

ОГЛЯД МЕТОДІВ ТА ТЕХНОЛОГІЙ ІНФОРМЕТРІЇ

Анотація. Наведено огляд методів та технологій інформетрії. Визначено її задачі та напрямки. Показано взаємозв'язок таких понять, як наукометрія (scientometrics), бібліометрія (bibliometrics), інформетрія (informetrie), вебометрія (webometrics), альтметрія (altmetrics). Проаналізовано наявні моделі і методи інформетрії, а також її актуальні проблеми. Розглянуто аналітично-пошукові наукометричні бази даних та системи, виявлено їхні недоліки та переваги. Визначено характеристики наукометричних систем, їхні складові та чинники, які впливають на наукометричні індекси. На основі проведеного дослідження сформульовано перспективні напрямки розвитку наукометричних систем.

Ключові слова: наукометрія, інформетрія, наукометричні системи, бібліометричні системи, метрики, індекси цитування.

ВСТУП

Наукові досягнення, технічний прогрес зумовлюють сталий розвиток держави за рахунок конкурентної економіки та промисловості, що сприяє покращанню якості життя громадян. Актуальними завданнями у цьому напрямку є удосконалення наявних та створення новітніх технологій і методів задля:

- управління науковим процесом;
- забезпечення якості наукового процесу;
- визначення пріоритетних напрямків розвитку науки;
- забезпечення оригінальності академічних робіт;
- визначення найвпливовіших науково-дослідницьких установ та вчених.

У світі для оцінювання наукової діяльності використовують багато технологій та методів. Найчастіше вживаними є: експертна оцінка; наукометричне оцінювання, яке здійснюють, ґрунтуючись на бібліометричних показниках; порівняння фінансових та кадрових (кількість вчених) ресурсів, які було витрачено для отримання результату.

Одним із факторів, що впливають на прийняття рішень, є наукометрична оцінка досліджень (наукової роботи). Вона повинна бути достовірною та відображати дійсний стан речей, інакше прийняті на її основі рішення будуть хибними та неефективними.

Зважаючи на великі обсяги наукової інформації, яку необхідно опрацювати для отримання наукометричних даних, створено програмно-технічні засоби, що містять спеціалізовані пошукові та аналітичні інструменти, а саме інформаційно-аналітичні системи (ІАС) — комп'ютерні системи, основним призначенням яких є ефективне зберігання, оброблення та аналіз даних. Різновидом ІАС є наукометрична система (НС) — бібліографічна і реферативна база даних (БД) з аналітичними інструментами, можливостями відстеження цитованості статей, підрахунку наукометричних показників. Міжнародна практика наукометричних досліджень базується на використанні двох універсальних комерційних БД: Web of Science Core Collection та Scopus.

На основі БД, що формуються глобальними НС, ухвалюють важливі стратегічні рішення щодо розвитку науки, техніки, технологій і визначають найвпливовіші академічні та науково-дослідні установи країн світу. Від достовірності вихідних даних з НС залежить успіх реалізації стратегічних планів розвитку наук та визначення справжніх лідерів у сфері науково-технічної діяльності [1].

ВИЗНАЧЕННЯ ЗАДАЧ ІНФОРМЕТРІЇ

На початку ХХ ст. почали з'являтися публікації у напрямку, який передував сучасній інформетрії та наукометрії. У цих роботах новий науковий напрям було описано новими термінами. Так, у 1923 р. Е.В. Гульм описав «статистичну бібліографію» (statistical bibliography) [2], у 1934 р. П. Оттл — «бібліо-

метрію» (bibliométrie), у 1948 р. С.Р. Ранганатан — «бібліометрію» (librametry), у 1966 р. В.В. Налимов — «наукометрію» (scientometrics), у 1969 р. А. Причард — «бібліометрію» (bibliometrics), у 1979 р. О. Наке, Л. Бертері, К. Зігель — «інформетрію» (informetrie), у 1997 р. Т.С. Альмінді, П. Інгерсен — «вебометрію» (webometrics), у 2010 р. Дж. Прієм і співавтори — «альтметрію» (altmetrics) [3].

Інформація створює документальний потік (ДП) — масив семантично пов'язаних документів, що є унікальним засобом зберігання, накопичення і передавання знань про навколишній світ. Він безперервно розвивається. Це інтелектуальна пам'ять цивілізації, штучно створена система, яка сформувала в процесі розвитку власні структури (тематичну, видавничу, видову, мовну, ціннісні зв'язки між документами тощо), закономірності розвитку — темпи зростання, старіння, ступінь розсіювання, дублювання та ін. Проблеми розвитку кількісних вимірювань потоку інформації загострюються з розширенням обсягів ДП, тому виникла необхідність перегляду методів, технологій, а також збільшення ресурсів для проведення досліджень. У роботі [4] визначено такі сучасні проблеми розвитку кількісних досліджень інформаційного потоку:

- удосконалення індикаторів та інструментів проведення вебметричних досліджень;
- веб-цитовання як фактор наукової престижності закладів вищої освіти, учених;
- удосконалення методів бібліометрії та формування рейтингів;
- впровадження нових методик оцінювання наукового внеску;
- складні прийоми динамічного картування реального потоку за запитамі;
- некоректне використання веб-статистики.

Перші карти науки, сформовані в Institute for Scientific Information (ISI) до 1990 р., виявили циклічний характер науки, неодномірність її структури, невизначений статус окремих галузей, а також уточнили спеціалізацію наукових напрямків на національному рівні [5].

Терміни наукометрія, бібліометрія, інформетрія, вебметрія, альтметрія використовують і сьогодні. Для кращого розуміння цього напрямку досліджень було введено також інші терміни. У 2004 р. Л. Бьорнеборн, П. Інгерсен [6] запропонували диференційовану термінологію та схему співвідношення понять (рис. 1). Як видно з цього рисунку, поняття пов'язані між собою та еволюціонують відповідно до розвитку інформаційних технологій. Інформетрія швидко розвивається, тому поняття альтметрія не відображено на схемі. Воно з'явилося вперше у 2010 р. [7, 8]. Альтметрія розширює уявлення про впливовість інформації та її формування.

