

МІЖНАРОДНИЙ ДОСВІД ОЦІНЮВАННЯ НАУКОВОГО ВНЕСКУ ЗА ДОПОМОГОЮ АНАЛІЗУ ЦИТУВАННЯ

У статті представлено огляд методів та інструментаріїв оцінювання внеску вчених, які спираються на аналіз цитування наукових робіт. Визначені переваги кожного інструменту та загальні недоліки аналізу цитування. Окреслені перспективи розвитку методів аналізу цитування. Підкреслена важливість відповідних занять у підготовці та перепідготовці науковців та менеджерів.

Ключові слова: оцінка наукового внеску; наукометричні бази даних; наукові пошукові системи; аналіз цитування; h-індекс.

Постановка проблеми. Розквіт сучасної світової науки супроводжується експоненціальним зростанням кількості дослідників та відповідним приростом бюджету. Від 80 до 90 % вчених, які існували коли небудь, все ще живуть [1]. Це створює нові виклики щодо ефективного розподілу людських та фінансових ресурсів. Тому необхідність удосконалення методів оцінки якості роботи вчених та наукових установ є актуальною проблемою.

Проблема визначення оптимальних напрямів підтримки науки, передусім, існує на державному рівні. Міністерства та державні фінансові інституції США використовують дані аналізу цитування для оцінки якості роботи професорів та дослідників під час призначення посад, видачі наукових грантів тощо [2]. Європейські закони також вимагають застосування незалежної системи оцінки наукової продуктивності дослідників [3]. Раціональний розподіл фінансів в окремих установах так само вимагає певної кваліфікації у володінні сучасними інструментами оцінки наукового внеску [4].

Перед вищою школою та наукою постають дві важливі задачі. По-перше, необхідно удосконалювати методи оцінки ефективності наукової діяльності; по-друге, розвивати навчальні програми щодо вивчення таких методів.

Аналіз досліджень та публікацій Кількість публікацій уже давно не є критерієм оцінки професійного рівня науковця. Залежно від наукової галузі від 50 % до 90 % статей, які опубліковані в академічних журналах, ніколи не цитуються, їх ніхто не читає, за винятком референтів та редакторів журналів [2].

Для визначення професійного рівня науковця використовують два підходи – експертні оцінки та аналіз цитування, причому останній перемагає зі зростаючим у часі відривом через швидкість застосування та дешево використання web-орієнтованих ресурсів і сервісів.

Основним інструментом аналізу цитування є наукометричні бази даних наукових публікацій. Існують

дві альтернативних групи таких баз даних: з одного боку реферативні бази, такі як Scopus та Web of Science, які індексують лише публікації з обмеженого списку журналів або інших джерел, причому списки періодично уточнюються експертами. З іншого – такі як Google Scholar або Microsoft Academic, що автоматично збирають інформацію з веб-простору за допомогою потужних пошукових систем.

Перша група спирається на комбіновану експертно-пошукову технологію, тоді як друга фактично є різновидом пошукової системи з доданими інструментами класифікації, аналізу та широким використанням автоматичних алгоритмів ідентифікації наукового змісту. Друга група демонструє потенціал для більш потужного вимірювання наукового впливу шляхом включення різноманітних наукових публікацій, наприклад, тез, доповідей на конференціях, тобто цитування більш широкого кола джерел. З іншого боку, автоматичне розпізнавання та виділення наукового змісту може призвести до проблем надійності та довіри внаслідок дублювання записів або неправильно визначених метаданих [1].

Аналіз цитат є також механізмом обчислення фактора впливу журналів – показника, який дослідники беруть до уваги при вирішенні того, в який журнал подавати свою роботу так, щоб збільшити кількість читачів. Аналіз цитування часто є єдиним способом для того, щоб неспеціалісти з урядів, фінансуючих організацій або навіть представники інших наукових дисциплін змогли визначити важливість наукового дослідження та його результатів [2].

