

УДК 001.811:004.7:021(477)

**Олександр Жабін,**  
наук. співроб. НБУВ

### **ВЕБОМЕТРИЧНІ ДОСЛІДЖЕННЯ БІБЛІОТЕЧНИХ САЙТІВ УКРАЇНИ**

У статті описано технології вимірювання якості (важливості та авторитетності) інформаційного наповнення веб-сайтів бібліотечних установ за показниками Alexa Rating, Google PageRank, Яндекс ТІЦ, що використовують статистичні вебометричні методики досліджень. Наведено впорядковані рейтингові таблиці сайтів низки книгозбірень за результатами цих показників. Надано рекомендації щодо шляхів підвищення рейтингу інформаційних ресурсів бібліотечним установам.

*Ключові слова:* вебометричний рейтинг, індекс цитувань, інформаційний ресурс.

Глобальна інфраструктура мережі Інтернет – неймовірно-величезне інформаційне джерело, а більшість населення розвинених держав стала його постійним користувачем.

Перед такою численною аудиторією потенційних споживачів не встояли й видавництва, які швидко стали переорієнтовувати видання книг та журналів з друкованої форми до цифрової.

Ці тенденції відбилися й на бібліотечному осередку, де інформаційний потік наповнення та зберігання знань людства засяяв новим цифровим забарвленням. Крім того, книгозбірні прискорили процеси оцифрування раніше накопичених документальних ресурсів.

Мережа Інтернет – ідеальне інтеграційно-інформаційне середовище спільного використання та обміну результатами інноваційної і наукової діяльності, особливо для молодих дослідників. Для початківця іноді досить проблематично оперативно виставити для широкого загалу у традиційній, друкованій формі свої наукові публікації, особливо у високорейтингових журналах. На перешкоді стають такі фактори, як черги на друк у періодичних наукових виданнях, фінансові проблеми або недостатня, на чийсь думку, «наукова вага» молодого вченого. Тому наукова спільнота

досить швидко зорієнтувалась у нових реаліях і оперативно стала виставляти на широкий загал результати своїх досліджень у електронній формі використовуючи широкі можливості світових інформаційних мереж. Розпочався неконтрольований процес заповнення глобального інформаційного простору численними інформаційними колекціями з різною науковою цінністю.

Постала проблема пошуку та вибору першочерговості інформаційного огляду сайтів з найбільш цитованими документами у процесі наукових досліджень. Тому серед науковців попит на використання послуг сайтів з пошуковим програмним забезпеченням, що використовує вебметричні інструменти аналізу цінності інформації став домінуючим, а на теренах глобального інформаційного простору з'явилися низка зведених топ-рейтингів онлайн-ресурсів, упорядкованих за різними статистичними методиками, що враховували численні кількісні показники та якісні властивості електронних документів. Поступово стали підвищуватися й рейтинги бібліотечних установ, де темпи накопичення цифрової інформація стали доминувати над темпами накопичення на паперових носіях.

Метою цієї статті є аналіз технології ранжування та формування топ-рейтингів сайтів за показниками Alexa Rating, Google PageRank, Яндекс ТІЦ, що використовують статистичні вебметричні методики досліджень якості (важливості та авторитетності) інформаційного наповнення. Головний критерій вибору показників обумовлений вільним доступом до результатів їхньої роботи та можливістю безкоштовного оперативного використання й будь-яким бібліотечним установам для оцінки рівня затребуваності своїх інформаційних ресурсів по відношенню до подібних у інших.

Серед компаній, що вимірюють дані якості ресурсу, надають послуги інформаційно-аналітичного огляду та пошуку з використанням елементів вебметричних технологій можна виділити таких представників, як Alexa Rating, Google PageRank, «Яндекс ТІЦ».

Alexa Rating – показник авторитетності ресурсу, надає комбінований рейтинг, що враховує, як число відвідувачів (у яких встановлено Alexa Toolbar), так і кількість переглядів сторінок за деякий проміжок часу (3 місяці, місяць, неділя, доба). Чим менше число Alexa Rank – тим краще.

Це система рейтингу певного інтернет-ресурсу по відношенню до всієї глобальної мережі, тобто Alexa Rank показує наскільки популярний сайт

по відношенню до всієї аудиторії інтернету. Тобто чим менше число Alexa Rank у вас, тим вищий ваш сайт в рейтингу (краще той, який ближче до одиниці). Наприклад сайт з рейтингом 5,550 кращий, ніж з рейтингом 6,000 [1]. Перший спосіб (загальний) дізнатися Alexa Rank, це вводити кожне доменне ім'я на сайті – <http://www.alexa.com/> в полі Search Alexa.