Оскільки альтметрія аналізує дані у веб-просторі та використовує методи і технології наукометрії та бібліометрії, будемо розглядати схему співвідношень понять на основі схеми, розробленої Л. Бьорнеборном та П. Інгерсеном у 2004 р. (рис. 2).

Інформетрія використовує різноманітний математичний апарат для аналізу, виявлення закономірностей, формулювання законів інформаційної діяльності та наукової інформації, а також прийняття рішень в інформаційній практиці [9]. Інформетрія запозичує інструменти (методики, моделі, аналогії) з математики, фізики, інформатики, математичної лінгвістики та інших наук. Кількісні показники, які розраховують з використанням зазначених інструментів, назвемо інформетриками.

Майбутнє інформетрії фахівці пов'язують головним чином з розвитком методів інтелектуального аналізу інформації, розробленням багатовимірних динамічних моделей наукової та соціальної комунікації [10, 11]. У 1983 р. було детально описано комплекс і можливості методів математичної статистики для вивчення закономірностей розвитку потоків наукової інформації, показав

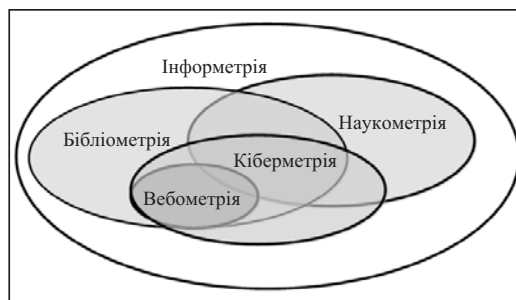


Рис. 1. Схема співвідношення понять інформетрія, бібліометрія, кіберметрія, наукометрія, вебметрія

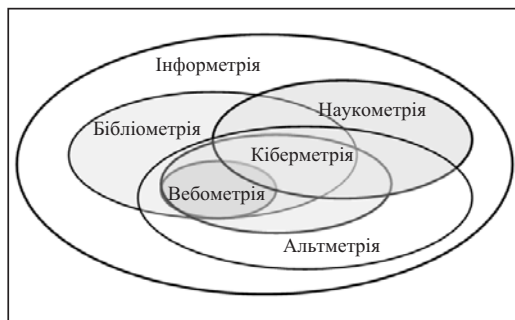


Рис. 2. Схема співвідношення понять інформетрія, бібліометрія, кіберметрія, наукометрія, вебометрія, альтметрія

мації для виявлення емпіричного матеріалу і побудови теоретичної моделі функціонування науки [10].

Наукометричні дослідження зазвичай поділяють на такі основні напрямки:

- удосконалення методів та показників;
- дослідження загального розвитку системи, конкретної галузі науки, структури та взаємозв'язків різних галузей;
- управління наукою, формування наукової політики [10];
- створення наукових знань (кількісні характеристики наукового потенціалу, комунікації в науці, наукова продуктивність, оцінка вчених та наукових установ, наукове співробітництво, структура наукових комунікацій та мереж тощо);
- макроекономічне середовище наукових досліджень, а саме дослідницька політика, інноваційні процеси, глобалізація тощо [11].

Завдання бібліометрії як частини інформетрії — кількісне відображення зв'язків між одиницями інформації про об'єкт у локалізованому предметному та інформаційному полі за обраними індикаторами [4]. Дослідження з бібліометрії надають можливість:

- оцінювати темпи розвитку предметної області;
- виявляти провідних фахівців, організації і країни з певною оцінкою їхнього внеску в розвиток галузі знання, тематичного спрямування;
- визначати перспективні напрямки розвитку науки, а також прогнозувати окремі параметри реального і цитованого ДП [4].

Термін кіберметрія (cybermetrics) визначає напрямок, який досліджує кількісні аспекти створення і використання інформаційних ресурсів, структур і технологій в Інтернеті. Вебометрія як складова кіберметрії аналізує рейтинги сайтів і зв'язки між ними, проводить моніторинг інформаційних ресурсів та сервісів Інтернету і т.п. Основні сфери вебометричних досліджень:

- контент-аналіз веб-сторінок;
- аналіз структури посилань на сайт;
- веб-аналіз використання (визначення та аналіз поведінки користувачів);
- веб-технологічний аналіз (робота пошукової системи).

Стрімкий розвиток інформаційних технологій, широке використання електронних інформаційних ресурсів, розвиток Інтернету та міжнародний рух відкритого доступу до наукових знань сприяє змінам у системі наукової комунікації. Змінюється не лише традиційна модель публікації і поширення наукових знань, а й модель їхньої оцінки. Виникають нові наукові проблеми та актуальні задачі інформетрії, бібліометрії, кіберметрії, наукометрії та вебометрії. У зв'язку з цим постають такі технологічні завдання:

- розв'язання неоднозначностей, глобальна ідентифікація авторів, наукових організацій, інших об'єктів;
- визначення авторського стилю написання академічних робіт для виявлення можливого плагіату та підтвердження авторства;
- забезпечення достовірності наукометричних показників;
- мережевий аналіз, розроблення нових методів оптимізації структури мережі.

но роль лінгвістичних методів, у тому числі сленгового аналізу [12]. Разом з інформетрією вживають термін інфометрія, який являє собою ту складову інформетрії, що орієнтована на результати аналітичних досліджень та візуалізацію галузей знання, наведених у вигляді «рисунків» — графіків, діаграм, картографії, мереж зв'язку, художніх образів та ін.

Наукометрія склалася на основі досліджень математичними методами різних аспектів наукової діяльності. Вона вивчає потоки інфор-

Актуальні задачі інформетрії [3, 5] мають такі складові.