Протягом десятиріч Web of Science від Thomson Reuters була єдиною міжнародною мультидисциплінарною базою даних. Однак у 2004 р. Elsevier представив Scopus, який став якісною альтернативою. Web of Science – це база даних цитувань і публікацій, яка охоплює усі галузі науки протягом багатьох років. Web of Science і Scopus доступні далеко не всім науковцям. Необхідність сплачувати за Scopus або Web

of Science створює певну елітарність у їх використанні [5].

Конкуренція між базами даних Web of Science і Scopus є інтенсивною і призводить до поліпшення послуг, які вони пропонують [6]. Web of Science забезпечує покриття до 1900 р., а Scopus контролює лише відносно свіжі джерела, повільно додаючи інформацію про цитування статей старших за 90-ті роки минулого сторіччя. У багатьох випадках Scopus забезпечує краще висвітлення ніж Web of Science [7]. Він містить багато корисних функцій, таких як пошук публікацій за допомогою булевих операцій, що робить його дуже точним для збирання даних про цитування та ідентифікацію авторів. З іншого боку, використання таких функцій ускладнює процедуру пошуку і може відштовхувати користувачів [5].

Порівняння Web of Science та Scopus за різними категоріями оцінки таких як: зміст, доступ, послуги, інтерфейс, пошук, результати пошуку, витрати, аналітичні інструменти показали, що Scopus та Web of Science наразі є достойними конкурентами. Кожний з них має переваги в окремих категоріях [8].

Реферативні бази даних цитування завжди критикувалися вченими за те, що вони індексують обмежену кількість назв журналів, охоплюють переважно англomовні джерела з Північної Америки та Західної Європи, а також ігнорують цитати з книг та більшості конференцій. Інша проблема полягає в тому, що вони не враховують те, що вчені частіше публікують статті в Інтернеті через журнали відкритого доступу, персональні сторінки, сервери електронного друку або інституційні репозитарії, щоб інші могли вільно отримувати доступ до матеріалу [2].

Кількість цитованих публікацій враховується під час прийняття рішень, пов'язаних з працевлаштуванням, просуванням по службі та перебуванням на посаді. Адміністратори вивчають публікації та цитати, щоб оцінити спільний внесок установи чи відділу. Для визначення рейтингів та пріоритетів були запропоновані інтегральні показники, що спираються на аналіз цитування.

Найбільш популярним показником є h-індекс. Цей критерій було запропоновано у 2005 р. Хорхе Хіршом в Університеті Каліфорнії в Сан-Дієго для оцінки впливу особи або рейтингу установи [9]. Як і інші показники, отримані з аналізу цитування, h-індекс має використовуватися з обережністю. Це пояснюється тим, що індекс ігнорує в якому контексті цитувався документ; наприклад, додає бали за негативні відгуки. Також він усереднює цитати і тому нечутливий до унікальних робіт, на які існує велика кількість посилань. Ці недоліки призвели до розробки альтернативних індексів, наприклад, a-index, який визначається як середнє число цитат, отриманих для публікацій, що включені в h-index. Іншою альтернативою є g-index, що розраховується як число g робіт, які разом отримали g^2 або більше цитат. Наприклад, дослідник отримає g-індекс 10, якщо опублікує 10 статей, які разом цитувалися принаймні 100 разів [2].

Одним з недоліків реферативних баз дослідники відзначають американське домінування на Web of Science та європейське на Scopus. Навпаки, Google

Scholar та Microsoft Academic, як правило, представляють географічно нейтральні результати [10].

У соціальних та гуманітарних науках, де книги і монографії є однією з основних платформ для наукових досліджень, Web of Science і Scopus відомі як слабкі [11]. Водночас Google Scholar показала себе надійною і може бути достатньою для оцінки досліджень. Більш того, у цьому випадку може бути використаний інший пошуковий інструмент Google Books, особливо враховуючи, що він має власний сервіс API. Це дозволяє уникнути ручного пошуку та перевірки великої кількості поданих матеріалів. Такий підхід може автоматично створювати цитування та цитовані асоціації між книгами на основі витягу бібліографічної інформації та цитованих посилань на книги. Google Books, у цьому випадку, може бути самодостатнім та корисним ресурсом. В інших випадках рекомендується використовувати комбінацію Google Scholar і Scopus [12].

