагогічній системі. Одночасно ставиться під сумнів один з постулатів лінійного мислення – уявлення про те, що процеси бурхливого росту (ріст якості знань, конвергенція когнітивної, аксіологічної і акмеологічної складових освіти, розвиненість структури метаосвітньої області і т.д.) відбуваються по експоненті.

Синергетика пропонує іншу модель більшості процесів лавиноподібного зростання – в режимі із загостренням, коли розглянуті величини хоча б частина часу змінюється за законом необмеженого зростання за кінцевий час. Синергетика вчить:

– складноорганізованим системам не можна нав'язувати шляхи їх розвитку, а скоріше треба зрозуміти, як сприяти їх власним тенденціям розвитку, як виводити, зокрема, педагогічні системи на ці шляхи; стає зрозуміло, яким чином і чому хаос може виступати в якості початку конструктивного механізму еволюції, як з хаосу власними силами може розвинути нову організація;

– у відповідні моменти – моменти нестійкості – малі збурення, флуктуації (в освітніх системах – при м'якому управлінні, коли воно втрачає характер сліпого

втручання методом проб і помилок або ж впертої зміни реальності, небезпечних дій проти власних тенденцій системи, і будується на основі знання того, що взагалі можливо в даному середовищі) можуть розростатися в макроструктури, тобто в особливих станах нестійкості соціального середовища дії кожної окремої людини можуть впливати на макросоціальні процеси;

– для складних систем, як правило, існує кілька альтернативних шляхів розвитку (вибір здійснюється в так званих точках біфуркації), зміцнюється надія на можливість вибору таких шляхів подальшого розвитку, які влаштовували людину й не були руйнівними для природи й соціуму (відповідно до прийдешнього порядку);

– відкриваються нові принципи суперпозиції, складання еволюційного цілого з частин, побудови складних структур, що розвиваються, з простих: об'єднання структур не зводиться до їх простого складання (інтерференції), а має місце дефект енергії, ціле при цьому вже не дорівнює сумі частин, воно не більше й не менше суми частин, воно інше якісно;

– з'являється новий принцип узгодження частин у ціле: встановлення загального темпу розвитку вхідних у ціле частин (співіснування структур різного віку в одному темпосвіті); виробляється новий підхід до побудови складних соціальних, освітніх цілісностей;

– при управлінні складними системами головне – не сила, а правильна "топологічна конфігурація, архітектура впливу на систему або середовище; малі (м'які), але правильно організовані – резонансні впливи на складні системи надзвичайно ефективні" [18];

– відкриваються нові закономірності й умови протікання процесів нелінійного самостимулюючого росту, ініціювання такого роду процесів у відкритих нелінійних середовищах (наприклад, в освітньому середовищі), і які існують вимоги, що дозволяють уникати імовірного розпаду складних структур поблизу моментів максимального розвитку.

Синергетичне світорозуміння ми зв'язуємо з образом відкритої освітньої системи, здатної до самоорганізації, процеси обміну в якій відбуваються не тільки

через границі, але й у кожній точці системи. Одночасно відкритість системи не є гарантія її самоорганізації й саморозвитку. Рівнозначний чинник самоорганізації й саморозвитку – позитивне співвідношення двох початків: то, що створює структури, і то, що розсіює їх. Із джерела: "... початок, що розсіює (дисипація) у відкритій системі може пересилювати, переборювати роботу конструктивного джерела, розмивати всі неоднорідності, створювані їм. У такому режимі структури не можуть виникати" [18]. У той же час стосовно до такого консервативного соціального інституту як освіта роль дисипації важко переоцінити. В освітніх системах (нелінійних і відкритих) дисипація відіграє роль "різця, яким скульптор поступово, але цілеспрямовано відсікає все зайве від кам'яної брили. А оскільки дисипативні процеси, розсіювання є, по суті справи, макроскопічний прояв хаосу, оскільки хаос на мікрорівні – це не фактор руйнування, а сила, що виводить на атрактор, тенденцію самоструктурування нелінійного середовища" [18] (атрактор – та реальна структура у відкритому нелінійному середовищі, на яке виходить процес розвитку в цьому середовищі в результаті загасання в ній перехідних процесів).

Фундаментальний концептуальний вузол нової картини світу – принцип «Нелінійність». Нелінійність в освітніх середовищах зазвичай зв'язується насамперед з необхідністю й можливістю опису процесів рівняннями, коефіцієнти яких залежать від властивостей середовища. Різні шляхи розвитку освітніх систем слід зв'язувати з біфуркаціями (розгалуженням, поділом) при зміні констант середовища, зміні керуючих параметрів. При цьому не зайве згадати відомий діалектичний закон «переходу кількості в якість», оскільки родоначальники синергетики, на наш погляд, в суворій відповідності із цим законом виділяють певний діапазон змін у середовищі і його параметрах, при якому не відбувається якісних змін картини процесу. Але якщо долається деяка гранична зміна, перевищується критичне значення параметрів, то режим розвитку системи якісно змінюється: вона потрапляє в область притягання іншого атрактора.

2.4 Методи кваліметричного аналізу наукової діяльності

Принцип невизначеності в науковій і освітній практиці в певній мері корениться у відносності наукового знання. Так, при здійсненні педагогічної кваліметрії необхідно враховувати той факт, що "... наукове знання (поняття, ідеї, концепції, моделі, теорії, висновки з них і т.п.) завжди відносно й обмежене. Задача вченого - встановити границі відповідності знання дійсності – інтервальної адекватності" [27], у межах якого знання є й залишається цілісним. Один з істотних ознак відносності знань взагалі й педагогічних, зокрема, впливає з її підтвердження експериментом, усе частіше заснованому на педагогічній кваліметрії, а вимірів абсолютно точних не буває, і в цьому зв'язку задача вчених і практиків - вказувати інтервал неточності. Істотна невизначеність результатів досліджень у педагогіці може бути знижена підвищенням точності вимірів шляхом модернізації вимірювальних процедур і доповнення кількісних описів результатів якісними.

Але невизначеність у пізнанні не обов'язково пов'язана з недосконалістю методів виміру. Наприклад, при вивченні у фізику властивостей мікрочастинок необхідно враховувати їхню двоїсту корпускулярно-хвильову природу. Виникає необхідність введення деяких обмежень у застосуванні до об'єктів мікросвіту понять класичної механіки. Не можна говорити про рух мікрочастинки по певній траєкторії й про одночасно точні значення її координати й імпульсу (що цілком можливо в класичній механіці). Це означає, що, якщо мікрочастинка перебуває в стані з точним значенням координати, то її імпульс є повністю невизначеним. Німецький фізик В.Гейзенберг, враховуючи хвильові властивості мікрочастинок і пов'язані із хвильовими властивостями обмеження в їхній поведінці, прийшов в 1927 р. до висновку, що одержав згодом назву "співвідношення невизначеностей Гейзенберга". Може бути записаним для координати й імпульсу мікрочастинки, це співвідношення виглядає в такий спосіб:

$$D_x D_p > h.$$

де h – постійна Планка,

D_x – невизначеність координати,

D_p – невизначеність імпульсу мікрочастинки.

Таким чином, у загальному (аналогічному) випадку добуток невизначеностей двох характеристик, параметрів не може бути менше деякої постійної величини.

Із принципом невизначеності нерозривно зв'язаний принцип додатковості. Він також був уперше сформульований Н.Бором для опису мікрооб'єктів: "...одержання експериментальної інформації про одні фізичні величини, що описують мікрооб'єкт (елементарну частку, атом, молекулу), неминуче пов'язане із втратою інформації про деяких інших величинах, додаткових до перших" [29] (що неминуче приводить до ускладнень при формуванні цілісного знання). Такими взаємно додатковими величинами можна вважати, наприклад, координату частки і її швидкість (або імпульс). У загальному випадку, який застосовується й до педагогіки, додатковими один до одного будемо вважати величини, яким відповідають оператори (формули для обчислень, дефініції), що не містять тих самих ключових слів, тобто не комутуючі між собою. Стан, у якого додаткові величини мали б одночасно точно певні значення, принципово неможливо, причому якщо одна з них точно визначена, то значення іншої повністю невиразно; якщо точність визначення однієї зростає, то точність визначення іншої стільки ж або ще значніше зменшується (у суворій відповідності зі співвідношенням невизначеностей Гейзенберга, що і є, цілком ймовірно, причиною того, що багато авторів обидва розглянуті принципи поєднують в один).