1. Використання нових інформаційно-комунікаційних технологій, наприклад Big Data.
2. Розвиток моделей і методів інформетрії:
 - формування нових напрямків дослідження (вебометрія, альтметрія);
 - розроблення сучасних методів картографії (мепінг) та візуалізації наукових даних;
 - створення нових індикаторів і модифікація наявних (h-індекс, SNIP, SJR);
 - формування нових БД цитування, таких, як Scopus і Google Scholar;
 - створення регіональних і національних БД цитування (Китай, Латинська Америка, Іспанія, Росія, Тайвань, Японія).
3. Вивчення впливу наукометричних індикаторів на поведінку вчених (штучне/договірне цитування).
4. Активне застосування бібліометрії та наукометрії в науковій політиці та управлінні фінансуванням науки, а також у національних програмах розвитку науки і національних системах оцінки результатів наукових досліджень.
5. Використання бібліометричних та наукометричних індикаторів у міжнародних і національних рейтингах установ.

ОГЛЯД ІНФОРМАЦІЙНО-АНАЛІТИЧНИХ ПОШУКОВИХ НАУКОМЕТРИЧНИХ БАЗ ДАНИХ ТА СИСТЕМ

Наукометрична БД є одним із різновидів інформаційно-аналітичних систем. Її можна визначити як платформу, яка складається з документів та їхніх метаданих. Усі дані аналізують та оброблюють, на їхній основі проводять кількісні вимірювання. Використовуючи кількісні показники, можна проаналізувати публікаційну активність та цитованість; впливовість того чи іншого видання, організації, науковця тощо [13]. Іншими словами — це бібліографічна та реферативна БД з інструментами для відстеження цитувань статей, опублікованих у наукових виданнях. Списки літературних посилань легко обробити та рейтингувати, що є об'єктивною допоміжною їхньою оцінкою. Суб'єктивне пряме оцінювання якості — більш складне завдання. Таку оцінку можливо отримати шляхом формалізації ознак, за якими можна визначити або описати якість наукової роботи. У сучасному науковому просторі використовують відомі методи та технології наукометрії, які є об'єктивною допоміжною оцінкою. Але більш оптимальним є комплексний підхід з використанням прямої оцінки якості [13].

Об'єкти вивчення для бібліометричного аналізу — публікації, згруповані за різними ознаками (сегменти документопотоку, мікропотоки): авторами, журналами, тематичними рубриками, країнами та ін. На основі метаданих публікацій формують інформаційне наповнення наукометричних БД, яке можна поділити на такі типи:

- змістовна, реферативна складова (проблематика наукових досліджень);
- інформація про суб'єктів — авторів і установи, міста, країни (аналітична складова бібліометричних досліджень).

Інформація зазначених типів дозволяє проводити дослідження ДП і формувати звіти за науковими напрямками, авторами, виданнями та регіонами.

На основі вказаного означення наукометричною БД можна вважати будь-яку БД, яка має джерельну базу, аналітичні інструменти та розраховує наукометричні показники. Прикладами є: Web of Science Core Collection (WoS CC), Scopus, Google Scholar (GS), Російський індекс наукового цитування (ПІНЦ) [14], Index Copernicus (IC) [15], Microsoft Academic (MA) [16] та ін. Google Scholar та Microsoft Academic не мають власної джерельної бази та за технологією належать до спеціалізованих пошукових систем. Зазначимо, що показники не кожної наукометричної бази є однаково авторитетними та використовуються при розв'язанні практичних задач.

Для проведення глобальних досліджень розвитку науки та визначення властивостей ДП (документів), складаючи світові рейтинги, використовують показники глобальних наукометричних систем: WoS CC, Scopus, Google Scholar. Ці бази є мультидисциплінарними та містять видання майже усіх країн світу. Саме мультидисциплінарна база дозволяє розширити перелік видань, в яких можливе

потенційне цитування статей з окремої дисципліни. Будь-яка БД, обмежена за науковим напрямом, не може претендувати на об'єктивний підрахунок наукометричних показників, оскільки значна кількість цитувань може надходити із суміжних за напрямом видань [13]. «Закон розсіювання» наукових публікацій у світовій періодиці було сформульовано С. Бредфордом (S. Bradford), англійським бібліографом і статистиком, у 30-х рр. минулого століття. За цим законом публікації з будь-якої галузі поділяють на три групи, перша з яких містить невелику кількість ключових журналів з цієї галузі, друга — значно більшу кількість журналів, суміжних із галуззю, а третя — велику кількість журналів, в яких поява публікацій з цього напрямку малоймовірна. Співвідношення кількості журналів у зазначених трьох групах виглядає як $1:n:n^2$, де n — стала, яка залежить від галузі [17]. Саме така закономірність є одним з основних принципів формування реферативних інформаційних служб, у тому числі WoS CC, Scopus, Google Scholar та ін.

Розглянемо найбільш авторитетні наукометричні бази Web of Science Core Collection (компанія Clarivate Analytics), Scopus (компанія Elsevier) та Google Scholar (компанія Google). Перші дві БД належать відомим корпораціям та індексують десятки тисяч періодичних видань, матеріалів конференцій, книг та інших наукових документів. Доступ до БД здійснюється за передплатою. До БД долучають лише ті видання, які формують «ядро» найкращих світових наукових періодичних видань в кожній галузі науки. Для цього створено певні критерії відбору. Наукові редакції, які мають на меті реєстрацію своїх видань у WoS CC або Scopus, повинні виконувати перелік важливих вимог, сформованих за критеріями відбору. Кожна система має певний перелік видань, що індексуються. Він поповнюється новими назвами, але з нього вилучають видання, в яких погіршилась якість публікацій.

