

## **ВИБІР ПОШУКОВОЇ СИСТЕМИ ВИЗНАЧЕННЯ НАУКОВОГО ІНДЕКСУ ЦИТОВАНOSTI**

**О.Р. ГАРАСИМ**

Піднято проблему вибору пошукової системи наукового індексу цитування для відкритих українських наукових архівів. Індекс цитованості є досить важливим, його можна отримати використовуючи сучасні пошукові системи, які спеціалізуються на індексації наукової літератури. Величина індексу цитування визначається кількістю посилань на публікацію або прізвище автора в інших джерелах. Для того, щоб здійснити пошук цитованості наукових документів електронного архіву розглянуто характеристики систем наукової індексації, основними з яких є Scopus, Російський індекс наукового цитування, Web of Science та Google Scholar. Для цих п'яти пошукових систем наукового індексу цитування побудовано дерево цілей яке складається з рівнів альтернатив, критеріїв та цілей. Методом аналізу ієрархії проведено попарне порівняння альтернатив за визначеними критеріями, з урахуванням найбільшої ваги критерію «охоплення українських видань».

### **ВСТУП**

Щороку в Україні збільшується кількість наукових публікацій. Наукові роботи дозволяють апробувати дослідження, а також знайти інших дослідників, які займаються вибраною тематикою. Питання якості та прикладне значення роботи дослідників стають одними із головних, оскільки кількість праць не може адекватно врахувати науковий внесок. Сьогодні вже існують системи, які спрямовано на визначенні числа цитованості наукового документу, як показника, який характеризує значущість роботи для подальшого розвитку науки.

Величина індексу визначається кількістю посилань на документ чи прізвище науковця. Для українських відкритих наукових архівів постає проблема отримання індексу цитованості для документів, які в ньому зберігаються.

### **ЗВ'ЯЗОК ВИСВІТЛЕНОЇ ПРОБЛЕМИ ІЗ ВАЖЛИВИМИ НАУКОВИМИ ТА ПРАКТИЧНИМИ ЗАВДАННЯМИ**

Індекс цитування наукових статей (ІЦНС) — реферативна база даних наукових публікацій, яка індексує посилання, що вказані у джерелах використаної літератури наукових робіт і представляє кількісні показники цих посилань [1].

Кількість публікацій та індекс цитування є одними із основних показників, що використовуються під час оцінки результативності наукової діяльності окремих учених, наукових підрозділів та організацій, а також для оцінки рівня наукових журналів. Основні параметри, які можуть бути використані для оцінки рівня наукових журналів:

- періодичність — частота виходу журналу;
- число статей за останній рік — кількість статей, які опубліковано в цьому виданні;
- наявність ISSN, EISSN — наявність ідентифікаторів;
- імпакт-фактор — визначає рейтинг цитованості видання;
- рік випуску — можливість вибору року за яким проводиться оцінення;
- тематична спрямованість;
- індексування в зарубіжних і російських базах даних [2].

Вихідною основою для аналізу структури цитувань та визначення бібліометричних характеристик є бази даних цитованості за періодикою, в яких збираються не тільки бібліографічні дані про журнальні публікації (автор, назва публікації, найменування журналу, рік, том, випуск, сторінки), але й списки цитованої в статтях літератури. Це дозволяє знаходити як публікації, цитування в конкретній статті, так і публікації, що цитують цю статтю. Таким чином користувач може проводити надзвичайно ефективний пошук всієї бібліографії за його запитами.

Як правило, поряд із бібліографічною та цитатною інформацією в індекс цитування включають відомості про авторів та організацію, в якій вони працюють, тому за допомогою такої бази можна інтегрувати показники на будь-якому рівні: для дослідника-автора, структурного підрозділу та інституту, міністерств і відомств, цілих адміністративно-географічних регіонів [3]. Система визначення ІЦНС дозволяє вирішити завдання статистичного аналізу діяльності науки на основі бібліометричних вимірювань.

**Актуальність теми** зумовлена появою та розвитком різних систем, що проводять аналіз наукових публікацій та здійснюють визначення числа цитованості. Показники цитованості активно використовуються для формування різного роду рейтингів та статистичного аналізу. Кожна із систем має свої переваги та недоліки, що підвищує складність вибору такої, котра найкраще відповідає вимогам організації.

**Мета і задачі дослідження.** Визначити пошукову систему наукового індексу цитованості, як таку, що найкраще задовольняє вимогам українського відкритого електронного архіву. Для досягнення мети було поставлено такі задачі:

- ознайомитися із основними пошуковими системами наукового індексу цитованості;
- дослідити переваги та недоліки пошукових систем наукового індексу цитованості;
- визначити критерії вибору пошукової системи наукового індексу цитованості;
- вибрати кращу систему пошуку наукового індексу цитованості.

## АНАЛІЗ ОСТАННІХ ДОСЛІДЖЕНЬ ТА ПУБЛІКАЦІЙ

Важливу роль під час оцінювання розвитку наукових напрямів відіграють пошукові системи наукового індексу цитованості основними з яких є: Scopus, Російський індекс наукового цитування, Web of Science (WoS), Google Scholar, Microsoft Academic Search.

