
К.В. Лобузiна

Нацiональна бiблiотека України iменi В.І. Вернадського



БІБЛІОТЕЧНІ ТА СЕМАНТИЧНІ ВЕБ-ТЕХНОЛОГІЇ В ОРГАНІЗАЦІЇ ДОСТУПУ ДО НАУКОВОЇ ПЕРІОДИКИ

Публікацію присвячено особливостям організації та перспективам розвитку інтегрованого інформаційного ресурсу наукової періодики Національної бібліотеки України імені В.І. Вернадського. Проаналізовано та визначено переваги для видавництва і авторів розміщення повних текстів наукових публікацій періодичних фахових видань у централізованому репозиторії відкритого доступу. Розглянуто сучасні можливості семантичних бібліотечних і веб-технологій для визначення наукометричних показників учених та наукових журналів.

Ключові слова: наукова періодика, централізований репозиторій, наукова інформація, семантичні бібліотечні та веб-технології, наукометричні показники.

Ресурси електронної періодики набули надзвичайної популярності серед наукової спільноти та провідних видавців наукової продукції у всьому світі. Це пов'язано з тим, що електронна форма наукової інформації виконує притаманні їй функції набагато ефективніше, ніж паперова, а саме: оперативне інформування наукового співтовариства про здобутки досліджень; включення публікацій у загальну систему обміну науковою інформацією; презентація автора наукових досліджень у наукових комунікаціях; визначення рейтингу наукових публікацій та їх авторів у загальній системі наукового знання (індекси цитування).

Наукова публікаційна активність учених сьогодні також є важливим критерієм оцінювання ефективності наукової роботи і діяльності наукової установи в цілому, одним із показників доцільності надання фінансування на проведення наукових досліджень [4, 5]. Відповідно до наказу МОН України «Про

опублікування результатів дисертацій на здобуття наукових ступенів доктора і кандидата наук» (2012 р.) рівень наукових публікацій за темою дисертаційного дослідження визначається тим, що вони опубліковані у наукових фахових виданнях, які включені до міжнародних наукометричних баз.

Функції міжнародних наукометричних баз виконують великі інтегратори (видавничі консорціуми) електронних періодичних видань *Elsevir*, *EBSCO-Host*, *ProQuest* тощо, які організують масиви реферативно-бібліографічної та повнотекстової наукової інформації. Реферативні бази даних мають професійний пошуковий інтерфейс, який забезпечує ефективний моніторинг наукових публікацій у заданому тематичному сегменті; виконання складних пошукових процедур; аналітичні процедури бібліометрії та наукометрії, необхідні для визначення тенденцій розвитку наукового знання [1]. Однак більшість із зазначених послуг для видавців наукових періодичних видань є не безкоштовними. Так, однією із необхідних вимог для включення наукового періодичного видання до наукометричних баз даних є наявність власного сайту наукового фахового видання та необхідність мати сплачений індекс *DOI* для кожної публікації [6, 7]. Система індексів *DOI* підтримується агентством *CrossRef*, яке забезпечує механізми зв'язку між посиланням на публікацію та її місцем у мережі Інтернет. Індекс *DOI* однозначно зв'яже із публікацією набір метаданих, до яких належать такі необхідні ідентифікатори: назва статті, прізвище та ініціали автора (авторів), рік видання, журнал, випуск, сторінки тощо.

Як альтернатива визнаним комерційним професійним наукометричним системам останнім часом швидко розвиваються наукові соціальні мережі та системи електронної періодики відкритого доступу. Ці системи мають більш демократичний характер та є загальнодоступними. За умов коректного представлення наукових публікацій у системах відкритого доступу, вони можуть сьогодні завдяки розвитку інструментів семантичного вебу надати надзвичайно корисну інформацію науковим дослідникам. Ідеї моделі семантичного вебу (*Web 3.0*) ґрунтуються на тому, що додавання певної структури даних (метаданих) неструктурованому контенту глобальної мережі робить його більш зрозумілим пошуковим роботам, що у свою чергу дає можливість перетворити ресурси Інтернету в аналог бази даних [3].