В Україні за допомогою наукометричних систем WoS CC та Scopus можна: офіційно визначити авторитетність науковця та впливовість видання для розміщення в ньому результатів власних досліджень (авторам); проаналізувати авторитетність власного наукового видання (видавництвам, редакціям), ефективність наукової роботи установи (адміністраторам) та країни (державним посадовцям). Проте внаслідок відносно малої кількості представлених наукових видань і статей авторів з України в кожній із зазначених систем потенційна ефективність їхнього використання значно знижується [13].

Google Scholar є вільно доступною пошуковою системою, орієнтованою на академічні документи та наукову літературу, яка відстежує цитування і підраховує показники для видань та авторів [18]. На відміну від WoS CC та Scopus система Google Scholar не має процедури рецензування під час відбору видань до бази, що робить неможливим контроль якості та легітимності джерел інформації. Вона автоматично індексує документи, доступні в Інтернеті, вимагаючи виконання технологічних інструкцій до опублікованих матеріалів на сайтах. За рахунок цього в системі Google Scholar наведено значно більше документів авторів з України, ніж у WoS CC та Scopus, але вони не завжди мають наукове наповнення.

Зазначимо деякі особливості системи Google Scholar:

- відсутність рецензування усіх документів;
- відсутність повного переліку індексованих ресурсів та інформації щодо частоти оновлення індексів;
- можливість зміни показників цитування у бік зменшення;
- самостійне включення ресурсів для індексування.

Україна не має власної наукометричної бази або системи, показники якої використовувалися для проведення регіональних наукометричних досліджень. Але у світі існує практика формування національних індексів наукового цитування. Першими роботу зі створення національного індексу наукового цитування розпочала Дослідна група з питань оцінки науки та наукових комунікацій Університету Гранади (Іспанія). У 1990-х роках ця група започаткувала два проекти з побудови БД цитувань: Impact index of Spanish social-science journals, Impact Index of Spanish Journal of Legal Studies. Серед країн слов'янського світу у 90-х роках ХХ ст. першою створювати такий індекс розпочала Сербія. Роботи виконував Центр оцінки освіти і науки за підтримки Міністерства науки Сербії. Сербський індекс наукового цитування охоплює всі галузі досліджень, що проводяться в країні.

Розроблення вітчизняного наукометричного інструментарію можна розглядати як формування власного ресурсу з аналітичними інструментами або в аспекті його конвергенції зі світовими науково-інформаційними платформами.

У 2014 р. видавнича служба Української науково-освітньої телекомунікаційної мережі «УРАН» впровадила національний сервіс «Український індекс наукового цитування» (УІНЦ) — технологічний комплекс для забезпечення моніторингу суб'єктів наукової діяльності в Україні.

Система забезпечує збирання, оброблення та надання доступу до даних щодо показників публікаційної активності індивідуальних та колективних суб'єктів наукової діяльності в Україні за даними наукометричних БД WoS CC та Scopus, а також проекту видавничої служби «УРАН».

Робота системи «Бібліометрика української науки» базується на основі БД Google Scholar, WoS CC, Scopus та призначена для надання суспільству цілісної картини стану вітчизняного наукового середовища. Ця система є національною складовою проекту Ranking of Scientists (Cybermetrics Lab). Її інформаційні ресурси формуються опрацюванням створених науковцями на платформі Google Scholar

Таблиця 1

Назва системи	Об'єкти, для яких підраховують показники	Основні показники	Переваги	Недоліки
Google Scholar (Google)	Автори Видання	h-індекс	Вільний доступ, вільно доступні усі показники Велика кількість джерел та їхнє глобальне представлення Зручний пошук статей, видань, учених	Показники нестабільні, можуть зменшуватися Не всі документи рецензовані Немає точного переліку включених джерел Точно невідома періодичність індексування джерел та оновлення індексної бази
Index Copernicus (Index Copernicus International, ICI)	Видання	Index Copernicus Value (ICV)	Вільний доступ, вільно доступний ICV Доступний пошук за назвами видань в ICI Journals Master List	Відсутній пошук статей видань, включених до ICI Journals Master List
Microsoft Academic (Microsoft, MS)	Автори Видання Установи	Кількість цитувань	Вільний доступ Зручний пошук документів, авторів, видань, організацій, їхнього групування	Розрахунок лише кількості цитувань Мала кількість в базі документів авторів з України
Scopus (Elsevier)	Автори Видання Установи	h-індекс SJR SNIP	Показники стабільні (не зменшуються) Показники видань та профілі учених вільно доступні Інтегрована з ORCID	Доступна за передплатою Мала кількість в базі документів авторів з України
Web of Science Core Collection (Clarivate Analytics)	Автори Видання Установи	h-індекс IF (імпаکت-фактор)	Показники стабільні Підтримує авторські профілі вільної реєстрації ResearcherID, які інтегровані з ORCID	Доступна за передплатою Мала кількість в базі документів авторів з України IF (імпакт-фактор) є закритим показником
Російський індекс наукового цитування (Наукова електронна бібліотека eLIBRARY)	Автори Видання	h-індекс 2-, 5-річні імпакт-фактори	Вільний доступ Зручний пошук документів, авторів, видань, організацій	Джерельна база містить в основному регіональні видання (країн СНД)

бібліометричних профілів та бібліометричних показників систем WoS CC і Scopus. Оновлення інформації про значення h-індексів в бібліометричних профілях учених здійснюється щомісячно, значення інших показників актуалізуються щоквартально.

Для проведення будь-яких досліджень з інформетрії використовують показники різних НС. Додатково показники можуть порівнюватися та корелюватися. У табл. 1 наведено порівняльні характеристики наукометричних баз (систем).