У контексті кваліфікаційної роботи важливий приклад додаткових категорій «один до одного», є динамічний і статистичний закони. Закони, які управляють поведінкою більших сукупностей і відносно індивідуального об'єкта, що дозволяють робити лише неоднозначні висновки про його поведінку, є ймовірно-статистичними. Вони характеризують поведінку не одного елемента в колективі, а поведінку колективу в цілому. Знання статистичної закономірності

(тенденції) дуже важливо при педагогічному проектуванні на всіх рівнях, але воно не дозволяє однозначно пророкувати поведінку окремих індивідуальних об'єктів, що входять у колектив. Відносно окремих елементів такі пророкування мають тільки імовірнісний характер. У фізиці є тому прекрасний приклад: максвелівський закон розподілу молекул газу по швидкостях нічого не говорить певного про швидкість кожної молекули окремо в певний момент – він лише встановлює частку молекул, які в цей момент мають цілком певну швидкість. Але класична наука завжди мала справу тільки з динамічними законами, абсолютизувала їх і вважалося, що статистичні закони виникають як результат неповноти нашого знання. Сьогодні ми не протиставляємо динамічні й статистичні закони з погляду освітнього статусу, «...вони не виключають один одного й не є по відношенню друг до друга гносеологічними наближеннями – вони доповнюють один одного, дозволяючи скласти більш багату виставу про різноманітність відносин детермінації, що існують в об'єктивному світі» [12].

З позицій методології кваліметрії цілісності освіти унікальною є ще одна пара додаткових друг до друга категорій: випадковість і необхідність. Можна виділити кілька моментів, що характеризують взаємозв'язок необхідності й випадковості. Нас найбільше цікавить розуміння випадковості як доповнення необхідності. Наше пізнання націлене в основному на розкриття істотних і необхідних зв'язків предмета вивчення, на пізнання загального, цілого, того, що характеризує цілий клас однорідних предметів і явищ. Але кожний предмет вивчення поряд із загальними має й індивідуальні риси, виникнення й прояв яких визначається випадковістю. Повне цілісне знання про той або інший об'єкт буде досягнуто тільки тоді, коли знання необхідних рис предмета буде доповнено знанням випадкових, унікальних рис цього явища. Ця вимога доповнює знання необхідними знаннями випадковості особливо важливо для таких наук, як історія, біологія, психологія, педагогіка.

На рівні фізичного або біологічного пізнання фізичний або біологічний об'єкт завжди суть комплексно диференційованих елементів, кожен з яких характеризується досить диференційованими функціями [23]. Відповідно, задача пізнання такого об'єкта полягає у виявленні законів взаємодії елементів з метою

пояснення природи його цілісності й, як наслідок, його властивостей, У контексті такого розуміння цілісності у свій час був сформульований Л.С.Виготським принцип психологічного аналізу по "одиницях" ("клітинкам"). Л.С.Виготський писав, що під одиницею він розуміє такий продукт аналізу, який, на відміну від елементів, має всі основні властивості, властиві цілому, і який є далі нерозкладною живою частиною цього цілого. Слід зазначити, що, виділивши в якості "одиниці" вивчення мислення "значення слова" у процесі подальшого дослідження Л.С.Виготський, проте, порушив питання про особливості його структурної організації, розклавши по компонентам, таким чином, "одиницю" на складові її елементи (рівні різному ступеня узагальненості) в описаному їм психологічному феномені "захід спільності понять" [70].

Дослідження показує, що структурно-інтегративна методологія дозволяє аналізувати природу цілісності психічного явища в рамках принципу субаддитивності ("ціле менше своїх частин"), який органічно доповнює принцип супераддитивності [30], і самі частини мають певну специфіку, природа якої накладає істотні обмеження на властивості цілого. Таким чином, ціле менше своїх частин у тому розумінні, що ціле виявляється в певній мірі залежним як від природи частин (елементів), так і від характеру їх взаємозв'язку. Ми переконалися, що методологічна установка, яка у свій час одержала назву «елементаризм» (шлях від складного до простого), на сучасному (системному) етапі розвитку наукового знання придбала зовсім інший зміст, виступаючи, скоріше, уже у вигляді інтегратизма (В.О.Енгельгардт) [30]: шлях наукової думки від простого до складного, спрямований від аналізу природи окремих компонентів складного об'єкта і інтегративного процесу їх співорганізації до розуміння природи внутрішньо єдиного цілого, що володіє якісно новими властивостями. Стає зрозуміло; категорія інтеграції передбачає, що в частин, з яких "зібране" ціле, є специфічні властивості, що забезпечують можливість виникнення між ними певних зв'язків. Ці властивості, згідно В.О.Енгельгардту, можна назвати десмосними (від грецького слова "десмос" – зв'язок). Ефект інтеграції проявляється, з одного боку, у тому, що частина, що входить до складу

нового, більш складного цілого, втрачає деяку частку властивостей або вони трансформуються й, з іншого боку, у тому, що в самій новій цілісності з'являються нові властивості, породжувані, головним чином, тими зв'язками, які виникли при входженні частин у це ціле, що утворювалося (Енгельгардт, 1970). Таким чином, інтегративний процес, що йде " усередині" структури складного об'єкта, є механізмом його існування й умовою появи в нього ряду якісно нових (системних) властивостей.

Сенсоутворювальним положенням структурно-інтегративної методології є визнання провідної ролі структурних характеристик об'єкта щодо тих конкретних властивостей, які вони виявляють у тих або інших умовах. У природничих науках ідея про те, що властивості (функції) об'єкта виявляються похідними стосовно закономірностей його структурної організації, була прийнята досить давно. Зокрема, визнавалося, що знайти закон існування того або іншого об'єкта - значить розкрити структуру даного об'єкта, тому що структура є основою його функціонування. Складніші справи в соціальних науках. Стосовно до вивчення структурної організації розумової діяльності ми пропонуємо облік наступних аспектів аналізу.

У перерахованих вище аспектах аналізу розумової діяльності кілька разів пролунало поняття "інтелект". Оскільки нам ще не один раз буде необхідно до нього вдатися, "розшифруємо" це поняття. "Інтелект – форма організації індивідуального ментального досвіду Ментальний досвід – система індивідуальних інтелектуальних ресурсів, що обумовлює особливості пізнавального відношення суб'єкта до світу" [312, 3362]. І оскільки пізнавальне відношення суб'єкта до світу безпосередньо пов'язане з його інтелектом, з'являється "шанс" з'ясувати психологічні джерела інтегративних процесів в освіті.

Знайомство з дослідженнями, що проводилися в останні десятиліття в різних областях закордонної й вітчизняної психології (психології пізнання, психології особистості, соціальної й інженерної психології), свідчить про зростання інтересу до проблеми структурної організації розумової сфери

людини. Поступово формується уявлення про те, що для розуміння механізмів психологічної (у тому числі інтелектуальної) зрілості важливо не тільки те, що суб'єкт відтворює у своїй свідомості в процесі пізнавального відображення, але й те, як він осмислює, що відбувається. У такого роду когнітивно орієнтованих психологічних підходах поступово накопичувався емпіричний матеріал, для опису якого доводилося прибгати до використання таких категорій, як "когнітивна структура"; "структурні властивості когнітивної системи"; "ментальний простір" суб'єктивний діапазон відображення, у рамках якого можливі різного роду структурні зміни, переміщення); "конструкти" – суб'єктивні оцінні шкали, представлені кожна у вигляді ознаки і його антоніма й визначальні в остаточному підсумку (в залежності від їхнього числа у індивіда й характеру відносин між ними) яку модель дійсності будує людей - спрощену (лінійну) або багатомірну (цілісну); "когнітивна складність"- поняття, що йде коріннями в теорію особистісних конструктів і обумовлене як "здатність конструювати соціальну поведінку на основі численних параметрів" [25]. Когнітивна складність асоціюється, в основному, з диференційованістю, але в деяких авторів можна знайти "розширення" поняття за рахунок залучення ієрархічної структури когнітивної системи – когнітивної інтегрованості [26].

Дотримуючись принципів невизначеності й додатковості, І.С.Алексєєв прийшов до твердження існування так званих квантових дій. У концепції додатковості оголюється діяльна структура знання. В останній присутня дослідницька програма, яка визначає магістральний шлях майбутнього розвитку науки.

