

Юлія ЯКОВЛЄВА,
молодший науковий співробітник НБУВ

Оцінка інформативності документів як напрям інтелектуалізації пошукових систем наукових бібліотек

У статті запропоновано впровадження нової парадигми інформаційного пошуку в інформаційно-пошукових системах наукових бібліотек, що передбачає, крім розвитку існуючих на сьогодні напрямів підвищення релевантності пошуку, також реалізацію оцінки інформативності знайдених документів.

Ключові слова: інформаційно-пошукові системи, інформатизація бібліотечних процесів, лінгвістичне забезпечення, штучний інтелект.

Максимальне задоволення інформаційних запитів читачів – основна мета діяльності бібліотек. Це положення лежить в основі роботи як бібліотеки в цілому, так і її окремих ділянок, у тому числі – в основі оцінки ефективності довідково-бібліографічного апарату як одного з видів інформаційно-пошукових систем.

Процес пошуку бібліографічної інформації досліджувався фахівцями в нашій країні протягом багатьох років¹. Пошук інформації вивчався, в основному, у сфері традиційних карткових каталогів. Однак вже тоді фахівцями були визначені перспективи пошуку інформації в автоматизованих системах відповідно до елементів класифікаційних індексів. Цей напрям досліджень був пізніше продовжений у роботах Державної бібліотеки СРСР ім. В. І. Леніна, пов'язаних зі створенням електронного каталогу. У сучасному бібліотекознавстві вивченням інформаційного пошуку займаються К. М. Зайцева², Е. Р. Сукиасян³, Н. І. Гендіна⁴,

О. О. Лаврьонова⁵, Г. А. Скарук⁶, І. П. Антоненко⁷, Н. Й. Дзюба⁸, К. В. Бардієр⁹.

Головна мета удосконалення інформаційно-пошукових систем (ІПС) наукових бібліотек – це інтелектуалізація інформаційного пошуку, тобто можливість вироблення рішення і рекомендації на рівні експерта. Іншими словами, такі ІПС повинні реконструювати методику вирішення експертом завдання інформаційного пошуку у відповідній галузі; формувати висновки, ґрунтуючись на знаннях, які зберігаються окремо від програмного коду, що керує процесом виводу; використовувати при вирішенні завдань евристичні методи; мати можливість

<http://www.nbu.gov.ua/articles/crimea/1997/DOC/sem11/sl1doc7.html>. – Загл. с екрана.

⁵ Лавренова О. А. Лингвистика информационных систем и межбиблиотечное сотрудничество [Электрон. ресурс] // Библиотеки и ассоциации в меняющемся мире: новые технологии и новые формы сотрудничества: 4-я междунар. конф. «Крым-97»: Материалы конф., Судак, 4–15 июня 1997 г. – Способ доступа: URL: <http://www.nbu.gov.ua/articles/crimea/1997/DOC/sem11/sl1doc10.html>. – Загл. с экрана.

⁶ Скарук Г. А. Тематический поиск в электронном каталоге: проблемы лингвистического обеспечения // Библиотековедение. – 2001. – № 3. – С. 53.

⁷ Антоненко І. П. Сучасна каталогізація: постановка проблем і шляхи їх розв'язання // Проблеми вдосконалення каталогів наукових бібліотек. – К., 1997. – С. 6–8.

⁸ Дзюба Н. И. Система лингвистического обеспечения АИБС детской библиотеки // Библиотеки и ассоциации в меняющемся мире: новые технологии и новые формы сотрудничества: 10-я юбил. междунар. конф. «Крым-2003»: Тр. конф. – М., 2003. – Т. 2. – С. 856–858.

⁹ Бардієр К. Структура лінгвістичного забезпечення сучасних інформаційних сервісів // Бібліотечний вісник. – 2002. – № 1. – С. 55–61.

¹ Сукиасян Э. Р. Homo Quaerens (Человек ищущий). К проблеме развития познавательных способностей читателя в процессе информационного поиска // Науч. и техн. б-ки. – 2002. – № 4. – С. 32–33.

² Зайцева Е. М. Лингвистическое обеспечение автоматизированных библиотечно-информационных систем: современные требования и направления развития // Науч. и техн. б-ки. – 2000. – № 3. – С. 54.

³ Сукиасян Э. Р. Электронные каталоги // Библиотека. – 2003. – № 3. – С. 31–33.