АНАЛІЗ НАУКОМЕТРИЧНИХ МЕТРИК (ПОКАЗНИКІВ)

Основними даними для підрахунків кількісних показників є публікації та їхні метадані, які є у НС. На цій основі здійснюють розрахунок загальної кількості публікацій та цитувань, визначають наукометричні показники та формують аналітичні дані. Об'єктивність розрахунків наукометричної БД прямо пропорційна кількості індексованих нею наукових видань та якості рецензування у процесі їхнього відбору. Внаслідок цього система, обмежена за національною ознакою, тематикою або з інших умов, не може претендувати на першість за якістю показників. На думку авторів [17], формування власної бази потребує значних фінансових витрат і має бути виправданим, тому забезпечення провідних наукових установ країни доступом до БД WoS CC або Scopus, збільшення публікацій учених з України в них є сьогодні актуальнішим. Інформетрики — лише формальні ознаки. Для коректного аналізу наукової публікації або дослідження обов'язково повинні враховуватися судження і висновки експертів.

Директор іспанської Cybermetrics Lab [19] за практичним призначенням розділяє індикатори на чотири категорії:

- бібліометрики: показники, пов'язані з офіційними виданнями в журналах та книгах (включаючи розділи, матеріали у процесі, тези або аналогічні статті);
- вебметрики: показники, отримані від веб-присутності, включаючи особисті або групові сторінки, веб-портали, (повнотекстові) документи в сховищах та інші комп'ютерні файли (програмне забезпечення, аудіо, відео тощо);
- альтметрики: показники, отримані з академічної або дослідницької інформації та розповсюджені соціальними веб-інструментами, включаючи блоги (мікроблоги), вікі та академічні (повні чи часткові) мережі обміну;
- метрики використання (usagemetrics): показники на початковому етапі розроблення — нові індикатори, які походять від кількості відвідувань, завантажень та інших даних щодо використання в Інтернет-просторі, документів з репозитаріїв, академічних порталів та веб-сайтів наукових проектів.

Основним фактором, який впливає на значення під час обчислення різних індексів, є джерела даних. Цитати, зареєстровані в конкретній БД, походять від посилань на документи, проіндексовані цією базою. Тому ширше охоплення джерел є одним з головних питань для підрахунку показників. У різних наукометричних БД значення h-індексу, а також інших показників відрізняються [20]. У порівнянні з WoS CC та Scopus система Google Scholar має найбільше охоплення документів і відповідно більші значення розрахованих показників. У деяких предметних галузях кількість цитат Google Scholar значно перевищує WoS CC або Scopus (які можливо порівняти) [21].

Підраховувати показники можна для будь-яких об'єктів: стаття, автор, видання, організація, географічний регіон, інше. Для статті основним показником є кількість цитувань. За деякими методиками враховується «вага» цитати — вона залежить від значення показника видання, в якому цитують статтю. Найбільш відомі показники видань:

- IF або JIF — імпакт-фактор, розраховується на основі даних за попередні два або більше років (Journal impact Factor);
- SJR — враховує «вагу» цитування, розраховується за попередні три роки (SCImago Journal Rank [22]);
- SNIP — нормалізований за напрямком дослідження, розраховується за попередні три роки (Source Normalized Impact per Paper);
- h-індекс — формується на основі даних за попередні роки;
- Eigenfactor — формується на основі даних за попередні п'ять років [23].

Класичний IF розраховується на основі даних з цитування наукометричної бази WoS CC для журналів за два попередні роки: $IF_{2016} = A/B$, де A — кількість цитувань статей, надрукованих протягом 2014–2015 рр. у журналах за 2016 р., B — загальна кількість статей, надрукованих за цей період.

Найчастіше для авторів та організацій розраховують кількість публікацій, їхніх цитувань та h -індекс (індекс Гірша). Він був визначений Х. Гіршем у 2005 р.; h -індекс дорівнює H , якщо H робіт із їхньої загальної кількості N цитують не менше H разів [17]. За таким принципом визначають h -індекс і для інших об'єктів. У табл. 2 наведено описи гіршеподібних індексів [24].

Існують також й інші додаткові показники: середнє цитування, індикатори з урахуванням самоцитування, індикатори екстремального цитування, яке має певні особливості у порівнянні з іншими, рангові індикатори, інше. Показники можуть бути абсолютні, відносні та нормалізовані. Кожна наукометрична БД визначає перелік показників, які розраховуються на основі власних даних. Для підрахунку показників додатково можна визначити проміжок часу та враховувати додаткові умови (кількість співавторів, «вага» джерела публікації, інше).

Враховуючи, що на основі значень наукометричних показників можна розв'язувати аналітичні завдання і приймати важливі рішення, технології та методології, які використовуються конкретною НС, повинні відповідати підвищеним вимогам та більш вагомих критеріям. Якість показників безпосередньо впливає на рівень достовірності даних, що, у свою чергу, впливає на успішну реалізацію прийнятого на їхній основі рішення.

Таблиця 2

Показник	Опис індексів	Властивості показників
g -індекс [25]	Максимальна кількість g найпопулярніших статей, які отримали разом не менше g^2 посилань. Враховує перевищення сумарного цитування ядра Гірша в порівнянні з мінімальними вимогами	$h \leq g$
hg -індекс [26]	Середнє геометричне h -індексу і g -індексу: $hg = \sqrt{h \times g}$	$h \leq hg \leq g$
h_α -індекс [27]	Індекс дорівнює h_α , якщо кожна з h_α публікацій має не менше $\alpha \times h_\alpha$ посилань, а кожна з решти — менше $\alpha \times h_\alpha$ цитувань, $\alpha \in \{1, 2, 3, \dots\}$	$h_1 = h$
h_I -індекс [28]	Індивідуальний індекс Гірша $h_I = h / N_A$, де N_A — середня кількість співавторів статей з ядра Гірша	$h_I \leq h$
e -індекс [29]	Квадратний корінь надлишкового цитування ядра Гірша $e = \sqrt{\sum_{j=1}^h c_j - h^2}$	—
A -індекс [30]	Середня кількість цитувань ядра Гірша $A = \frac{1}{h} \sum_{j=1}^h c_j$	—
R -індекс [31]	Квадратний корінь із сумарного цитування ядра Гірша $R = \sqrt{\sum_{j=1}^h c_j}$	$R = \sqrt{h \times A}$
AR -індекс [32]	Модифікація R -індексу, що враховує вік публікацій: $AR = \sqrt{\sum_{j=1}^h \frac{c_j}{a_j}}$, де a_j — вік j -ї публікації	$AR \leq R$
m -індекс [33]	Медіана кількості цитувань ядра Гірша	$h \leq m$
m_q -індекс [34]	Відносний індекс Гірша $m_q = h/y$, де y — вік першої статті автора	$h = m_q \times y$
i_{10} [18]	Кількість статей, кожна з яких отримала не менше десяти цитувань	—