З початку 80-х років спостерігається розвиток уявлень про постнекласичний тип раціональності (один з аспектів аналізу цієї проблеми описується в першому розділі). На передній план вийшли проблеми соціокультурної обумовленості наукового пізнання, аналіз взаємодії науки з іншими феноменами людської культури, дослідження пізнавальних процедур у зв'язку з історично мінливими цінностями й світоглядними орієнтаціями. Об'єкт дослідження все частіше становився унікальною системою, що саморозвивається, включають людину в

якості особливого компонента (наприклад, освітні системи). При вивченні "людинорозмірних систем" пошук істини виявляється пов'язаним з визначенням стратегії й можливих напрямків перетворення систем, у першу чергу - їх структур, тобто відносин між елементами. Останнє в педагогіці торкається безпосередньо гуманістичних цінностей. Робимо висновок: об'єктивне пояснення й опис стосовно до "людинорозмірних" об'єктів не тільки допускає, але й припускає включення аксіологічних факторів до складу пояснюючих положень. Стає необхідною своєрідне стикування (інтеграція) специфічних для науки внутрішніх ціннісних установок (на пошук предметного й об'єктивного знання, цінність новизни) із цінністю загальносоціального характеру. Конкретним механізмом такої інтеграції служать соціально-гуманітарна й екологічна експертиза великих науково-технічних програм, коли простежуються можливі наслідки реалізації програми під кутом зору гуманістичних цінностей.

Таким чином, якщо на етапі некласичної методології науки увага центрувалася на об'єктних структурах діяльності (засоби, операції з об'єктом і ін.), те в постнекласичній методології потрібно, крім цього, урахувувати особливості суб'єктних структур діяльності в їхньому історичному розвитку (особливості суб'єкт – суб'єктних відносин, цілей і цінностей діяльності, їх співвідношення з домінуючими цінностями культури певного історичного типу). Інтегративна діяльнісна структура – особливий тип цілісності. Стосовно до вивчення структури особистості виділяють три типи цілісності:

- дифузійна цілісність, для якої характерна глобальна, недиференційована активність;

- диференційована цілісність, що відрізняється видільної складових її частин, кожна з яких функціонує більш-менш автономно;

- інтегрована цілісність, яка проявляється, коли диференційовані частини виявляються в стані стабільної, різноспрямованої взаємозалежності. При аналізі інтегративної структури діяльності в цілісній освіті слід говорити про третій тип цілісності й відповідно будувати стратегію його розвитку.

3 ОПИС РОЗРОБЛЕНИХ АЛГОРИТМІВ

3.1 Алгоритми побудови рейтингу публікаційної активності

Системне і операціоналізоване вирішення завдання аналізу публікаційної активності й оцінки індивідуального внеску в обмін науковими результатами для цілей керування науковою діяльністю в цілому вимагає обліку базових принципів у рамках науково-технічної політики наукової організації або університету.

Насамперед, слід забезпечити диференційованість оцінки. Очевидно, що для рішення даного завдання в розрізі різних наукових напрямків або афілійованих організацій досить робити вибірку (наприклад, по ключових словах, приналежності до організації й ін.).

Важливим елементом диференційованого підходу, за нашою думкою, є облік тривалості публікаційної активності. Дійсно, зіставлення наукометричних показників дослідника, що працює у своїй області протягом багатьох років і дослідника, що нещодавно захистив дисертацію, є некоректним. Облік даного зауваження звичайно виражається у введенні поправочного коефіцієнта пропорційного тривалості публікаційної активності або віку дослідника. Вважається достатнім зважування за допомогою розподілу на тривалість публікаційної активності. У даному випадку може вирішуватися й інше важливе завдання – стимулювання молодих вчених. Для наукового колективу, якщо буде потреба, у якості оцінки тривалості публікаційної активності слід брати максимальну індивідуальну тривалість.

Важливим завданням є розробка підходу по забезпеченню консервативності оцінки. У ряді робіт у цьому зв'язку пропонується використовувати функції з убутної похідної. Забезпечення убутного вкладу обумовлене необхідністю обмежити спекуляцію наукометричними показниками й урахувати деяку специфіку їх динаміки (наприклад, кількість цитувань статті лінійно залежить від числа співавторів [25], а динаміка індексу Гірша апроксимується сигмоїдальною функцією). Зокрема, вибір подібної функції

(наприклад, $\sqrt[4]{N/n}$, де N – кількість цитат, n – кількість публікацій. Дозволяє зробити облік цитування більш консервативним, а внесок кожної наступної цитати помітно менше внеску попередньої. Іншим варіантом є використання логоріфмічного перетворення, при цьому логарифмування наукометричних даних також вирішує завдання наближення розподілу до нормального.

Типи комплексності і обґрунтованості оцінки можуть бути реалізовані використанням повних наборів даних по показниках із трьох основних, закріплених вимогами моніторингу наукометричних баз – Web of Science, Scopus. У залежності від стратегічних завдань по стимулюванню реферування в тієї або іншій базі даних, наборам показників для оцінки необхідно привласнити відповідні ваги (у випадку ранжирування – пріоритетні ранги).

Важливим критерієм вибору тих або інших наборів показників, на нашу думку, є можливість їх автоматизованого вилучення з доступних джерел і інтеграція наукометричної оцінки в інформаційну систему організації (університету).

Спираючись на описані вище підстави, нами пропонується оцінка публікаційної активності на основі формування агрегованого рейтингу. Основні етапи методики полягають у наступному (на прикладі Scopus).

Відбір наукометричних показників, що приймають участь у побудові рейтингу, виходячи із типів оцінки й управлінських завдань.

Формування масиву й вивантаження показників, загальнодоступних для перевірки в системі Scopus, у розрізі областей дослідження.

Коректування значень показників (зважування):

$$x' = x_i / (y_i + 1), \quad (3.1)$$

де x_i – вихідний показник по i -му авторові;

y_i – коректування по i -му авторові (наприклад, число років від першої публікації).

Коригування дозволяє одержати питомі, порівнянні по авторам, величини, а також дозволяє формувати певні стимули публікаційної активності, зокрема:

– віднесення кількості публікацій і індексу Гірша до років від першої публікації з однієї сторони обумовлює необхідність підтримки публікаційної активності на винному рівні, а з іншого – дозволяє стимулювати більш молодих авторів;

– віднесення загального числа цитат до публікацій ставить пріоритет на публікації малого числа великих цитованих робіт порівнюючи із численними маловпливовими роботами (попереджає дроблення наукових результатів на кілька публікацій);

– віднесення числа публікацій до співавторів з однієї сторони орієнтує на залучення продуктивних співавторів і участь у вагомих роботах, а з іншого – попереджає номінальне співавторство.

Перетворення значень показників у ранги в результаті відображення на рангову множину (сортування по убутанню):

$$\langle x' \rangle \rightarrow \langle 1, 2, \dots, N \rangle \quad (3.2)$$

ранг i -го автора визначений, як:

$$\text{sort} \downarrow x = \text{sort} \downarrow (x'), x \in 1, 2 \dots N \quad (3.3)$$

Ранжирування показників по значимості для оцінки індивідуальної публікаційної активності, потім граничне агрегування [33], яке, як раніше відзначене, забезпечує некомпенсаторність підсумкової оцінки.

На даному етапі слід розділити наукометричні показники на дві групи: до першої ставляться показники, які апіорі можна ранжувати по значимості (наприклад, входження в top-10 % більш значиме, аніж в top-25 % і т.п.), а до другої ставляться показники умовно рівної значимості.

Нехай n -мірний вектор рангів усіх n показників (атрибутів профілю автора в

системі Scopus) *i*-го автора, називаний альтернативою, що визначається, як:

$$a_i = \langle r^1, r^2, \dots, r^n \rangle, a_i \in A. \quad (3.4)$$

Тоді, для першої групи показників ранжирування впливає виробляти, виходячи зі значимості показника, тобто більш кращої є альтернатива, що має більший ранг по більш значимому показнику (наприклад, вектор з більшим рангом по показникові входження top-10 %, перебуває в ранжируванні вище, чим вектор з більшим рангом по показникові входження top-25% і т.п.).

Рішення завдання ранжирування наукометричних показників по значимості може опиратися на внутріорганізаційну науково-технічну політику, державні стратегічні орієнтири й т.п.

Коефіцієнт конкордації (W) для даного ранжирування рівний 28.8, що гарантує статистичну значимість ранжирування при $\alpha=0,05$. При цьому експерти також оцінили значимість джерел наукометричних даних (реферативних баз) для оцінки індивідуального внеску в обмін науковими результатами: на першому і другому місці Scopus і Web of Science, відповідно, третє Google Scholar.

Із числа експертів 100 % мають учений ступінь і 50 % мають учене звання. 30 % мають досвід керівництва науковим проектом державного значення, 60 % керували науковим проектом на рівні організації або підрозділу. 60 % експертів здійснюють фактичне й систематичне керівництво науковим колективом/проектом.

Загальне рішення може бути також знайдене на основі аналізу експертних оцінок.