⁴ Гендина Н. И. Лингвистическое обеспечение информационно-поисковых систем: теория и технология [Электрон. ресурс] // Библиотеки и ассоциации в меняющемся мире: новые технологии и новые формы сотрудничества: 4-я междунар. конф. «Крым-97»: Материалы конф., Судак, 4–15 июня 1997 г. – Способ доступа: URL:

накопичувати знання в процесі функціонування; мати здатність пояснювати пропонувані рішення; забезпечувати «дружній» інтерфейс із користувачем бажано природною мовою.

Базуючись на позитивних рисах пошукового апарату бібліотек і тенденціях розвитку інформаційних технологій, можна сформулювати напрями інтелектуалізації інформаційно-пошукових систем:

- розвиток традиційних засобів лінгвістичного забезпечення;
- організація ітеративного природномовного користувацького інтерфейсу, який дозволить використати всі можливості ІПС і тим самим забезпечити якісне задоволення інформаційної потреби користувача ІПС;
- організація зворотного зв'язку в ІПС на основі використання різнопланових статистичних даних для оцінки релевантності результатів пошуку.

Необхідними умовами реалізації оцінки релевантності результатів пошуку є, по-перше, забезпечення механізму відстеження і аналізу результатів пошуку, і, по-друге, використання технології ранжування результатів пошуку за релевантністю знайдених документів.

Перший з наведених аспектів – відстеження і аналіз «успішних» пошуків – дозволить покращити ефективність пошуку для подібних запитів. Найефективніше такий алгоритм спрацюватиме, якщо є можливість аналізувати використання результатів пошуку, тобто відстежити, які саме з запропонованих пошуковою системою документів зацікавили користувача. Такий механізм активно застосовується в онлайн-магазинах, наприклад, у книжковому онлайн-магазині поруч з описом потрібної книги користувач знайде посилання на книги, котрі замовив або купив інший покупець, який шукав цю книгу. В АІБС такий механізм можливий у разі функціонування автоматизованого замовлення. Навіть при відсутності такої функції відстеження і аналіз результатів пошуку буде можливим, якщо в ІПС реалізовано механізм відбору користувачем множини потрібних документів з усього масиву запропонованих системою документів. Такий механізм зробить зручнішим користування системою, і, крім того, дані про видання, відібрані в підмножину, певним чином відобразатимуть попит на літературу і можуть використовуватись у системі моніторингу використання бібліотечних фондів.

Другий аспект, а саме реалізація ранжування результатів пошуку за релевантністю потребує збору і оцінки статистичних даних про використання

знайдених документів. Варто наголосити на важливості аналітичного етапу. Збір статистики як такий не є виміром, лише оцінка важливості і релевантності знайденого документа дозволить реалізувати алгоритм ранжування і тим самим підвищити якість функціонування ІПС.

Нами запропоновано методику ранжування результатів пошуку в інформаційно-пошукових системах науково-технічних бібліотек, що базується на багатокритеріальній оцінці видань за їх інформативністю¹⁰. Ранжування результатів інформаційного пошуку пропонується здійснювати на основі багатокритеріальної оцінки релевантності знайдених документів, враховуючи:

- фактори старіння науково-технічної літератури (вік знайденого видання);
- статистичні дані про обіг літератури в бібліотеці;
- значущість автора (індивідуального або колективного) на основі даних файлів авторитетних записів.

На нашу думку, наведені критерії якнайкраще відображають інформативність документів. Перший критерій – старіння літератури – як закономірний постійний процес зменшення з часом необхідності її використання для отримання вміщеної в ній інформації, однозначно пов'язаний з інтенсивністю використання документів. Значущість використання даного критерію для ранжування можна підвищити, якщо для нормування його числових значень використовувати одну з функцій, які апроксимують закономірності старіння літератури. Аналіз публікацій, присвячених теорії старіння інформації, дозволяє зробити висновок, що для апроксимації закономірностей старіння літератури найчастіше використовуються від'ємні експоненціальні та від'ємні показникові функції¹¹.