Найважливіша властивість показників — достовірність, тобто властивість відображати дійсний стан речей. Існує низка факторів, що негативно впливають на достовірність (якість) наукометричних показників, яка має такі складові: аналітичну, змістовну, глобальне представлення. Для кожної складової визначимо перелік негативних чинників, які можна усунути повністю або частково у процесі формування інформаційної БД наукометричної системи.

На аналітичну складову негативно впливають: помилки під час перенесення метаданих (джерелом даних є сама публікація); неоднозначна ідентифікація суб'єктів та ін. Внаслідок цього спотворюється інформація про публікації, видання та їхнє цитування, а також авторів публікацій, організації, регіони, країни.

Існує низка технологій для забезпечення однозначної ідентифікації авторів та публікацій. Проблеми ідентифікації автора або документа призводять до викривлення даних щодо публікацій і цитування. Відомо багато випадків, коли статті зараховуються не тому авторові; профіль автора «розривається» на декілька за кількістю варіантів написання прізвища та імені (ППП). Для формування авторських метрик важлива коректна ідентифікація автора. Аналогічно показник з цитування не зараховується, коли посилення в переліку використаних джерел не збігається зі справжнім описом документа.

Однозначну ідентифікацію забезпечують такі інструменти:

- документа — цифровий ідентифікатор об'єкта (Digital Object Identifier, DOI), стандарт позначення представленої в мережі інформації про об'єкт;
- автора — унікальний ідентифікатор автора (Scopus AuthorID, Pubmed AuthorID, ResearcherID, ORCID, інші) [35].

ORCID (<http://orcid.org>) (Open Researcher and Contributor ID) — реєстр унікальних ідентифікаторів учених. ORCID унікальний і не залежить від наукових дисциплін і національних кордонів. Цей реєстр є методом ув'язування ідентифікаторів на різних платформах. Однак проблему однозначної ідентифікації текстового представлення об'єктів під час проведення наукометричних досліджень досі не розв'язано.

Якість змісту наукових робіт забезпечується оригінальністю та вимогами до публікацій видань, що включені до НС. Щоб підтвердити оригінальність роботи, можна виконати її перевірку на наявність у тексті подібностей та провести експертне оцінювання тексту для визначення авторського стилю.

Показники можна вважати достовірними, якщо їх визначають, ґрунтуючись на репрезентативній інформації. Репрезентативність залежить від правил формування вибірки, яка за критичними параметрами для дослідження повинна відтворити загальний об'єкт дослідження. Саме з цієї причини важливо забезпечити глобальне подання інформації (джерело). Детальну схему факторів, які негативно впливають на якість показників, наведено на рис. 3.

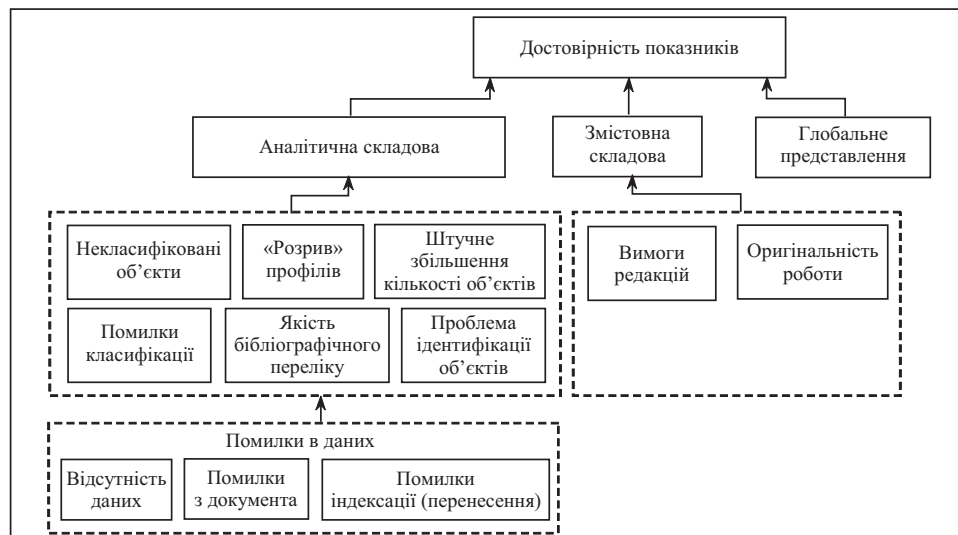


Рис. 3. Схема факторів, які впливають на складові НС

Серед інших напрямків інформетрії прикладом вебметричних досліджень є робота Cybermetrics Lab (Іспанська національна дослідницька рада, CSIC). Вона започаткувала Web або Webometrics, який у світі вважають найбільшим академічним рейтингом вищих навчальних закладів. На основі цих даних на сайті проекту публікують рейтингові переліки [36], які використовують у глобальних академічних рейтингах. Починаючи з 2004 р. ця лабораторія проводить незалежне та об'єктивне оцінювання академічних установ на основі їхньої присутності в інформаційному просторі Інтернету. Більшість університетів зараз підтримують «Рух Відкритого Доступу» (Open Access) до наукової інформації. Перше його визначення сформульовано на міжнародній Будапештській конференції з відкритого доступу (Budapest Open Access Initiative, BOAI) у 2002 р. Ідеологія відкритого доступу має на меті:

- реалізацію права автора щодо доведення інформації до наукової спільноти;
- створення в глобальному інформаційному просторі сегментів академічних ресурсів (репозитарії, видавництва, архіви відкритого доступу, журнали відкритого доступу);
- надання авторам доступу до достовірних ресурсів;
- консолідацію даних (директорії та агрегатори відкритих ресурсів);
- забезпечення захисту авторських прав.