Для цього має бути проведене анкетування експертів, у число яких увійшли керівники науковою працею різного рівня, що дозволить визначити погоджену порівняльну значимість наукометричних показників, доступних в Scopus і (табл. 3.1).

Таблиця 3.1 – Ранжування наукометричних показників порівняльної значимості

№	Найменування показника	Джерело
1	Кількість публікацій в top-10% журналів	Scopus
2	Кількість публікацій в top-25% журналів	Scopus
3	Кількість статей у закордонних журналах	Google Scholar
4	Кількість цитувань у закордонних журналах	Google Scholar
5	Імпакт-фактор журналів, у яких були опубліковані статті	Scopus, Google Scholar
6	Імпакт-фактор журналів, у яких були процитовані статті	Google Scholar
7	Кількість статей з журналів переліку ВАК	Google Scholar
8	Кількість цитувань із журналів переліку МОНУ	Google Scholar
9	Індекс Гірша	Scopus
10	Кількість публікацій автора	Scopus
11	Кількість цитувань автора	Scopus
12	Кількість співавторів	Scopus

У свою чергу, для другої групи показників необхідно застосувати граничне правило ранжирування, оскільки, відомо, що такий агрегований рейтинг буде задовольняти умові некомпенсаторності й Парето-оптимальності ранжирування [29].

Позначено через $v_m(a)$ – кількість рангів m у векторі a , $0 \leq v_m(a) \leq n$. Тоді альтернатива $a_x \in A$ вважається (строго) кращою, ніж альтернатива $a_y \in A$, якщо знайдеться такий ранг $1 \leq t \leq n$, що $v_m(a_x) = v_m(a_y)$ для всіх рангів $1 \leq m \leq t - 1$ і $v_t(a_x) > v_t(a_y)$.

Таким чином ранг результату ρ_i , i -го автора з s при багатомірному ранжируванні з урахуванням n наукометричних показників і умові $a_i > a_{i+1}$, $i \in 1, \dots, s$ вочевидь рівний $\rho_i = i$ (тобто вектор з більшим числом одиниць

розташовується в ранжируванні вище, ніж вектор s меншим числом одиниць, інакше співставляється кількість двійок і т.п.).

За описаним алгоритмом здійснюється відображення множини векторів альтернатив A на рангову множину: $A \rightarrow \langle 1, 2 \dots N \rangle$.

Фрагмент отриманого ранжирування на масиві експериментальних даних представлений на рисунку 3. Отримане в такий спосіб ранжирування також є статистично значимим при $\alpha=0,05$ [29].

Запропонована методика, включаючи алгоритм збору, предобробки й ранжирування наукометричних даних являє собою основу методичного інструментарію аналізу й оцінки публікаційної активності.

3.2 Алгоритм оцінки публікаційного потенціалу вченого

У загальному значенні, під публікаційним потенціалом автора слід розуміти його можливість по поліпшенню публікаційної активності й, зокрема, по підвищенню наукометричних показників без зниження якості. Для цілей справжньої роботи визначення індивідуального публікаційного потенціалу на основі публікаційного рейтингу, розглянутого вище, опирається на квартильний захід.

Квартилі – значення, які ділять таблицю даних (або її частину) на чотири групи, що містять приблизно рівну кількість кортежів (значень). Загальний обсяг поділяється на чотири рівні частини: 25 %, 50 %, 75%, 100 %.

У нашому випадку, належати до першого квартилю (або верхньому, $q1$) будуть автори, що ставляться до верхніх 25% місць у публікаційному рейтингу ($R_{25\%}$); до другого квартилю $q2$ – стосовно до наступних 25 % місць ($R_{50\%}$) і так далі.

Іншими словами, у першому квартилі будуть перебувати 25% авторів, що займають кращі місця в публікаційному рейтингу, а в четвертому квартилі ($q4$) –

25% що займають гірші місця, відповідно.

З обліком цього, публікаційний потенціал автора визначається нами, як прогнозна оцінка можливостей даного автора по переходу зі свого квартиля за рейтингом публікаційної активності до кращого квартилю. Необхідні умови для подібного переходу в тривіальному випадку являють собою такий приріст наукометричних показників автора, який би забезпечив істотний підйом у публікаційному рейтингу до місця, займаного останнім автором у кращому квартилі.

Оскільки при побудові публікаційного рейтингу перевага віддається більш молодим авторам (по року першої публікації), то визначення публікаційного потенціалу на його основі, зберігає дану особистість. Це, втім, узгодиться зі специфікою життєвого циклу: передбачається, що більш молодий автор розташовує більшим публікаційним потенціалом.

З метою оцінки публікаційного потенціалу нами пропонується адаптація алгоритму інтелектуального аналізу взаємозв'язків (пошуку асоціативних правил) «A priori» [11], модифікованого К. Боргельтом (С. Borgelt) [6] до наукометричних даних у системі Scopus. За допомогою даного алгоритму, як буде показано далі, можуть бути сформульовані деякі набори асоціативних правил, придатні для прогнозування ймовірних майбутніх публікаційних результатів автора.

Для подальшого розгляду введено деякі визначення. Атрибути профілю вченого (наприклад, індекс Гірша, кількість публікацій, список журналів і ін.): $A_i = \langle a_i; w_i \rangle$, де a_i назва атрибута, w_i вага атрибута, $1 \leq i \leq m$, $i \in \mathbb{N}$.

Профіль ученого:

$$P = \{ \langle A_1; v_1 \rangle, \langle A_2; v_2 \rangle, \dots, \langle A_m; v_m \rangle \}, v_i \in \mathbb{R}.$$

Рейтинг-аркуш – послідовність із P , упорядкована за зменшенням значення атрибутів, узятих відповідно до убутання їх ваг: $P = (P_1, P_2, \dots, P_n)$.

Квартильна функція профілю – ставить у відповідність заданому профілю вченого $P \in \mathbb{R}$ його квартиль Q по рейтинг-аркушу R :

$$Q: R \rightarrow Q, Q = \{Q_1, Q_2, Q_3, Q_4\}, \forall i, j: Q_i > Q_j \Leftrightarrow i < j.$$

Квартильна функція атрибута – ставить у відповідність заданому профілю вченого $P \in R$ його квартиль q по атрибуту A_i : $q: R \times N \rightarrow \mathbf{q}, \mathbf{q} = \{q_1, q_2, q_3, q_4\}$, $\forall i, j: q_i > q_j \Leftrightarrow i < j$.

Оцінка профілю вченого $\sum_{i=1}^n$ функція, що формує множину квартилів відповідних до заданого профілю $P \in R$ і всім його атрибутам: $v: R \rightarrow \mathbf{q} \cup Q$, $v(P) = \mathbf{q} \cup Q(P) \cup (P, i)$.

Множина оцінок усіх профілів учених:

$$V = U^n \rightarrow V_j; V_j = v(P_j).$$

Шаблон публікаційної активності – правило виду

ЯКЩО ant TO cons,

для якого виконуються наступні умови:

$$\text{ant } \mathbf{q} \subseteq \Lambda \text{ ant} \neq \emptyset \wedge \text{cons} \in Q \wedge \text{cons} \neq \emptyset;$$

$$\forall V_j (\text{ant} \subseteq V_j \Rightarrow \text{cons} \in V_j);$$

$$\sigma \geq \text{minsupp} \wedge \delta \text{ minconf}.$$

де σ – умовна ймовірність події, у якій оцінка профілю що містить cons також містить ant = $P(\text{cons}|\text{ant})$.

Достовірність шаблону δ – відношення ймовірності події в якому оцінка профілю містить ant та cons, до ймовірності події в якому оцінка профілю містить ant :

$$\delta = P(\text{ant} \cup \text{cons}) / P(\text{ant}).$$

Множина всіх шаблонів публікаційної активності: $A\mathcal{R}$ Множина шаблонів публікаційної активності, що підходять профілю вченого

$$P \in R : ar(P) = \{T \mid ant \subseteq V(P), T \in A\mathcal{R}\}.$$

Публікаційний потенціал ученого із профілем:

$$P \in R : pt) = \max \text{cons}(P).$$

Таким чином, відповідно до передумов застосовуваного метода інтелектуального аналізу, методика оцінки публікаційного потенціалу й алгоритм виявлення авторів, що мають такий потенціал, будуть полягати в наступному.

Крок 1 – для заданого набору авторів будується публікаційний рейтинг на основі запропонованої методики оцінки публікаційної активності.

Крок 2. На основі отриманого публікаційного рейтингу набір авторів розбивається на чотири квартиля. Перший квартиль (перші 25% авторів) інтерпретується як кращий на основі оцінки публікаційної активності, другий квартиль – як гарний і т.д. Кожний автор здобуває синтетичний атрибут, що містить мітку приналежності до квартилю.