Google Scholar є безкоштовною пошуковою системою пошуку наукових публікацій, таких як книги, статті, препринти. Використання Google Scholar рекомендовано, передусім, для пошуку нетрадиційних форм публікацій [3]. Він має ширший за інших пошук, але не завжди чітку межу між науковим і звичайним пошуком. Тобто це потужний інструмент, але такий, що вимагає навичок [13]. Якщо вчений хоче визначити перелік посилань на власні роботи, що містяться в книгах або розділах книг, тоді Google Scholar або Google Books будуть оптимальним вибором. Коли необхідно задокументувати свої цитати, що містяться в рецензованих статтях, Web of Science і Scopus мають вичерпне висвітлення академічних журналів, але обмежене охоплення книг, дисертацій або збірників [4, с. 14].

Перспективною альтернативою Google Scholar є Microsoft Academic. Це фактично портал, побудований на Microsoft Academic Graph, який є завантажуваним, безкоштовним додатком. Microsoft Academic офіційно був запущений у березні 2016 р. і є наступником проекту Microsoft Academic Search, який припинено у 2012 р. Останні версії portalу інтегровані з пошуковою системою Bing [1].

На відміну від Google Scholar, де доступ користувача до внутрішньої організації даних відсутній, Microsoft Academic представляє дані у вигляді відкритих структурованих метаданих, що полегшує пошук та обробку. Дані в основному збираються з каналів видавців і веб-сторінок, індексованих Bing. Крім того, пропонується сервіс API, який дозволяє дуже точно провести їх аналіз. Користувачі можуть отримувати необроблені метадані, а також обробляти їх за допомогою API. Такий сервіс є недоступним для Google Scholar, він значно полегшує пошук та обробку даних, що є основною перевагою Microsoft Academic над Google Scholar. Однак існують обмеження щодо доступних метаданих. Microsoft Academic не надає тип документа, області дослідження в структурі даних є занадто специфічними, а також існують помилки в роках публікації та кількості авторів [15].

Microsoft Academic, за оцінкою дослідників, знаходить більше робіт, ніж Scopus і Web of Science і менше, ніж Google Scholar [16]. Однак недоліком є те,

що разом з охопленням більш широкого кола дослідницьких робіт (наприклад, книг), Microsoft Academic, як і Google Scholar, також включають так звані бездомні публікації, тобто публікації, які є дублікатами інших публікацій, але з дещо іншим варіантом назви або автора. Microsoft Academic прогресує, має потенціал для розвитку та представляє хорошу альтернативу для аналізу цитування.

Постановка завдання Метою цієї роботи є узагальнення досвіду використання та сучасних тенденцій розвитку методів та засобів оцінки наукового внеску дослідників або установ. Стаття досліджує досвід інших країн, який може бути корисним для підготовки майбутніх менеджерів, що спеціалізуються в управлінні науковою діяльністю.

Виклад основного матеріалу Показники цитування все частіше використовуються для підтримки експертної оцінки наукового внеску як наукових підрозділів, так і окремих дослідників. Тим не менш, традиційні методи аналізу цитування можуть бути неадекватними [12]. Разом з безумовними перевагами цей підхід має суттєві загальні недоліки, такі як:

- можливість так званих гомографів, тобто об'єднання посилань на двох різних вчених, які мають одне прізвище та ініціали [2]. Для запобігання такому явищу Scopus та Web of Science змушені вимагати, щоб автори перевіряли, чи система правильно приписує авторство [3];

- провокування самоцитування та кронізму, коли колеги домовляються цитувати один одного, щоб підвищити рівень цитування [2];

- церемоніальне цитування, коли автор посилається на авторитетних науковців або журнали безвідносно до змісту роботи. Багато людей цитують роботи зі стратегічних, а не принципових причин. Як наслідок, показники цитування не надають точної оцінки внеску науковця [2, с. 17];