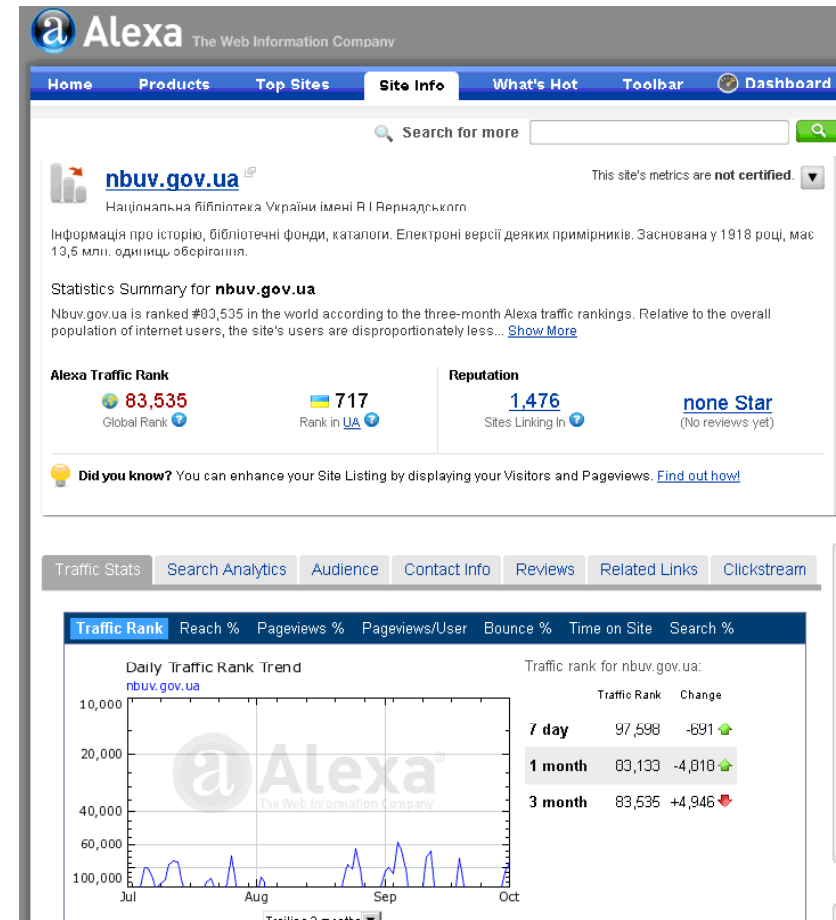


Рис.1. Показники НБУВ від Alexa Rank

На рис.1 відображено результат розрахунків показника Alexa Rank після вводу доменного імені сайту Національної бібліотеки України імені В. І. Вернадського (НБУВ).

Інший спосіб дізнатися свій рейтинг – скачати й встановити панель Alexa Toolbar, відображеного на рис. 2.

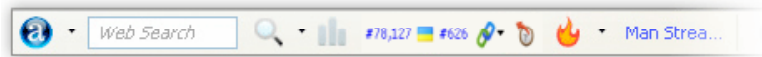


Рис. 2. Панель Alexa Toolbar

Як збільшити значимість показника Alexa Rank? Це питання пов'язане з відвідуваністю сайту. Чим більше відвідуваність вашого сайту, тим нижче буде значення Alexa Rank. Значення нижче, а значимість більше [2].

Компанією Google для визначення авторитетності сайту використовується технологія визначення показника PageRank, розроблена у Стенфордському університеті.

PageRank (далі просто PR) це числова величина – міра «важливості» сторінки в пошуковій системі Google. Залежить від числа зовнішніх посилань на дану сторінку та від їх ваги (важливості). Іншими словами – від кількості і якості посилань. А якщо говорити математичною мовою, то PR це алгоритм розрахунку авторитетності сторінки, що використовується пошуковою системою Google. PR не є основним, але є одним з допоміжних чинників при ранжируванні сайтів у результатах пошуку [3].

Спростивши викладене, можна сказати, що PR інтерпретує зв'язки між сторінками сайтів, тобто на посиланнях зі сторінки «А» на сторінку «В» як голос «А» на користь «В». Враховується не тільки кількість голосів, але і якість голосуючих сторінок. Сторінкам привласнюється певний рейтинг – від 0 до 10.