Крок 3. По кожному атрибуту профілю (наукометричному показнику), що приймає участь в побудові рейтингу проводиться квартильна дискретизація аналогічна попередньому пункту, тобто формується набір синтетичних атрибутів, що представляють собою мітки виду «attrib_QX», де *attrib* – ім'я атрибута, а «X» – ознака квартиля, наприклад, «*h_index_Q1*». Крім того, значення атрибута еквівалентні мінімуму (частіше за все для наукометричних показників це нуль) свідомо здобувають мітку «Q4». Дане корекційне правило приводить до викривлення квартильної структури по деяких атрибутах, але забезпечує вірну інтерпретацію наукометричного профілю автора.

Крок 4. Для отриманих наборів синтетичних атрибутів по кожному автору,

що представляють собою транзакцію, виконується пошук асоціативних правил. Потім проводиться вибірка стійких правил, що виявляють перехід авторів із власних квартилей у кращі по поєднанню наукометричних атрибутів.

Крок 5. На основі отриманих правил виявляється набір авторів їм відповідних, тобто, що мають потенціал перехід до кращого квартилю рейтингу публікаційної активності. При цьому прогнозним заходом публікаційного потенціалу автора є захід вірогідності підходящого стійкого правила.

Даний алгоритм може бути легко реалізований в рамках інформаційної системи науково-освітньої організації й використана для завдань керування науковою діяльністю. Повна автоматизація може бути досягнута інтеграцією модуля автоматичного збору (вивантаження) наукометричних даних [27] з використовуваних реферативних баз.

На закінчення розглядається алгоритм формування індивідуальних рекомендацій. Під рекомендацією розуміється комбінація з набору альтернатив по поліпшенню показників публікаційної активності до певного рівня. Фактично, рекомендація для автора представляє собою тезову інтерпретацію результатів наукометричного аналізу (і зокрема, оцінки публікаційного потенціалу).

Формалізовані визначення виглядають у такий спосіб: Індивідуальна рекомендація – для даного профілю $P_j \in R$ і його цільового квартиля $target$ представляє собою список журналів для публікацій список учених для співробітництва: $\langle J, S \rangle$.

Функція кластеризації розбиває заданий набір профілів на $1 \leq k \leq n$ кластерів і повертає номер кластера, до якого належить даний профіль $P_{jp}: \mathbb{N}^2 \rightarrow \{1, 2, \dots, \rho\}$. Список профілів учених цільового квартиля в одному кластері з P_j :

$$\mathcal{S}\{P_i | Q(P_i) = target, \rho(j, k) = \rho(i, k)\}, P_i \in R, 1 \leq i \leq n.$$

Список журналів, у яких публікувалися вчені зі списку $S:|S| P_i$ journals.
Формування рекомендації відбувається на основі наступних альтернативних шаблонів:

– сконцентруватися на поліпшенні наступних показників (у порядку зниження значимості): $X, Y, Z\dots$;

– по збільшенню X : слід досягтися рівня – не менш ніж Y , тобто підвищити свій показник на $Z\dots$;

– необхідний приріст до наступного звітного періоду складає $+F\dots$;

– по X слід зберегти позитивну динаміку (підвищення на Y) \dots ;

– рекомендуються наступні журнали для публікації: $X, Y, Z\dots$

4 ОПИС РОЗРОБЛЕНОГО ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ

4.1 Проектування програмної системи

Програмна система призначена для інтелектуального аналізу наукометричних даних – деякої сукупності значень наукометричних показників публікаційної активності для набору авторів. Користувач програмної системи представлено двома ролями (акторами): автор (викладач, дослідник, що має публікації) – використовує основний функціонал програмної системи для одержання індивідуальних рекомендацій і керівник – використовує програмну систему для аналізу публікаційної активності авторського колективу, а також готує програмну систему до роботи з автором.



Рисунок 4.1 – Діаграма варіантів використання

Функціональні можливості (варіанти використання) програмної системи). Формування публікаційного рейтингу. Прецедент, що надає користувачеві можливість перегляду таблиці рейтингу авторів, побудованої по наукометричним даним, раніше завантаженим в систему (на основі оригінальної методики оцінки публікаційної активності).

Оцінка публікаційного потенціалу. Дозволяє користувачеві ознайомитися з результатами прогнозу можливостей авторів по переходу до кращого квартилю в

публікаційному рейтингу (на основі пошуку асоціативних правил).

Пошук груп публікаційної активності. В рамках даного прецеденту виконується формування груп авторів, близьких по своїй публікаційній активності з виділенням їх типових представників (на основі кластеризації з виділенням оптимального числа кластерів або з розбивкою на задане число кластерів).

Одержання індивідуальних рекомендацій. Дозволяє користувачу одержати набір рекомендацій із просування в рейтингу публікаційної активності й реалізації публікаційного потенціалу (на основі аналізу місця автора в публікаційному рейтингу щодо кращого, типового й гіршого авторів, а також добору журналів, у яких опубліковано роботи авторів).

Підготовка наукометричних даних. Узагальнений прецедент, доступний тільки керівникові, що реалізує програмні можливості по підготовці й предобробці наукометричних даних авторів, їх завантаженню в програмну систему й, по необхідності, відновленні.

Налаштування аналітичних моделей. Виконується керівником. Прецедент, що реалізує програмні можливості по налаштуванню робочих методик, а також параметрів моделей інтелектуального аналізу наукометричних даних, надалі використовуваних системою для роботи. Зокрема, сюди ставиться вибір джерела даних, контроль предобробки даних, добір мінімальної підтримки й вірогідності для шаблонів публікаційної активності, вибір способу визначення числа кластерів і ін.

Для програмної системи на даному рівні деталізації проектування істотні класи предметної області: автор (*Author*), стаття (*Article*) і журнал (*Journal*). Дані класи асоційовані відповідно до логіки предметної області й мають істотні для програмної системи атрибути.

Автор у програмній системі працює з пойменованим набором наукометричних даних надаваних сховищем наукометричних профілів (*Profilestore*) і аналітичних моделей (*Model*).

Діаграма класів, проектованої програмної системи показано на рисунку 4.2.

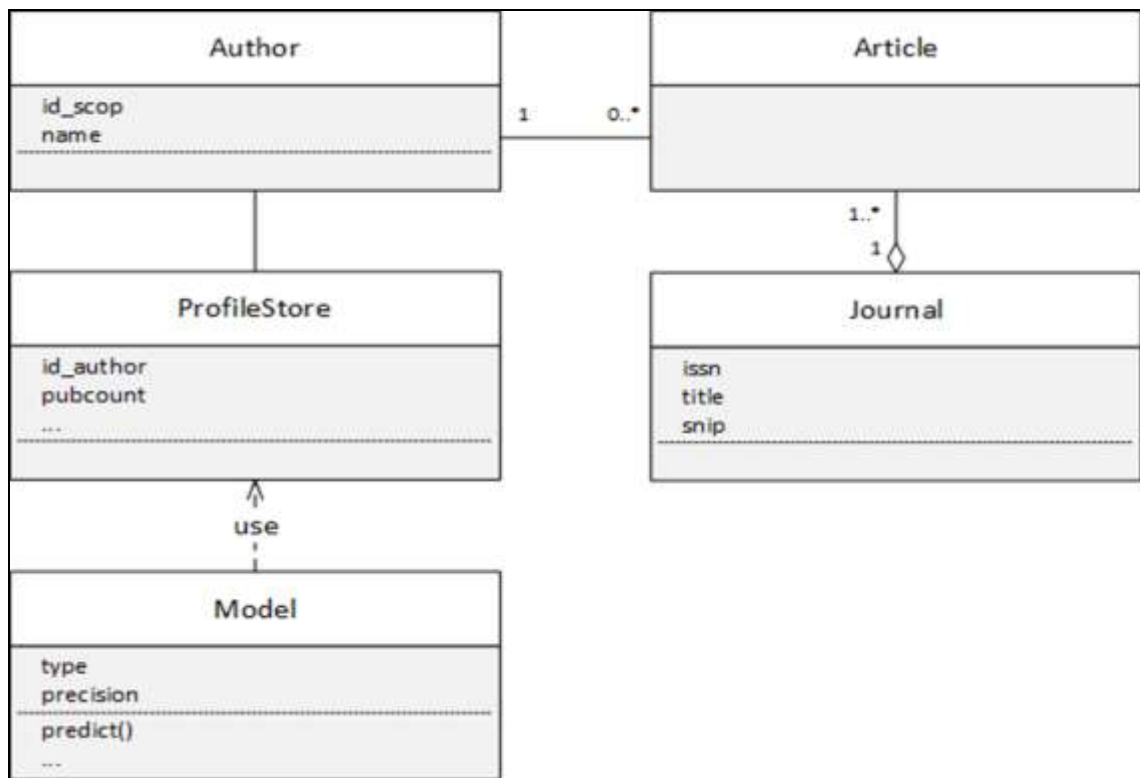


Рисунок 4.2 – Діаграма класів проектованої програмної системи

Користувач може створювати, видаляти й змінювати наявні в системі набори даних і моделей. Набір створюється шляхом відбору даних і моделей із завантажених і підготовлених керівником (модератором) раніше, у залежності від області досліджень і інших критеріїв. На основі наявних наборів даних, певних атрибутами наукометричного профілю автора в системі Scopus, програмна система будує публікаційний рейтинг.