Так само не викликає сумнівів цінність другого критерію, а саме – даних системи моніторингу використання бібліотечних фондів у науковій бібліотеці. Нами запропоновано методику селективного моніторингу для виявлення видань підвищеного попиту шляхом вибіркового аналізу відповідей

¹⁰ Яковлева Ю. В. Ранжирування результатів пошуку в інформаційно-пошукових системах бібліотек // Документознавство. Бібліотекознавство. Інформаційна діяльність: Проблеми науки, освіти, практики: Зб. матеріалів міжнар. наук.-практ. конф., Київ, 25–26 трав. 2004 р. – К., 2004. – С. 115–117.

¹¹ Горькова В. И. Информетрия (Количественные методы в научно-технической информации) // Итоги науки и техники. Сер. Информатика. Т. 10. – М.: ВИНТИ, 1988. – 328 с.

книгосховища про незадоволені запити¹². Числовими значеннями даного критерію є кількість вимог (відповідей книгосховища) на видання за визначений період аналізування даних, наприклад за останній рік. Для нормалізації значень з метою використання критерію в завданні багатокритеріальної оцінки числові значення діляться на деякі нормуючі коефіцієнти, за які беруться максимальні (мінімальні) значення критеріїв з області припустимих рішень.

Використання третього критерію – даних файлів авторитетних записів – можливе за умов розробки методів оцінки значущості авторів наукових публікацій. Однією з перспектив розвитку цього напрямку може бути автоматичний семантичний аналіз авторитетних записів, результатом якого є ранжування (бальна оцінка) індивідуального або колективного автора. Так, академічне видання буде мати найвищу оцінку, вишівське видання – дещо нижчу і так далі. Питання введення такої оцінки потребує окремого дослідження за участі спеціалістів з інформаційних технологій та бібліотекознавців-практиків.

Оцінка інформативності документів на основі запропонованої методики ранжування результатів пошуку є важливим напрямом розвитку інформаційно-пошукових систем наукових бібліотек, але, на нашу думку, сьогодні вже можна дослідити й інші перспективні напрями, пов'язані зі збільшенням питомої ваги електронних наукових публікацій. Зокрема, стає можливим упровадження автоматичної обробки текстів на базі методу індексування цитування, який добре зарекомендував себе у бібліотечній практиці.

Як відомо, індекс цитування – визнана у науковому світі міра «значущості» праць певного вченого¹³. Величина індексу визначається кількістю посилань на цю працю (чи прізвище) в інших джерелах. Показники наукового цитування Science Citation Index (SCI), Social Science Citation Index (SSCI) і Arts & Humanities Citation Index, а також Інтернет-SCI (Web of Science), що готуються і публікуються Інститутом наукової інформації (Institute of Scientific Information (ISI), Філадельфія, США), широко використовуються ученими різних країн

світу, у тому числі науковцями України та Росії. Інформаційні продукти ISI і, насамперед, SCI – унікальний інструмент для вирішення цілої низки завдань.

Використання даних цитування для ретроспективних і сучасних літературознавчих досліджень, виявлення видатних учених у будь-якій галузі, аналіз інтелектуальної структури тієї чи іншої дисципліни є важливою науково-дослідною та комунікативною методикою. На її базі можлива реалізація технології автоматичної обробки текстів електронних наукових публікацій.

Вичленовування з тексту наукових публікацій переліку використаної літератури і посилань для накопичення та подальшого аналізу дозволить проводити якісну оцінку наукової значущості документів. Це завдання досить неординарне, але його вирішення полегшиться за умови зберігання текстів електронних документів у структурованому вигляді. У такому випадку бібліографія до публікації є окремим об'єктом, придатним до автоматичного аналізу. Засобом реалізації такого завдання може бути мова XML (eXtensible Markup Language)¹⁴, затверджена у 1998 р. міжнародною організацією World Wide Web Consortium (W3C). XML – це розширювана мова розмітки, призначена для опису в текстовій формі структурованих даних.

Використання технологій XML в електронних бібліотеках досить перспективне. На сьогодні важливим питанням є розробка єдиних стандартів для мов розмітки, створюваних на основі XML, що дозволить зменшити витрати на створення конверторів XML–XML. Переваги XML роблять його зручним інструментом для розробки гнучких засобів пошуку й аналізу інформації, зокрема засобів аналітичної обробки списків літератури, наведених у наукових електронних публікаціях з метою визначення інформативності документа.

Аналіз структурованих текстів за допомогою технологій XML та методики індексування цитування в сукупності з іншими методами інтелектуалізації пошукових систем забезпечить якісніше задоволення інформаційних потреб користувачів інформаційно-пошукових систем наукових бібліотек.