Scopus являє собою найбільшу в світі єдину реферативну базу даних, яка індексує більше 17000 найменувань науково-технічних і медичних жур-

налів приблизно 4000 міжнародних видавництв. Щодня ця база даних оновлюється і містить записи щодо першого тому та першого випуску журналів провідних наукових видавництв. Вона забезпечує неперевершену підтримку в пошуку наукових публікацій та пропонує посилання на всі цитати з широкого обсягу доступних статей.

Розробникам, які відповідають за інформаційне наповнення, доводиться оцінювати значну кількість джерел, щоб гарантувати якісне відображення наукової літератури, включаючи публікації у відкритому доступі (Open Access), праці наукових конференцій, а також матеріали, доступні тільки в електронній формі.

Пошукова система Scopus також пропонує апарат вимірювання ефективності досліджень, що є засобом контролю ефективності досліджень, які допомагають оцінювати авторів, галузі наук і журнали.

Щоб робота індексувалася в Scopus вона має бути опублікована в одному з журналів, що входять до цієї бази. База охоплює близько 25048 видань, але серед них є тільки 41 українське, тобто 0,16%. До того ж з 41 лише 18 активних видань, що постійно оновлюють інформацію [4, 5].

Російський індекс наукового цитування (РІНЦ) спочатку був задуманий як база даних і пошукова система, яка здатна забезпечити повноцінне покриття російського наукового простору. Зробити це можливо, лише використовуючи потужності ІТ-технологій. Несподівано виявилось, що середня кількість бібліографічних посилань на друковану статтю — 2,74, а на безкоштовну електронну версію — 7,3, тобто в 2,6 рази більше, ніж на друковану. Електронні статті цитуються в 4,5 рази частіше друкованих і це співвідношення швидко зростає.

Рейтинг цитування авторів неминуче призводить і до рейтингу наукових видань. Так само, як для окремого автора або наукового колективу, можна вивести рейтинг російських журналів. Він може бути загальним або зведеним за окремою тематикою, упорядкованим за загальною кількістю отриманих журналом цитувань.

Автор будь-якої наукової публікації може відшукати в РІНЦ свою роботу і одразу побачити скільки разів цитувалася його наукова робота в статтях інших авторів. Це найпростіша частина електронного сервісу. Але РІНЦ надає і послуги більш високого інформаційного рівня: надання переліку всіх статей, включених до бази РІНЦ, в яких цитується або є посилання на роботи авторів. Оскільки в «Науковій електронній бібліотеці» на elibrary.ru в повному тексті представлено більше 300 журналів (половина з яких є у відкритому доступі), можна одразу подивитися публікації, в яких міститься посилання на вашу роботу [6].

Google Академія є вільно доступною пошуковою системою, яка індексує повний текст наукових публікацій всіх форматів і дисциплін. Дата виходу в статусі бета-версії — листопад 2004 року. Індекс Google Scholar включає в себе більшість рецензованих он-лайн журналів Європи та Америки найбільших наукових видавництв. За функціями він схожий на вільно доступні системи Scirus від Elsevier, CiteSeerX та getCITED, а також на інструменти засновані на підписці, такі як Elsevier Scopus та Thomson ISI's Web of Science [7].

Google Scholar — це пошукова система, яка спеціалізується на індексації одного типу матеріалів — наукових публікацій (статей, книг, препринтів тощо). Google Scholar повідомляє користувачеві назву, фрагмент тексту і гіперзв'язок до документа.

Робот Google Scholar відвідує тільки сайти, що мають відношення до науки, і збирає у свій індекс інформацію про місцезнаходження і зміст наукових робіт. До бази даних потрапляють також відомості про безкоштовні повнотекстові статті, і ті, у яких доступні лише реферати чи бібліографічні описи.

Google Scholar містить відомості як про статті, які опубліковано в журналах, так і про ті, які зберігаються в репозиторіях або знаходяться на сайтах наукових колективів чи окремих вчених. Google Scholar містить відомості не тільки про он-лайнкові, а й про друковані статті.

Робот Google Scholar індексує онлайнкові наукові публікації. Якщо в такому документі в списку літератури виявляється посилання на офлайнвий документ, бібліографічний опис такого друкованого документа теж потрапляє в базу даних Google Scholar. У списку результатів пошуку офлайнкові статті мають позначку [Citation]. Список результатів пошуку містить гіперзв'язки, що ведуть до веб-сторінок, де є відомості про статтю (як мінімум — бібліографічний опис). У списку відсутні гіперзв'язки до офлайнвих першоджерел.

У списку результатів пошуку може бути декілька посилань на матеріали, пов'язані з однією й тією ж статтею (наприклад, посилання на сайт видавництва, на сайт агрегатора, на реферативну базу даних, на персональний сайт автора). У списку результатів пошуку посилання на безкоштовні повні тексти публікацій мають позначки [PDF]. У списку може бути декілька посилань на декілька повнотекстових версій однієї й тієї ж статті (наприклад, на остаточну версію на сайті видавництва і на препринт на сайті автора).