4.2 Проектування аналітичної моделі

Аналітична модель являє собою, створений і настроєний модератором інтелектуальний об'єкт, який використовується для оцінки й прогнозування приналежності автора до тієї або іншої публікаційної групи. Передбачається, що автор може бути віднесений до чотирьох основних публікаційних груп, що

відповідають квантилям публікаційного рейтингу, а також до однієї із природніх публікаційних груп (наприклад, у випадку кластеризації з виділенням оптимального числа кластерів).

Перелік атрибутів, збережених в Profilestore, що представлений у табл. 4.1, 4.2.

Таблиця 4.1 – Перелік наукометричних атрибутів відкритого профілю автора в системі Scopus

Найменування атрибута	Опис
pubcount	Число публікацій автора
citcount	Число цитат автора
hindex	Індекс Гірша
authcount	Число співавторів
yearstart	Рік першої публікації

Таблиця 4.2 – Перелік предоброблених наукометричних атрибутів

Найменування атрибута	Характеристика
top10_sum	Загальна кількість публікацій в TOP10% видань по імпаکت-факторові віднесене до років від першої публікації
snip_mean	Середньозважений імпакт-фактор журналів у яких є публікації
h_index	Індекс Гірша віднесений до числа років від першої публікації
doc_count	Загальна кількість публікацій віднесене до числа років від першої публікації
doc_per_au	Кількість публікацій віднесене до співавторів

При цьому відповідно до методики для формування рейтингу й реалізації

інших варіантів використання застосовується попередня обробка й обчислення додаткових атрибутів.

На основі обраної моделі й наявних у сховище даних, методом `predict()` проводиться оцінка й прогнозування належності автора до публікаційної групи, оцінка публікаційного потенціалу.

Програмна система складається із трьох основних компонентів, а також використовує зовнішню інтелектуальну систему, що обробляє дані на платформі KNIME Analytics Platform.

Основні компоненти надають і використовують відповідні інтерфейси (рис. 4.3):

- користувачський інтерфейс програмної системи й/або інтерфейс програмування (компонент *Gateway*);
- компонент зберігання даних (компонент *DataSystem*);
- базова обчислювальна підсистема (компонент *BaseSystem*), що реалізує взаємодію зі сховищем даних, обчислення й обробку результатів роботи зовнішньої інтелектуальної системи KNIME Analytics Platform.

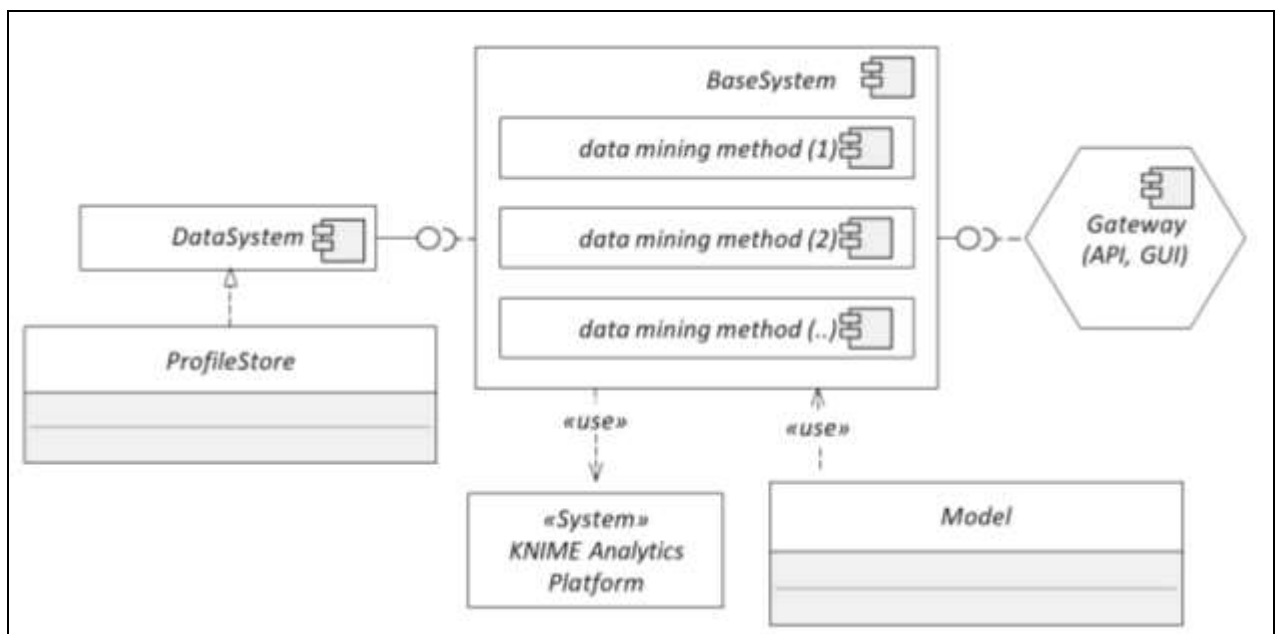


Рисунок 4.3 – Діаграма компонентів програмної системи

Розгортання (див. рис. 4.4) прототипу програмної системи передбачається на одному обчислювальному вузлі (персональному комп'ютері) із попередньо

встановленою актуальною версією KNIME Analytics Platform.

Базова обчислювальна підсистема (компонент *Basesystem*) передбачає модульну організацію таким чином, що кожний модуль незалежно реалізує один з варіантів використання.

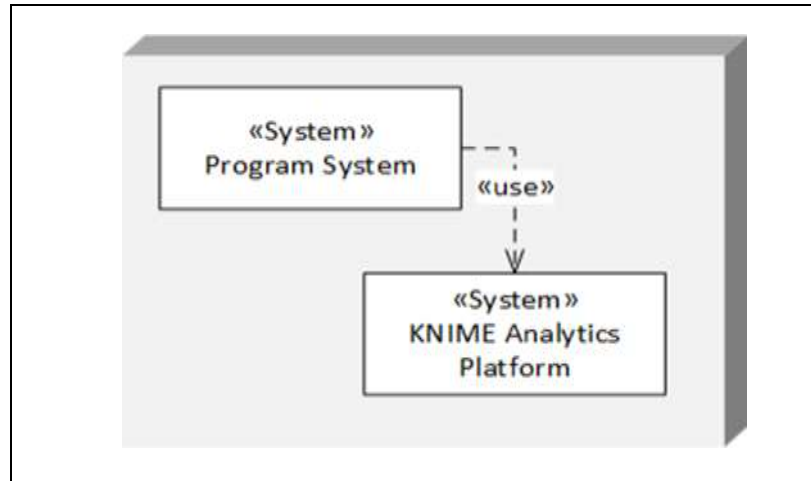


Рисунок 4.4 – Діаграма розгортання програмної системи

Структура основних модулів: ядром функціонального модуля є обчислювальний потік (KNIME workflow), окремий екземпляр якого опрацьовує дані на платформі KNIME. З найменування модулів ясно, що: *scopus_rating*, реалізує формування публікаційного рейтингу на основі даних системи Scopus; *scopus_rules* – оцінку публікаційного потенціалу; *scopus_cluster* – пошук груп публікаційної активності; а *scopus_preparing* – підготовку й предобробку наукометричних даних.

Проектування організації користувацького інтерфейсу. Схема організації робочого вікна представлена на рисунку 4.5.

Виходячи зі звичного користувацького дизайну розповсюджених операційних систем лінійки Windows, робоче вікно повинне містити головне меню (1), що надає доступ до основних параметрів програмної системи, довідці й допомоги, а також область статусу (5), у якій буде відображена балка основних подій і інша службова інформація.