- негативне цитування, коли посилання вказують на неправильні результати, але внаслідок відсутності механізму виявлення їх неможливо відрізнити від інших [2];

- якщо автор публікує статті зі співавторами, це як правило, збільшує кількість посилань на такі статті. Обчислення цитат не враховує кількість співавторів. Якщо, наприклад, один раз наводиться стаття з п'ятьма авторами, кожен з них отримує одне цитування, а не одну п'яту частину цитати [17];

- дослідники часто публікують статті в журналах з інших дисциплін, що унеможливує однозначну оцінку внеску науковця [18], і навпаки, багато дослідників посилаються на документи інших наукових напрямів, що може ігноруватися при розрахунку бібліометричних показників [19];

- нарешті, впровадження аналізу цитат може суттєво стримувати розвиток науки. На кількість цитування впливає кількість дослідників, які працюють у сферах, пов'язаних зі статтею [18]. Досліднику вигідніше опублікувати ще одну роботу на тему, в якій працює велика група вчених (тому що вони цитують один одного), ніж відкривати нове поле досліджень, де просто немає кому цитувати [17].

Окрім науковців, аналізуючи динаміку цитування за роками, доходять висновку, що цитування нових статей не дає точної оцінки наукового внеску. Дина-

міка цитування така, що лише з часом виявляється реальне значення роботи. Це протирічить сучасним тенденціям враховувати лише свіжі публікації. Особливо це стосується революційних наукових робіт в нових галузях.

Існує багато прикладів, коли видатні вчені практично не отримували відгуку від колег в перші роки після публікації. Основні роботи геніального американського математика, лауреата Нобелівської премії Джона Неша, який зробив значний внесок у розвиток диференціальної геометрії та теорії ігор, отримали лише 16 цитат у перші п'ять років після публікації [17].

Перераховані недоліки викликають необхідність пошуку більш надійних методів оцінки наукового внеску. Покладання виключно на наукометричні бази не надає точних оцінок та не сприяє широкому розповсюдженню нових знань. Дослідники самі оскаржують корисність таких кількісних показників для оцінки наукової роботи, підкреслюючи той факт, що зловживання та недосконало розроблені показники можуть мати шкідливий вплив на дослідження [1].

Сьогодення вимагає змін в існуючій елітарній системі наукової комунікації на основі рецензованих статей, опублікованих у журналах. Склалася тенденція, коли вчені починають протестно поширювати результати досліджень на якомога більшій кількості платформ і в максимально доступній кількості способів. Наприклад, вони публікують власні роботи у відкритих журналах, інституційних репозиторіях, на персональних сторінках та хмарних сервісах, бо хочуть, щоб їхні колеги були обізнані, обговорювали та використовували результати досліджень, і в кінцевому підсумку цитували авторів [2].

Перехід наукової комунікації в онлайн-форму дає змогу збирати дані щодо використання публікацій іншими дослідниками в безпрецедентному масштабі. Ці дані можуть бути використані для вивчення розвитку як окремих наукових напрямів, так і науки загалом.

Науковці закликають до універсальної, відкритої та сумісної стандартизованої інфраструктури даних, розробки відповідних показників, стандартів інфраструктури даних та відповідної семантики. Це збільшить інтеперабельність джерел даних, масштаби та надійність показників. Зібрані дані повинні враховувати дисциплінарні відмінності (дисциплінарна нормалізація), можливість перевірки даних та прозорість процедури і показників [1].

Традиційний аналіз цитування припускає, що всі цитати мають однакову цінність і вагу, тобто фактично ігнорується зміст. Сучасні методи штучного інтелекту створюють можливість визначати цінність цитування, інтерпретуючи кожен цитату на основі її контексту як на синтаксичному, так і на семантичному рівнях. Це створює безпрецедентні можливості визначити, яким чином розвиваються знання [20].

Через вибухове збільшення наукових публікацій стало практично неможливо переглянути всі роботи, які б цікавили науковця. Фактично йому залишився доступним рандомний пошук, який назавжди дає очікуваний результат. Технології обробки великих даних дозволяють відстежувати передачу, зростання та трансформацію знань від одного дослідника до іншого шляхом аналізу цитування їхніх робіт.