Необхідною передумовою об'єктивного визначення рейтингів, обчислення наукометричних та бібліометричних показників є повнота представлення публікаційної активності вчених. Вирішення проблеми підтримки сайтів окремих фахових видань відповідно до міжнародних вимог не може професійно вирішити питання організації інтегрованого науково-інформаційного сервісу, який би забезпечував дослідників необхідним рівнем пошукових можливостей та наукометричним інструментарієм. Підтримувати таку платформу професійно можуть лише спеціалісти з інформаційного менеджменту наукових ресурсів. Тому, на нашу думку, варто розмежувати ці два завдання, бо вирішити їх одночасно, застосовуючи однакові інформаційно-технологічні рішення, як показує міжнародний досвід,



ефективно неможливо. Створення окремих сайтів наукових періодичних видань кожним видавництвом або науковою установою призводить до розпорошеності наукових публікацій та необхідності мати централізований агрегатор цих ресурсів, який повинен забезпечувати інтегрований пошук та аналіз інформації в межах всього інформаційного масиву. Агрегація розрізнених інформаційних ресурсів може здійснюватись двома основними способами: 1) централізоване збирання метаданих корпоративним агрегатором та надання доступу до повних текстів через сайти окремих видань або 2) збирання та опрацювання повнотекстових ресурсів централізованим репозиторієм.

Перша модель, як показує світовий досвід, є ефективною лише за умов, коли окремі сайти видавництва мають типові програмні забезпечення, яке дозволяє в автоматичному режимі постачати метадані наукових публікацій до агрегатора. Це вимагає від кожного видавництва впровадження відповідної інформаційної платформи для свого видання на основі безкоштовно доступних програмних систем аналогічних *Open Journal* або *Dspace*, які мають готові механізми автоматичного агрегування метаданих через протокол *OAI-PMH*. У тому числі, цей протокол дає змогу інтегрувати бібліографічні дані до інформаційних сервісів *Google Scholar*. Правильне постачання якісних метаданих значно підвищує цитованість наукових робіт та покращує якість визначення індексів цитування. Потрібно зазначити, що аналіз сайтів наукових періодичних фахових видань НАН України показав, що відповідну платформу з необхідним комплектом метаданих мають лише два журнали. На практиці ці безкоштовні журнальні платформи виявляються лише умовно безкоштовними. Насправді їх адаптація та налаштування відповідно до вимог кожного видавництва вимагає наявності професійного веб-програміста, виділеного сервера або можливості аренди хостингу на вже налаштованому сервері, українізації англomовного інтерфейсу, розробки моделі та дизайну журналу тощо. Все це для кожного видавництва є проблемою. Як наслідок, ми маємо сьогодні безліч комерційних установ або особистих пропозицій, які, налаштувавши відповідні журнальні платформи, пропонують як корпоративні агрегатори свої послуги науковим видавництвам, і аж ніяк не безкоштовно.

Друга модель має свої переваги у тому сенсі, що основний пласт проблем, пов'язаних із представленням наукових фахових видань у світових наукових комунікаціях бере на себе адміністрація централізованого репозиторія. Сьогодні в Україні найбільш авторитетним та розвиненим є загальнонаціональний портал «Наукова періодика України». У 2014 році для забезпечення інтеграції української наукової періодики до відкритих наукових веб-ресурсів та організації повнотекстового й інформаційного пошуку за метаданими «Наукову періодика України» було переведено на нову платформу.