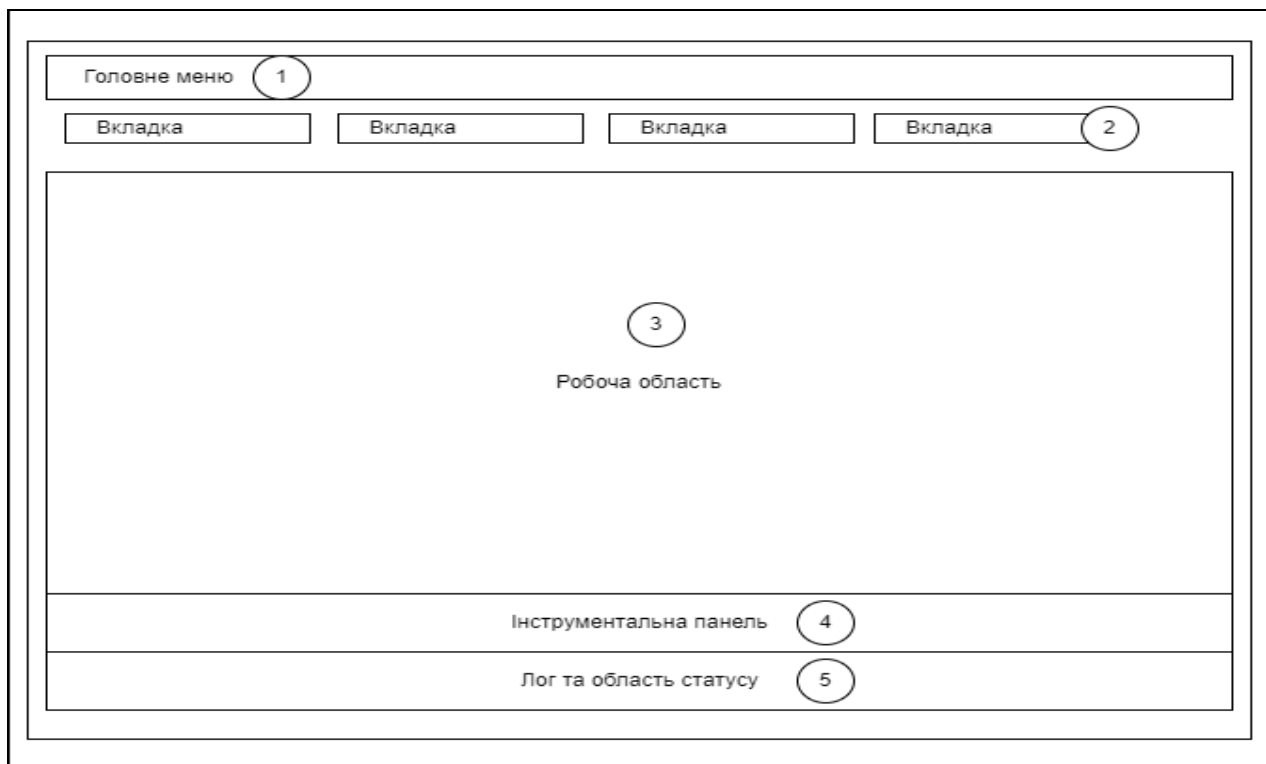


Рисунок 4.5 – Схема робочого вікна користувацького інтерфейсу

Основна робоча область (3) вікна надає користувачеві різноманітні способи відображення результатів наукометричного аналізу, зокрема – таблиці, текстові відомості, графічну інформацію. Інформація, представлена в робочій області повинна надаватися залежно від обраного варіанта використання й застосованих інструментів, при цьому кнопки інструментарію розміщуються на інструментальній панелі (4).

Вибір доступних варіантів використання системи здійснюється шляхом перемикування вкладок (2), кожна з яких відповідає за свій варіант використання й надає свій інструментарій на панелі (4).

Дана схема інтерфейсу є тільки достатній варіант представлення функціональних можливостей системи.

У силу модульної організації системи й у випадку її реалізації у вигляді, наприклад, мікросервісів користувацький інтерфейс може бути виконаний на основі незалежного web-представлення.

5 ОПИС МОЖЛИВОСТІ ВИКОРИСТАННЯ ОТРИМАНИХ РЕЗУЛЬТАТІВ

Відповідно до розглянутих вище проектних матеріалів, у рамках справжньої роботи представлений реалізований прототип програмної системи.

У якості інтегрованого середовища розробки обрана Embarcadero Delphi 10 Delphi надає найшвидший спосіб написання, компіляції, складання й розгортання системних і кросплатформених застосунків, поєднує гнучкість сучасної мови Object Pascal, з native-компіляторами й бібліотеками компонентів для швидкої розробки для Windows, macOS, ios, Android і Linux на базі єдиного коду.

Вибір даного середовища обумовлений мінімальними тимчасовими витратами на розробку прототипу програмної системи й можливостями подальшої доробки прототипу до повноцінного кросплатформеного програмного комплексу.

З тих же міркувань, простоти й швидкості розробки, у якості інструментарію роботи зі сховищем даних обрана технологія Activex (ADO), що включає клієнтські додатки для доступу й керування даними з різних джерел через постачальника OLE DB. З урахуванням того, що ADO підтримує основні функції для створення клієнта й сервера веб-додатків, також передбачається подальша доопрацювання прототипу сховища до веб-сервісу.

Оскільки, по суті, графічний інтерфейс і сховище є лише способами вистави результатів і зберігання розрахованих даних, а основним обчислювальним ядром є робочі потоки KNIME, представлені незалежними здійсненими файлами, розумно використовувати два етапи тестування системи.

Зокрема, для тестування розроблено три набори даних, для яких заздалегідь, у напівавтоматичному режимі, проведені всі процедури згідно з алгоритмами побудови рейтингу, оцінки публікаційного потенціалу й розрахунку метрик для формування індивідуальних рекомендацій.

Крім того, коректність самих файлів робочих потоків перевірялася за допомогою вбудованих засобів контролю цілісності й зв'язності потоків даних

для KNIME. Результати тестів методом чорного ящика представлені в таблиці нижче й можуть бути визнані задовільними й достатніми для позитивного рішення про готовність системи.

Таблиця 5.1 – Результати функціонального тестування

№ тесту	№ тестового набору	Коректність обчислень в підсистемі KNIME	Коректність обробки й представлення в програмній системі	Результат тесту
1	1	Вірно	Вірно	пройдений
2	2	Вірно	Вірно	пройдений
3	3	Вірно	Вірно	пройдений

Несистематичні помилки, пов'язані з якістю самих вхідних даних і їх придатністю для обчислення, не можуть бути передбачені заздалегідь і будуть оцінені нами в період досвідченої експлуатації програмної системи. Таким чином, функціонально й технічно розроблений програмний продукт задовольняє пропонованим вимогам і його можна допустити в експлуатацію.

У рамках апробації програмної системи в частині реалізованих методик побудови публікаційного рейтингу й оцінки публікаційного потенціалу.

У таблиці представлені групові оцінки (табл. 5.2).

Таблиця 5.2 – Публікаційний потенціал авторів у галузі «computer science»

Вихідний кuartиль	Кількість авторів	Середня підтримка правил по кuartиллю, %	Середня достовірність правил по кuartиллю, %	Целевий кuartиль (consequent)
Q2	13	6,0	90,2	Q1
Q3	48	9,0	65,8	Q2
Q4	4	7,6	60,2	Q2

Загалом програмна система на наявних даних буде 4163 правила за мінімальної підтримки та достовірності 5,0 %. Вибірка з надійністю щонайменше 50,0 % виявляє 78 правил. Проте, для переходу авторів у публікаційні кuartілі

рейтингу «Q1» та «Q2», а саме більшою мірою цікавлять з погляду аналізу публікаційного потенціалу, виділено 20 найкращих правил, які відповідають заданим критеріям достовірності. Відповідно до виділених правил 28 авторів можуть бути потенційно (з різним ступенем достовірності) перекваліфіковані на кращий кuartиль.

Для цієї мети в сховище програмної системи імпортовані дані наукометричних профілів авторів, віднесених Scopus до області дослідження «computer science» на 1 липня 2021 р.

ВИСНОВКИ

Таким чином, відповідно до передумов застосовуваного метода інтелектуального аналізу, методика оцінки публікаційного потенціалу й алгоритм виявлення авторів, що мають такий потенціал, будуть полягати в наступному.

У роботі були вивчені й проаналізовані формальні способи інтелектуальних систем: способи представлення знань залежно від конкретних галузей застосування систем; формальні мови, що дозволяють представляти знання в пам'яті комп'ютера.

Для заданого набору авторів будується публікаційний рейтинг на основі запропонованої методики оцінки публікаційної активності.