Популяризація «Руху Відкритого Доступу» та розвиток Інтернет-технологій сприяли формуванню сегментів, які публікують в Інтернеті відкриті наукові дані. Учені активно обмінюються посиланнями та описами наукових даних у соціальних та професійних мережах і обговорюють викладені ідеї. Саме це сприяло появі нових альтернативних метрик — альтметрик. Вони не заміняють традиційні наукометричні показники, але є додатковими даними, які демонструють активний інтерес до інформації та потребують менше часу для їхнього отримання.

Альтметрика — наймолодша та одна з найбільш цікавих метрик. Альтметрики — це показники, які доповнюють традиційні показники на основі цитування в мережі Інтернет. Суть цієї методики оцінювання полягає у підрахунку згадувань про публікацію в Інтернеті: на офіційних сайтах наукових установ та університетів; у соціальних мережах типу Facebook, LinkedIn, Twitter тощо [37].

У світі розробляють системи, що враховують показники альтметрики, методи підрахунку, а також створюють нові метрики. Альтметрика має багато проблем, однак, враховуючи актуальні питання інформетрії та стрімку еволюцію наукового спілкування, швидкість та широту охоплення даних, вона є важливим напрямом для досліджень [7].

ВИСНОВКИ

Реалізація якісного представлення, інтеграції та адаптації інформації в інформаційному просторі в умовах прискореного розвитку досліджень наукометрії, вебметриї, альтметриї, інформетрії та забезпечення коректних даних для визначення наукометричних показників є актуальним та необхідним завданням для усієї наукової спільноти.

Огляд сучасного стану інформетрії показав, що для якісного аналізу даних необхідно більше застосовувати наукоємні методи та технології з різних галузей, такі, як Big Data, Machine Learning та Natural Language Processing. Використання цих технологій дозволить удосконалити методи оцінювання та аналізу даних на основі кількісних характеристик з урахуванням якісних характеристик об'єктів. Якісне оцінювання документа та його оригінальності може бути реалізоване шляхом комплексного аналізу текстових даних на лексичному, синтаксичному та семантичному рівнях. Таким чином, буде забезпечено підвищення об'єктивності методів наукометрії, вебметриї, альтметриї, інформетрії для оцінювання наукової інформації, наукової діяльності та ін.

СПИСОК ЛІТЕРАТУРИ

1. Акоев М.А., Маркусова В.А., Москалева О.В., Писляков В.В. Руководство по наукометрии: индикаторы развития науки и технологии (под ред. Акоева М.А.). Екатеринбург: Изд-во Урал. ун-та, 2014. 250 с.
2. Редькина Н.С. Библиометрия: история и современность. *Молодые в библиотечном деле*. 2003. № 2. С. 76–86.
3. Галыавиева М.С. Информетрия и библиотеки: точки пересечения. URL: <http://kpfu.ru/portal/docs/F1469809912/Galyavieva.MS.pdf>. 22.12.2017.
4. Гордукалова Г.Ф. Библиометрия, наукометрия и вебометрия — от числа строк в работах Аристотеля. *Научная периодика: проблемы и решения*. 2014. № 2 (20). С. 40–46.
5. Ефименко И. Можно ли измерить науку: философия, язык и культура современной наукометрии. URL: <https://www.hse.ru/mirror/pubs/share/150125175>. 14.01.2018.
6. Björneborn L., Ingwersen P. Towards a basis framework for webometrics. *Journal of the American Society for Information Science and Technology*. 2004. Vol. 55, N 14. P. 1216–1227.
7. Priem J., Taraborelli D., Groth P., Neylon C. Altmetrics: A manifesto. 26 Oct. 2010. URL: <http://altmetrics.org/manifesto>.
8. Priem J., Piwowar H.A., Hemminger B.H. Altmetrics in the wild: An exploratory study of impact metrics based on social media. URL: <http://jasonpriem.org/self-archived/PLoS-altmetrics-sigmetrics11-abstract.pdf>. 14.01.2018.
9. Pindlowa W. Wokół informetrii, bibliometrii i naukometrii. *Aktual. Probl. Inf. i Dok.* 1989. Vol. 34, N 1–2. P. 3–7.
10. Асеев Г.Г. Соотношение различных метрических исследований в науковедении. *Системы обработки інформації*. 2017. Вип. 1 (147). С. 119–126.
11. Galyavieva M.S. On the formation of the concept of informetrics. *Scientific and Technical Information Processing*. 2013. Vol. 40, N 2. P. 89–96. DOI: 10.3103/S014768821302007X.
12. Хайтун С.Д. Наукометрия: состояние и перспективы. Москва: Наука, 1983. 344 с.
13. Чайковський Ю.Б., Сілкіна Ю.В., Потоцька О.Ю. Наукометричні бази та їх кількісні показники. Частина I. Порівняльна характеристика наукометричних баз. *Вісн. НАН України*. 2013. № 8. С. 89–98.
14. Российский индекс научного цитирования. URL: https://elibrary.ru/projects/citation/cit_index.asp.
15. Index Copernicus. URL: <https://indexcopernicus.com>. 14.01.2018.
16. Microsoft Academic. URL: <https://academic.microsoft.com>. 14.01.2018.
17. Чайковський Ю.Б., Сілкіна Ю.В., Потоцька О.Ю. Наукометричні бази та їх кількісні показники. Частина II. Фактори, що впливають на кількісні показники наукометричних баз. *Вісн. НАН України*. 2013. № 9. С. 84–92.
18. Google Scholar. URL: <http://scholar.google.com>.
19. Aguillo I.F. Informetrics for librarians: Describing their important role in the evaluation process. *El Profesional de la Información*. 2016. Vol. 25, N 1. P. 5–10.
20. Bar-Ilan J. Which h-index? A comparison of WoS, Scopus and Google Scholar. *Scientometrics*. 2008. Vol. 74, N 2. P. 257–271.
21. Peters I. Informetrics, bibliometrics, altmetrics: What is it all about? *Proceedings of the American Society for Information Science and Technology*. 2014. Vol. 51, Iss. 1. P. 1–4.
22. Scimago Journal & Country Rank. URL: <http://www.scimagojr.com>. 14.01.2018.
23. Eigenfactor: About. URL: <http://www.eigenfactor.org/about.php>. 14.01.2018.
24. Штовба С.Д., Штовба Е.В. Обзор наукометрических показателей для оценки публикационной деятельности ученого. *Управление большими системами. Спец. вып. «Наукометрия и экспертиза в управлении наукой»*. 2013. № 44. С. 262–278.
25. Egghe L. Theory and practice of the g-index. *Scientometrics*. 2006. Vol. 69, N 1. P. 131–152.
26. Alonso S., Cabrerizo F., Herrera-Viedma E., Herrera F. hg-index: a new index to characterize the scientific output of researchers based on the hand g-indices. *Scientometrics*. 2010. Vol. 82, N 2. P. 391–400.
27. Eck N.V., Waltman L. Generalizing the h- and g-indices. *Journal of Informetrics*. 2008. Vol. 2, N 4. P. 263–271.
28. Schreiber M. A modification of the h-index: The h(m)-index accounts for multi-authored manuscripts. *Journal of Informetrics*. 2008. Vol. 2, N 3. P. 211–216.
29. Zhang C.T. The e-index, complementing the h-index for excess citations. *PLoS ONE*. 2009. Vol. 4, N 5. URL: <http://www.plosone.org/article/info%3Adoi%2F10.1371%2Fjournal.pone.0005429>.
30. Rousseau R. New developments related to the Hirsch index. *Science Focus*. 2006. Vol. 1, N 4. P. 23–25. URL: <http://eprints.rclis.org/6376/>.