Упровадження бібліотечної технології опрацювання ресурсів «Наукової періодики України» забезпечило включення електронної бібліотеки фахових періодичних видань до єдиного пошукового апарата науково-

інформаційних ресурсів. Розроблені принципово нові технологічні рішення надали користувачам системи (автори, редакції, бібліотекарі, читачі) такі можливості:

- повна інформація про публікацію: журнал, випуск, рік тощо;
- можливість встановлювати термін затримки для надання відкритого доступу до повних текстів наукових публікацій;
- правильний бібліографічний опис наукових публікацій;
- пошук публікацій за «Авторським покажчиком» та «Покажчиком назв публікацій»;
- оперативний добір всіх публікацій певного автора за гіперпосиланням у форматі перегляду знайденого бібліографічного опису;
- проведення розширеного пошуку за ключовими словами;
- перехід за гіперпосиланням від бібліографічного опису наукової публікації до повного опису журналу для з'ясування його фахової спрямованості та спеціальності ДАК України;
- перехід за гіперпосиланням від бібліографічного опису наукової публікації до змісту журналу для перегляду всіх публікацій випуску;
- попередній перегляд повного тексту публікації без його завантаження на комп'ютер користувача;
- інтеграція інформації «Наукової періодики України» та реферативної бази даних «Україніка наукова»: пошук повних текстів через реферативну інформацію та наведення наявних рефератів для бібліографічних описів повних текстів наукових публікацій;
- автоматичний перехід до бібліометричного портрету публікації у системі *Google Scholar*, перегляд індексу цитування публікації та її авторів, пошук подібних публікацій;
- повнотекстовий пошук наукових публікацій з використанням лінгвістичних інструментів *Google* (синонімія, нечіткий пошук) у межах усього архіву наукової періодики України;
- запобігання плагіату (можливість відшукати будь яку фразу і порівняти тексти), а також передруку одного і того ж матеріалу в декількох виданнях.

Наявність на новій технологічній платформі повноцінних бібліографічних записів для кожної наукової публікації забезпечило можливість автоматичного формування метаданих, необхідних для коректного індексування пошуковими роботами *Google Scholar* (*Google Scholar Crawler*). Протягом 2014 року було проведено технологічне узгодження рішень. На сьогодні *Google Scholar Crawler* проіндексував бібліографічні дані приблизно 400 тис. повних текстів публікацій «Наукової періодики України» із домену Національної бібліотеки України імені В.І. Вернадського. У процесі опрацювання наукової інформації інтелектуальними засобами системи *Google Scholar* важливим елементом представлення наукових публікацій є наявність повного тексту у відкритому доступі. Це створює передумови для автоматичного аналізу роботою пристатейної бібліографії (*References*). Аналіз списку бібліографічних посилань наукової публікації надалі вико-



ристовуються системою для автоматичного обчислення індексів цитування.

Індексування пошуковими роботами *Google Scholar* величезного інтегрованого масиву наукової періодики України забезпечило постачання до глобальної наукової інформаційної системи якісних метаданих, необхідних для аналізу публікаційної активності та наукових взаємозв'язків українських учених. Національної академією наук України на офіційному рівні прийнято як одну із оцінок ефективності наукової діяльності наявність бібліометричного профіля вченого та оцінка його індексу цитування за даними системи *Google Scholar* [2]. Попередня оцінка результатів індексування дає змогу стверджувати, що персональні профілі вчених, які публікуються у журналах, що знаходяться у вільному доступі, на порталі НБУВ, значно збагатились бібліографічною інформацією (кількість бібліографічних посилань зросла у 3—4 рази). Спостерігається також зростання показників цитування публікацій та виявлення раніше недоступних семантичних зв'язків представлених результатів наукової діяльності українських учених.

Здійснена робота з усією очевидністю показала перспективність використання бібліотечних технологій для упорядкування та структурування наукових веб-ресурсів, вирішила проблеми інтеграції наукових фахових періодичних видань України до пошукового та аналітичного інтерфейсу глобальних наукових комунікацій. Інтелектуальні можливості безкоштовної наукометричної системи *Google Scholar* дають змогу українським вченим інтегрувати свою діяльність до сучасних наукових веб-технологій, покращують відкритість результатів наукової діяльності та персональну присутність у глобальному науковому співтоваристві. Ці технології забезпечують також можливість в оперативному режимі слідкувати за цитуванням наукових публікацій, виявляти тематично пов'язані напрями наукової діяльності та визначати коло колег, які активно працюють у споріднених галузях науки.