На основі отриманого публікаційного рейтингу набір авторів розбивається на чотири квартиля. Перший квартиль (перші 25% авторів) інтерпретується як кращий на основі оцінки публікаційної активності, другий квартиль – як гарний і т.д. Кожний автор здобуває синтетичний атрибут, що містить мітку приналежності до квартилю.

По кожному атрибуту профілю (наукометричному показнику), що приймає участь в побудові рейтингу проводиться квартильна дискретизація. Дане корекційне правило приводить до викривлення квартильної структури по деяких атрибутах, але забезпечує вірну інтерпретацію наукометричного профілю автора.

Для отриманих наборів синтетичних атрибутів по кожному автору, що представляють собою транзакцію, виконується пошук асоціативних правил. Потім проводиться вибірка стійких правил, що виявляють перехід авторів із власних квартилей у кращі по поєднанню наукометричних атрибутів. На основі отриманих правил виявляється набір авторів їм відповідних, тобто, що мають потенціал перехід до кращого квартилю рейтингу публікаційної активності. При цьому прогностичним заходом публікаційного потенціалу автора є захід вірогідності підходящого стійкого правила.

ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛЕННЯ

1. Загорулько Ю.А., Загорулько Г.Б., Боровикова О.І. Технологія створення тематичних інтелектуальних наукових інтернет-ресурсів, що базується на онтології // Програмна інженерія. – 2016. – Т.7, №2. – С. 51–60.
2. Загорулько Г.Б. Розробка онтології для Інтернет-Ресурсу підтримки прийняття рішень у слабоформалізованих областях / Г.Б. Загорулько // Онтологія проектування. – 2016. – Т. 6, №4(22). – С. 485-500. – DOI: 10.18287/2223-9537-2016-6-4-485-500.
3. Зайцева А.С. Моделювання терміносистеми надзвичайних ситуацій / А.С. Зайцева, Ю.В. Сложенікіна // Онтологія проектування – 2018. – Т. 8, №4(30). – С.562-570.
4. Історія Львівського університету https://1.bp.blogspot.com/-sqa_wBUJakU/YPcWlQ9n7kI/AAAAAAAAAMZQ/5aUkRwjzcLAZyjhQ1eAqgm0x-cz9LNPQCLcBGAsYHQ/s960/senat_uu.jpg
5. Біографія В. Налімов <https://m.rusmir.media/2017/10/31/nalimov>
6. Налімов В.В. Наукометрія: Изучение развития науки как информационного процесса / В.В. Налімов, З.М. Мульченко. – М.: Наука, 1969. – 192 с.
7. Aggarwal Charu C, Zhai Cheng Xiang. Mining text data. Springer Science & Business Media, 2012.
8. Alan Griffiths, H. Claire Luckhurst, and Peter Willett. Using Interdocument Similarity Information in Document Retrieval Systems. Department of Information Studies, University of Sheffield, Western Bank, Sheffield S10 2TN, United Kingdom. Information Retrieval 1997 p.365-373.
9. Artale A., Franconi E., Guarino N., Pazzi L. Part-Whole Relations in Object-Centered Systems: An Overview // Data and Knowledge Engineering. 1996. V.20. P. 347-383.
10. Whissell John S., Clarke Charles L.A. Improving document clustering using Okapi BM25 feature weighting. Information retrieval. 2011. Т. 14, № 5, pp. 513-523.
11. Fouad M. Data mining and fusion techniques for Wsns as a source of the big

data / M.M. Fouad, N.E. Oweis, T. Gaber, M. Ahmed, V. Snasel // *Procedia Computer Science*. – Elsevier, 2015. – Vol.65. – P. 778-786.

12. Methods of multidimensional classification in problems of linguistic localization / Shubin I., Kozyriev A. // *Proceedings of the III International Conference "Innovative Technologies in Science and Education"*. November 14, 2019 in Amsterdam, The Netherlands, 2019. pp 398-402

13. Beloborodov A., Kuznetsov A., Braslavski P. Characterizing Health-Related Community Question Answering // *Proc. of the 35th European Conf. on IR research (ECIR'13): LNCS.. Vol. 7814. 2020. , P. 680-683.*

14. Beloborodov A., Braslavski P., Driker M. Towards Automatic Evaluation of Health-Related CQA Data // *Proc. of the 5th International Conf. of the CLEF Initiative (CLEF'14): LNCS. - Sheeld, UK. Vol. 8685. 2020. , P. 7-18.*

15. Sieg A., Mobasher B., Burke R. Inferring User's Information Context from User Profiles and Concept Hierarchies // *Classification, Clustering, and Data Mining Applications*. | 2019., | P. 563-573.

16. Chirita P., Firan C., Nejdl W. Personalized Query Expansion for the Web // *Proceedings of SIGIR'07 Conference*. | 2017. | P. 7-14.

17. Speretta M., Gauch S. Personalized Search Based on User Search Histories // *Proceedings of the IEEE/WIC/ACM International Conference on Web Intelligence*. - 2019 . P. 622-628.

18. Chetverikov G., Puzik O., Vechirska I. Multiple-valued structures of intellectual systems // *Proceedings of the with Internations Computer Sciences and Information Technologies (CSIT)*. 2016, 7589907. -pp. 204-207

19. Ould-Amer N., Mulhem P., LIG at CLEF 2015 SBS Lab // *Working Notes of CLEF'2019 Conference*. | 2019.

20. O. Lassila. Introduction to RDF Metadata // *World Wide Web Consortium*, – November 1997. <http://www.w3.org/TR/Note-rdf-simple-intro-971113.html>.

21. Тарасевич Ю.Ю., Шиняева Т.С. Временная динамика индекса Хирша // *Математическое моделирование и программирование*. – 2018. – № 1 – С. 32-45.

22. The Interaction Domain// World Wide Web Consortium, - November 2003.
<http://www.w3.org/Interaction/>.
23. M.F. Bondarenko, Z.V. Dudar, N.T. Protsay, V.V. Cherkashyn, V.A. Chykyna, Y.P. Shabanov-Kushnarenko, “Algebra of predicates and predicate operations” Radio electronics and informatics, no. 1, 2004, pp. 5-22.
24. The protégé-2000 project// Stanford, California, - 2000-2014.
<http://protege.stanford.edu/index.html>.
25. O. Lassila. The XML Family of Specifications: A Practical Guide// Addison-Wesley, - 2002. <http://shop.barnesandnoble.com/booksearch/isbninquiry.asp?isbn=0201703599>
26. Semantic Web Development // World Wide Web Consortium, - January 2000. <http://www.w3.org/2000/01/sw/>
27. Semantic Web // World Wide Web Consortium, - September 2016.
<http://www.w3.org/2016/sw/>.
28. Naming and Addressing: Uris, Urls // World Wide Web Consortium, - October 2013. <http://www.w3.org/Addressing/>
29. G.G. Chetverikov, I.D. Vechirska, S.S.Tanyanskiy, “The methods of algebra finite predicates in the intellectual system of complex calculations of telecommunication companies,” International Conference Proceedings Crimean Microwave and Telecommunication Technology (CriMiCo), 6959425, 2014, pp.
30. Shubin, I., Snisar, S., Zhyrnov, V., Slavhorodskyi, V.// Practical Application of Formal Representation of Information for Intelligent Radar Systems 2018 International Scientific-Practical Conference on Problems of Infocommunications Science and Technology, PIC S and T 2018 - Proceedings, 2019, pp. 433-436, 8632103
31. DTD for XML Schema // World Wide Web Consortium, - August 2002.
<http://www.w3.org/TR/xmlschema-1/#nonnormative-schematd>
32. XSL Transformations (XSLT), W3C Recommendation // World Wide Web Consortium, - November 1999. <http://www.w3.org/TR/1999/Rec-xslt-19991116>.

33. Resource Description Framework (RDF) Model and Syntax Specification, W3C Recommendation // World Wide Web Consortium, – February 1999. <http://www.w3.org/TR/2019/Rec-rdf-syntax-19990222/>.
34. Web Ontology // World Wide Web Consortium, – March 2014. <http://www.w3.org/Help/siteindex.html#webont>.
35. DAML+OIL Reference Description // World Wide Web Consortium, - March 2020. <http://www.w3.org/TR/daml+oil-reference>
36. G. Wiederhold. Mediators in the architecture of future information systems// IEEE Computer, -March 2016. - p. 38-49.
37. Gruber T. Towards principles for the design of ontologies used for knowledge sharing// International Journal of Human-Computer Studies, 43: 907-928,- 2014.
38. N. Guarino. Some ontological principles for designing upper level lexical resources. In Proceedings of the First International Conference on Language Resources and Evaluation, Granada, 2008. pp. 3–15