Поділ на одиниці за шкалою від 0 до 10 базується на логарифмічній залежності або на чомусь вельми схожому (спеціальна формула розрахунку), але тільки не на рівномірному розділенні. Перерахунок значущості сторінок відбувається під час так званого «Google Dance». Можна дотримуватися приблизно такої градації: PageRank від 4 до 5 – найбільш типовий для більшості сайтів середньої «розкрученості», 6 – дуже добре «розкручений» сайт, 7 – величина, практично недосяжна для сайтів,

але іноді зустрічається. Значення 8, 9, 10 мають виключно популярні і значущі проекти. Наприклад, зараз у каталогі dmoz.org Pagerank = 8, у англійської Вікіпедії Pagerank = 9. Значення 10 мають всього декілька десятків сайтів, наприклад: <http://www.whitehouse.gov>, <http://www.adobe.com> [4].

Отже, кожна унікальна сторінка сайту, проіндексована Google, має вагу PageRank. Для вимірювання PR існує величезна кількість онлайн-сервісів, що дозволяють визначити даний в автоматичному режимі. Крім того можна встановити Google Toolbar у браузері користувачів.

ТІЦ (Тематичний індекс цитування) – алгоритм пошукової системи Яндекс, призначений для визначення авторитетності сайтів шляхом підрахунку кількості посилань на нього ресурсів схожої тематики.

Індекс цитування (або ІЦ) – прийнята в науковому світі міра «значущості» праць якого-небудь вченого. Величина індексу визначається кількістю посилань на цю працю (або прізвище) в інших джерелах. Проте для дійсно точного визначення значущості наукових праць важлива не тільки кількість посилань на них, але і якість цих посилань. Так, на роботу може посилатися авторитетне академічне видання, популярна брошура або розважальний журнал. Значимість у таких посилань різна. Яндекс (ТІЦ) визначається з урахуванням якісної характеристики посилань – «вагою» посилання, що розраховується за спеціальним розробленим алгоритмом, тобто показник визначається не загальною кількістю посилань, а сумою їх впливів. Велику роль відіграє тематична близькість ресурсів, які посилаються на нього [5].

В таблиці наведено кількісні значення вебметричних рейтингів 14 сайтів бібліотек загальнодержавного значення України та 10 національних бібліотек країн світу на основі даних від «Alexa Rating», «Google PageRank», «Яндекс ТІЦ» (станом на 15 травня 2012 року) [6]. Проаналізувавши табличні дані зведеного рейтингу книгозбірень, що пріоритетно впорядковані за округленим значенням показника Alexa Rating, зробимо декілька висновків:

1. Немає одностайності в оцінках рейтингу деяких бібліотек від «Alexa Rating», «Google PageRank», «Яндекс ТІЦ». Так Львівську національну наукову бібліотеку імені Василя Стефаника Google PageRank може поставити на перше місце, «Яндекс ТІЦ» – на третє, а Alexa Rating поставив аж на тринадцяте.

Таблиця

## Вебометричні рейтинги сайтів бібліотек

N п/п	Бібліотека	Alexa Rating (тис.)	Google PageRank	Яндекс ТІЦ
Бібліотеки загальнодержавного значення				
1.	Національна бібліотека України імені В. І. Вернадського	68	8	2900
2.	Харківська державна наукова бібліотека імені В. Г. Короленка	242	6	750
3.	Національна парламентська бібліотека України	1.125	6	450
4.	Державна бібліотека України для юнацтва	1.273	5	475
5.	Національна бібліотека України для дітей	2.245	6	1100
6.	Державна науково-педагогічна бібліотека України імені В. О. Сухомлинського	2.182	6	500
7.	Державна наукова архітектурно-будівельна бібліотека імені В. Г. Заболотного	2.900	4	10
8.	Книжкова палата України імені Івана Федорова	3.064	6	130
9.	Центральна державна науково-технічна бібліотека гірничо-металургійного комплексу України	3.587	4	20
10.	Державна науково-технічна бібліотека України	3.872	5	475
11.	Державна наукова сільськогосподарська бібліотека	7.300	5	110
12.	Львівська національна наукова бібліотека імені Василя Стефаника	10.285	8	550
13.	Національна історична бібліотека України	12.966	5	160