31. Liang B.J.L., Rousseau R., Egghe L. The R- and AR-indices: complementing the h-index. *Chinese Science Bulletin*. 2007. Vol. 52, N 6. P. 855–863.
32. Jin B. The AR-index: complementing the h-index. *ISSI Newsletter*. 2007. Vol. 3, N 1. P. 6.
33. Bornmann L., Mutz R., Daniel H. Are there better indices for evaluation purposes than the h-index? A comparison of nine different variants of the h-index using data from biomedicine. *Journal of the American Society for Information Science and Technology*. 2008. Vol. 59, N 5. P. 830–837.
34. Hirsch J.E. An index to quantify an individual's scientific research output. *Proc. National Academy of Sciences of the USA*. 2005. Vol. 102, N 46. P. 16569–16572.
35. Главчева Ю.М., Кухаренко В.М., Рибалко О.В. Куратор змісту (за ред. Кухаренка В.М.). Харків: НТУ «ХПІ», 2016. 157 с.
36. Welcome to Ranking Web of Universities. URL: <http://www.webometrics.info/en>. 14.01.2018.
37. What are altmetrics? URL: <https://www.altmetric.com/about-altmetrics/what-are-altmetrics/>. 14.01.2018.

Надійшла до редакції 07.02.2018

Ю.Н. Главчева, О.В. Канищева, Н.В. Борисова
ОБЗОР МЕТОДОВ И ТЕХНОЛОГИЙ ИНФОРМЕТРИИ

Аннотация. Приведен обзор методов и технологий информетрии. Определены ее задачи и направления. Показана взаимосвязь таких понятий, как наукометрия (scientometrics), библиометрия (bibliometrics), информетрия (informetrie), вебометрия (webometrics), альтметрия (altmetrics). Проанализированы существующие модели и методы информетрии, а также ее актуальные проблемы. Рассмотрены аналитически-поисковые наукометрические базы данных и системы, выявлены их недостатки и преимущества. Выделены характеристики наукометрических систем, их составляющие и факторы, влияющие на наукометрические индексы. На основе проведенного исследования сформулированы перспективные направления развития наукометрических систем.

Ключевые слова: наукометрия, информетрия, наукометрические системы, библиометрические системы, метрики, индексы цитирования.

Yu.M. Hlavcheva, O.V. Kanishcheva, N.V. Borysova
SURVEY OF INFORMETRY METHODS AND TECHNOLOGIES

Abstract. The paper overviews the methods and technologies in informetry. The authors define the tasks and directions of informetry, show the interaction of terms such as scientometrics, bibliometrics; informetry, webometrics and altmetrics. Available informetry models, methods and current problems in a rapidly developing information society are analyzed. The authors review the available analytic-and-search science-based databases and systems, identify their disadvantages and advantages. As a result of the analysis, the characteristics of scientometric systems, their components, and factors that affect scientometric indices have been identified. On the basis of the study, the authors show the perspective directions of development of scientometric systems.

Keywords: scientometrics, informetry, scientometric systems, bibliometric systems, metrics, citation indexes.

Главчева Юлія Миколаївна,
 заступник директора науково-технічної бібліотеки Національного технічного університету «Харківський політехнічний інститут», e-mail: glavcheva@khpі.edu.ua.

Канищева Ольга Валеріївна,
 кандидат техн. наук, доцент кафедри Національного технічного університету «Харківський політехнічний інститут», e-mail: kanichshevaolga@gmail.com.

Борисова Наталя Володимирівна,
 кандидат техн. наук, доцент кафедри Національного технічного університету «Харківський політехнічний інститут», e-mail: borysova.n.v@gmail.com.