Наративний текст, який був віками найефективнішим способом передачі наукової інформації, перетворюється в одиниці знань, концепції, наукові декларації, систему контекстно пов'язаних тверджень. Це в свою чергу вимагає нових способів написання та цитування. Авторизація однієї одиниці знань може привести до посилання на пов'язані блоки знань з цієї бази знань. Отже, кожна цитата має стати не просто текстом, а інтерфейсом до глобальної системи знань.

Висновки і перспективи подальших розвідок. Існує нагальна потреба в освоєнні науковцями та менеджерами від освіти операцій і функцій з пошуку цитування в різних наукометричних базах не лише для визначення наукового рейтингу, але й для отримання права на посаду, просування по службі, подан-

ня заявок на отримання грантів, керування науковими проектами, проведення консультацій тощо.

У навчальну програму підготовки та перепідготовки науковців і професійних менеджерів потрібно додавати курси, що надають знання з аналізу цитування як інструменту дослідження у визначенні основних тенденцій розвитку науки, стратегій і методів управління, а також прикладної психології, прийняття експертних рішень та аналізу наукової продуктивності.

Через тенденцію побудови відкритого універсального середовища обміну науковими ідеями, заснованого на онлайн платформах, досягненнях штучного інтелекту і технологій обробки великих даних, попереду нас очікують кардинальні зміни в аналізі цитування. Тому науковці та менеджери приречені вчитися цитуванню впродовж всього життя.