¹² Яковлева Ю. В. Селективний моніторинг використання бібліотечних ресурсів // Ресстрація, зберігання і обробка даних. – 2002. – № 1. – С. 89–96.

¹³ Маршакова И. В. Система цитирования научной литературы как средство слежения за развитием науки / ВИНТИ АН СССР; Отв. ред. А. И. Михайлов. – М.: Наука, 1988. – 287 с.

¹⁴ The World Wide Web Consortium [Electronic Resource]. – Way of access: URL: <http://www.w3.org/>. – Title from the screen.

[Електрон. ресурс] // Электронные библиотеки: науч. электрон. журнал. – 1999. – Т. 2. – Вып. 3. – Способ доступа: URL: <http://www.iis.ru/el-bib/1999/199902/elib/elib.en.html>. – Загл. с экрана.

Отже, для досягнення якісно нового результату функціонування інформаційно-пошукових систем наукових бібліотек ми вважаємо за необхідне розвиток та впровадження нової парадигми інформаційного пошуку, що передбачає, крім розвитку існуючих на сьогодні напрямів підвищення релевант-

ності пошуку, реалізацію оцінки значущості документів. Пошуковий апарат наукових бібліотек має базуватися на розвиненому аналітичному апараті, який забезпечує проведення в інформаційних масивах якісних бібліометричних, наукометричних та інформетричних досліджень.

УДК 002.6 + 002.5

Надія ЗАЙЧЕНКО,
науковий співробітник НБУВ

Реферативні ресурси - інформаційний продукт сучасної наукової бібліотеки

У статті подано стислий огляд світових реферативних ресурсів (баз даних, реферативних журналів), а також реферативних видань в Україні за період 1999–2004 рр. Наведено статистичні дані щодо загальнодержавної реферативної бази даних «Україніка наукова», висвітлено її сучасний стан, технологію формування та використання. Розглянуто форми та види кооперації, на основі яких здійснюється реалізація проекту про національну систему реферування наукової літератури.

Ключові слова: інформаційний продукт, інформаційні ресурси, наукова література, реферативна база даних, реферативний журнал, реферативні ресурси, форми кооперації.

Діяльність наукових бібліотек та інформаційних установ завжди була спрямована на забезпечення наукових досліджень світовими документними ресурсами, сприяння пропаганді та розповсюдження наукових знань. Але постійна потреба вчених в оперативній, повній та достовірній інформації про наукові досягнення на сьогодні не може бути забезпечена тільки традиційними методами, зважаючи на економічні причини та технологічні умови надання інформації, які постійно змінюються. Необхідно зазначити, що кінець ХХ ст. ознаменувався новим технологічним стрибком, який призвів до стрімкого росту машиночитаних (первинних і вторинних) джерел, зумовив появу бібліографічних, реферативних, повнотекстових баз даних (БД), які функціонують у режимах офлайн та онлайн на веб-сайтах бібліотек, інформаційних установ, навчальних закладів, а також БД на компакт-дисках, котрі, у свою чергу, почали активно використовуватися для задоволення інформаційно-бібліографічних запитів читачів. Зріс рівень підготовки користувачів бібліотек, змінилися їхні потреби, підвищилися вимоги до якості наданих послуг. У даній ситуації основним завданням

бібліотек є реалізація права користувача на одержання оперативної та повної інформації, доступ до ресурсів не тільки в бібліотеці, а й за її межами. Окрім традиційного МБА значна кількість бібліотек сьогодні активно використовує у своїй діяльності необмежені ресурси інтернету, значно підвищивши таким чином інформаційну повноту обслуговування і надаючи читачам інформацію у зручнішій формі порівняно з традиційними.

Завдяки впровадженню передових бібліотечно-інформаційних технологій, комп'ютеризації та автоматизації бібліотечних процесів провідні наукові бібліотеки поступово перетворюються з книгозбірень на інформаційні центри. Окрім того, що вони використовують у своїй діяльності *інформаційні продукти*, створені органами науково-технічної інформації (НТІ), науково-дослідними установами, навчальними закладами та ін., найбільші з них беруть безпосередню участь і у процесі їх створення.

Інформаційні продукти – це інформаційні ресурси всіх видів, програмні продукти, бази та банки даних та інша інформація, представлена у формі товару.