З точки зору видавництва наукових журналів інтеграція їх публікацій до єдиного науково-інформаційного ресурсу «Наукова періодика України» надає такі переваги:

1) реклама наукового фахового видання, обов'язкове посилання на сайт видавництва журналу, інформація про редколегію, тематичний профіль та контакти;

2) повноцінний бібліографічний опис наукових публікацій із необхідним комплектом метаданих, можливість переглянути цитованість окремої публікації та цитованість авторів публікації;

3) інтегрований ресурс працює на розвиток науки та користувачів, що опосередковано сприяє використанню наукової продукції видавництв, її цитованості та затребуваності, що, як наслідок, підвищує рейтинг видань та їх видавництв;

4) наукові видавництва мають серед потенційних споживачів їх інформації щодобово понад 5 тисяч сеансів відвідування — це щоденне читання наукових фахових видань та цитування їх публікацій, яке не може бути забезпечено самостійно окремими сайтами журналів;

5) у рейтингах пошуку *Google* запити із порталу НБУВ надаються першими і містять необхідні метадані про публікацію (автор, назва, журнал). Посилання веде із пошукової системи *Google* до персональної сторінки публікації на порталі НБУВ, де на неї можна коректно послатися (у тому числі вказати видання, в якому опублікована стаття);

6) завдяки проіндексованим *Google Scholar* метаданим публікацій (які обов'язково містять назву журналу) з'явилась можливість створити наукометричний профіль журналу та слідкувати за цитованістю опублікованих у ньому статей;

7) сервіси для авторів та видавництв надаються в межах інформаційної платформи НБУВ безкоштовно на умовах вільного доступу до повних текстів публікацій. Ці функції, які наукова бібліотека традиційно надає безкоштовно, охоче переберуть на себе комерційні установи (бо ця ніша обов'язково буде заповнена), які будуть продавати ці послуги за гроші, що аж ніяк не покращить становище та статус наукових фахових видань України та наукових видавництв.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Васильєв О., Чьочь В. Системи пошуку наукової інформації (Огляд) // Бібліотечна планета. — 2009. — № 2. — С. 11–16.
2. Костенко Л., Жабін О., Кузнецов О., Кухарчук Є., Симоненко Т. Бібліометрика української науки: інформаційно-аналітична система // Бібліотечний вісник. — 2014. — № 4. — С. 8–12.
3. Лобузінa К.В. Бібліотека 3.0: знання, сховища даних, експерти / К.В. Лобузінa // Бібліотекознавство. Документознавство. Інформологія. — 2012. — № 1. — С. 26–35.
4. Мриглюд І., Мриглюд О. Наука України в світовому інформаційному просторі // Вісник НАН України. — 2007. — № 7. — С. 3–18.
5. Одоцюк І. Оцінка результатів наукової діяльності в Україні: нормативно-правовий аспект // Проблеми інноваційно-інвестиційного розвитку. — 2012. — № 3. — С. 38–42.
6. Радченко А.І., Яцків Т.М. Упровадження системи цифрових ідентифікаторів DOI: поступ і досвід // Наука України у світовому інформаційному просторі / НАН України. — К. : Академперіодика, 2014. — Вип. 10. — С. 41–46.
7. Яцків Я.С. Сучасний стан та проблеми української наукової періодики // Наука України у світовому інформаційному просторі / НАН України. — К. : Академперіодика, 2013. — Вип. 8. — С. 5–16.

Lobuzina K.V.

LIBRARY AND SEMANTIC WEB-TECHNOLOGIES IN ORGANIZATION OF ACCESS TO SCIENTIFIC PERIODICALS

The article is devoted to the peculiarities of the organization and prospects of integrated information resource of scientific periodicals The Vernadsky National Library of Ukraine. Benefits to publishers and authors placement of full-text scientific publications in a centralized repository of open access are analyzed and defined. The modern capabilities of semantic library and web-technology to determine scientometric indicators of scientists and scientific journals are reviewed.

Key words: scientific periodicals, open access, centralized repository, scientific information, semantic library and web-technologies, scientometric indicators.