Список використаних джерел

1. Paszcza B. Comparison of Microsoft Academic (Graph) with Web of Science, Scopus and Google Scholar. (Master's Thesis), University of Southampton. – 2016. <http://eprints.soton.ac.uk/id/eprint/408647>.
2. Meho L. I. The rise and rise of citation analysis. // *Physics World*. – 2007. – Т. 20 (1). p. 32–36.
3. Chirici G. Assessing the scientific productivity of Italian forest researchers using the Web of Science, SCOPUS and SCIMAGO databases // *iForest – Biogeosciences and Forestry*. – 2012. – Т. 5 (3). p. 101–107.
4. Piotrowski C. Bibliometrics and citation analysis for the psychologist–manager: A review and select readings // *The Psychologist–Manager Journal*. – 2013. – Т. 16 (1). – p. 53–71.
5. Ahmar A. S., Kurniasih N., Irawan D. E., Sutiksno D. U., Napitupulu D., Setiawan M. I., Simarmata J., Hidayat R., Busro, Abdullah D., Rahim R., Abraham J. Lecturers' Understanding on Indexing Databases of SINTA, DOAJ, Google Scholar, SCOPUS, and Web of Science : A Study of Indonesians // *J. Phys. Conf. Ser.* – 2018. – Т. 954, № 1. – p. 12026.
6. Aghaei Chadegani A., Salehi H., Yunus M. M., Farhadi H., Fooladi M., Farhadi M., Ale Ebrahim N. A comparison between two main academic literature collections : Web of Science and Scopus databases. // *Asian Social Science*. – 2013. – Т. 9 (5). – p. 18–26.
7. Li J., Burnham J. F., Lemley T., Britton R. M. Citation analysis : Comparison of Web of Science, Scopus, SciFinder, and Google Scholar // *Journal of electronic resources in medical libraries*. – 2010. – Т. 7 (3). – p. 196–217.
8. Adriaanse L., Rensleigh C. Comparing Web of Science, Scopus and Google Scholar from an environmental sciences perspective // *South African Journal of Libraries and Information Science*. – 2011. – Т. 77 (2). – p. 169–178.
9. De Groote S. L., Raszewski R. Coverage of Google Scholar, Scopus, and Web of Science : A case study of the h-index in nursing. // *Nursing Outlook*. – 2012. – Т. 60 (6). – p. 391–400.
10. Shah S., Mahmood K., Hameed A. Review of Google Scholar, Web of Science, and Scopus search results: The case of inclusive education research. // *Library Philosophy and Practice (e-journal)*. – 2017. <http://digitalcommons.unl.edu/libphilprac/1544>.
11. Martín-Martín A., Orduna-Malea E., Thelwall M., Delgado López-Cózar E. Google Scholar, Web of Science, and Scopus : a systematic comparison of citations in 252 subject categories // *Journal of Informetrics*. – 2018. – Т. 12 (4). – p. 1160–1177.
12. Kousha K., Thelwall M., Rezaie, S. Assessing the citation impact of books : The role of Google Books, Google Scholar, and Scopus. // *Journal of the American Society for Information Science and Technology*. – 2011. – Т. 62 (11). – p. 2147–2164.
13. Badke W. B. Coming Back to Google Scholar. // *Online Searcher*. – 2013. – Т. 37 (5). – p. 65–67.
14. Piotrowski C. Citation analysis for the modern instructor : An integrated review of emerging research. // *Journal of Educators Online*. – 2013. – Т. 10 (2). – p. 1–24.
15. Hug S. E., Ochsner M., Brändle M. P. Citation analysis with Microsoft Academic. // *Scientometrics*. – 2017. – Т. 111 (1). – p. 371–378.
16. Harzing A. W., Alakangas, S. Microsoft Academic : Is the Phoenix getting wings? // *Scientometrics*. – 2017. – Т. 110 (1). – p. 371–383.
17. Jensenius F., Htun M., Samuels D., Singer D., Lawrence A., Chwe M. The Benefits and Pitfalls of Google Scholar. // *Political Science & Politics*. – 2018. – Т. 51 (4). – p. 820–824.
18. Xiao B., Cheung C., Thadani D. Assessing the Quality and Knowledge Contribution of Top IS Journals : A Comparative Citation Analysis. // *ECIS 2011 Proceedings*. – Helsinki. – 2011. – 227 p.
19. Patience G. S., Patience C. A., Blais B., Bertrand F. Citation analysis of scientific categories. // *Heliyon*. – 2017. – Т. 3. – № 5. – e00300.
20. Ding Y., Zhang G., Chambers T., Song M., Wang X., Zhai, C. Content-based citation analysis : The next generation of citation analysis. // *Journal of the Association for Information Science and Technology*. – 2014. – Т. 65 (9). – p. 1820–1833.

В. В. Ромакин,

ЧНУ ім. Петра Могили, г. Николаев, Україна

МЕЖДУНАРОДНЫЙ ОПЫТ ОЦЕНКИ НАУЧНОГО ВКЛАДА С ПОМОЩЬЮ АНАЛИЗА ЦИТИРОВАНИЯ

В статье представлен обзор методов и инструментариев оценки вклада ученых, основанных на анализе цитирования научных работ. Определены преимущества каждого из инструментов и общие недостатки анализа цитирования. Обозна-

чены перспективы развития методов анализа цитирования. Подчеркнута важность соответствующих занятий в подготовке и переподготовке ученых и менеджеров.

Ключевые слова: оценка научного вклада; наукометрические базы данных; научные поисковые системы; анализ цитирования; h-индекс.

V. Romakin,

Petro Mohyla Black Sea National University, Mykolayiv, Ukraine

INTERNATIONAL EXPERIENCE IN ACADEMIC IMPACT EVALUATION BASED ON CITATION ANALYSIS

The paper presents a review of academic impact evaluation tools and techniques based on citation analysis. The advantages of each tool and the general disadvantages of citation analysis are identified. The prospects for the development of citation analysis are outlined. It emphasizes the importance of including academic impact evaluation tools and techniques in the scientists and managers training or retraining.

Key words: academic impact evaluation; scientometric databases; academic search engines; citation analysis; h-index.

Рецензенти: Сисоева С. О., д-р пед. наук, професор;
Мещанинов О. П., д-р пед. наук, професор.

© Ромакин В. В., 2019

Дата надходження статті до редколегії 29.05.2019