Продовження таблиці

N п/п	Бібліотека	Alexa Rating (тис.)	Google PageRank	Яндекс т. щ
14.	Національна наукова медична бібліотека України	18.102	5	140
Національні бібліотеки країн світу				
1.	Бібліотека Конгресу США	3	9	2600
2.	Національна бібліотека Франції	17	9	700
3.	Національна бібліотека Японії	20	8	110
4.	Національна бібліотека Великої Британії	27	9	900
5.	Національна бібліотека Італії	31	8	50
6.	Національна бібліотека Китаю	45	8	80
7.	Російська державна бібліотека (Москва)	73	8	8300
8.	Національна бібліотека Німеччини	80	8	140
9.	Національна бібліотека Росії (Санкт-Петербург)	126	9	4700
10.	Національна бібліотека Білорусі	488	8	1700

2. Вебометричне оцінювання інформаційного ресурсу веб-сайтів книгозбірень автоматизованими системами не підвладне людському фактору, вельми точно та показує істинний стан популярності по відношенню до інших бібліотек. Так, наприклад, Харківська державна наукова бібліотека імені В. Г. Короленка за рейтингом випередила таку потужну книгозбірню, як Національна парламентська бібліотека України.

3. Провідні книгозбірні України показали досить хороший рівень рейтингів на тлі показників національних бібліотек світу.

Окремо треба відзначити провідне місце у зведеному рейтингу Національної бібліотеки України імені В. І. Вернадського. Це – результат

багатолітньої кропіткої роботи колективу книгозбірні протягом останнього десятиліття по наповненню веб-ресурсу повнотекстовим інформаційним науковим контентом, основу якого складають загальнодоступні ресурси, а саме:

- репозитарій «Наукова періодика України» (0,6 млн статей з 1,7 тис. журналів);
- електронні тексти (80 тис. документів, з них 60 тис. авторефератів дисертацій);
- аналітичні матеріали (4 тис. бюлетенів оперативної інформації та інформаційно-аналітичних оглядів) [7].

Цей рейтинг підтверджують і понад 65 тисяч користувачів, які щодоби відвідують онлайнві ресурси книгозбірні.

Таким чином, для досягнення чинних місць у топ рейтингах визначення цінності ресурсу бібліотекам необхідно:

- нарощувати наповнення затребуваними відкритими повнотекстовими ресурсами веб-сайтів;
- проводити моніторинг ефективності використання своїх інформаційних ресурсів;
- використовувати елементи вебметричних технологій для усвідомлення реального стану цінності накопиченого контенту у співвідношенні з подібним у інших;
- наводити детальний опис для кожного виставленого документу або колекції;
- прагнути до інтуїтивноосмисленого та прозороструктурованого подання веб-інформації на постійній (мінімальні зміни у структурі контенту даних) основі.

Підсумовуючи вищевикладене, можна зробити такі висновки:

1. Слід відзначити відсутність єдиних уніфікованих критеріїв визначення інформаційної цінності ресурсу сайту.
2. Діє низка веб-технологій з відповідним програмним інструментарієм для визначення авторитетності та веб-активності інформаційних ресурсів з вільним доступом до своїх сервісів.
3. Доступність для індивідуального використання на основі вільного поширення, ряду програмних продуктів для визначення цінності інформаційного ресурсу.
4. Бібліотеки, що активно накопичують інформаційний ресурс, надають на своїх веб-сайтах відкритий доступ до них та аналізують ефективність їх

використання, стали посідати більш високі щаблі у зведених рейтингах затребуваності у користувачів.

### Список використаних джерел

1. Все об Alexa Rank [Електронний ресурс]. – Режим доступу : URL : <http://bloginat.ru/alexa-rank>. – Заголовок з екрану.
2. Alexa Rank – Что это такое? [Електронний ресурс]. – Режим доступу : URL: <http://irinanovikova.com/catalog/alexa-rank-cto-eto-takoe.html>. – Заголовок з екрану.
3. Что такое PageRank. [Електронний ресурс]. – Режим доступу : URL : <http://optsite.narod.ru/> – Заголовок з екрану.
4. Все о Google PageRank. [Електронний ресурс]. – Режим доступу : URL : <http://www.pr0.ru/> – Заголовок з екрану.
5. Что такое тИЦ. [Електронний ресурс]. – Режим доступу : URL: <http://help.yandex.ru/catalogue/?id=873431>. – Заголовок з екрану.
6. Вебметричний рейтинг. [Електронний ресурс]. – Режим доступу : URL : <http://www.nbuv.gov.ua/portal/librating.html>. – Заголовок з екрану.
7. Статистика порталу Бібліотеки. Інформаційні ресурси порталу. [Електронний ресурс]. – Режим доступу : URL : <http://www.nbuv.gov.ua/library/webstat.html> – Заголовок з екрану.