Пошукова програма Google Scholar працює за тими ж правилами, що й пошукова програма Google, але індекси Google та Google Scholar — це різні бази даних. Google Scholar виконує не тільки інформаційні, але й наукометричні функції.

Із списку результатів пошуку за гіперзв'язком «Cited by», можна отримати відомості про те, скільки і які саме документи посилаються на конкретну публікацію в межах бази даних Google Scholar. Показник у «Cited by», відбиває ступінь авторитетності і популярності публікації [8, 9].

Web of Science — база даних, яку публікує Філадельфійський інститут наукової інформації, визначаючи журнали з найвищим рейтингом (Thomson Reuter Master Journal List).

Система Web of Science являє собою сукупність різноманітних баз даних, що функціонують на платформі ISI Web of Knowledge. Для визначення індексу цитованості вченого використовуються БД Science Citation Index Expanded та Social Sciences Citation Index. У WoS закладено можливість пошуку процитованих робіт не тільки за першим автором, а й за співавторами, за умови, що джерело, в якому міститься стаття, розписано в Web of Science. Можна задати ретроспективу індексу цитованості і хронологічні рамки процитованих робіт.

Існує два способи визначення індексу цитованості з використанням WoS. Перший (спрощений) складається з двох етапів: спочатку здійснюється пошук публікацій шуканого автора, потім автоматично (кнопка <Create Citation Report>) визначається їх цитованість. Індекс цитованості при цьому буде «усіченим» — обмежений тільки статтями і тільки з тих журналів, які

опрацьовуються в Web of Science. Другий (більш складний спосіб) вимагає ручного налаштування, але індекс цитованості буде більш повним і буде відображати цитованість монографій, дисертацій та авторефератів, патентів, статей зі збірників і тих журналів, що не опрацьовуються в WoS.

Для того, щоб отримати найбільш достовірні дані щодо цитованості вченого, слід мати повний список його публікацій, оформлений відповідно до чинних державних стандартів на бібліографічний опис документів. Це необхідно з декількох причин:

- для обліку різних варіантів написання прізвища автора іноземними мовами;
- для видалення посилань на публікації з однаковими прізвищами [10].

База даних є платною, на сайті (<http://science.thomsonreuters.com/mjl/>) можна завантажити тільки 3 тематичні списки журналів у форматі PDF (публікації за напрямленням — гуманітарні науки, природничі науки та соціальні) і простежити, чи входять журнали, в яких публікувалися статті, до цих списків. Визначення імпаکت-фактора журналу можна здійснювати безпосередньо на сайті цього журналу.

База даних є досить великою і містить близько 7000 назв журналів, тому для зручності пошуку можна використовувати тематичний пошук, а після з'ясування наявності журналу в списку проводити пошук імпаکت-фактора безпосередньо в повному списку [11].

Microsoft Academic Search є вільною академічною пошуковою системою наукових досліджень розробником якої є Microsoft. Вона надає безліч способів для вивчення наукових публікацій, авторів, конференцій, журналів, організацій за ключовими словами. Academic Search створюється як конкурент Google Scholar. Об'єкти в результатах пошуку сортуються за двома критеріями: релевантність до запиту і глобальне значення. Актуальність оцінки об'єкта обчислюється його атрибутом, а також важливість оцінки об'єкта визначається його відносинами з іншими об'єктами [12, 13].

## **ВИДІЛЕННЯ ПРОБЛЕМИ**

Пошукові системи наукового індексу цитованості мають досить широке застосування і використовуються як для пошуку актуальних наукових праць із використанням пошукових фільтрів, так і для визначення кількісного показника цитованості праць із можливістю простежування розвитку наукової думки. Ці системи мають різне походження, різні алгоритми індексування документів та підходи визначення кількості посилань. Такий стан систем пошуку наукового індексу зумовлює складність і неоднозначність проведення оцінок і порівнянь між системами. Вибираючи систему для відкритого наукового архіву на теренах України передусім необхідно з'ясувати низку моментів, які впливають на здійснення найоптимальнішого вибору:

- Який обсяг індексування наукових україномовних публікацій проводить система? Для повноцінного аналізу документів, що наявні в науковому електронному архіві, існує необхідність в індексації кожного з документів незалежно від того, де його було опубліковано. Такий механізм дозволяє найбільш ефективно використовувати дані.
- Наявність можливостей автоматизації запитів до системи та отримання числа цитованості.

- Наявність процедур розширеного пошуку. Підтримка системою різних мов пошуку.
- Відображення повного списку публікацій за автором.
- Правдивість показника цитованості наукового документа. Система має враховувати існування посилань з однієї статті, яка може існувати у різних форматах, перевіряти на достовірність посилання (наприклад за науковим напрямом, відкидаючи однакові прізвища).
- Регулярність проведення індексування документів.
- Вимоги до розширення системи. Можливість інтеграції з інформаційними системами.

Будемо розглядати п'ять пошукових систем наукового індексу цитованості кожна з яких має свої переваги та недоліки (Scopus, РІНЦ, Web of Science, Google Scholar, Microsoft Academic Search). Виділимо критерії порівняння: доступність використання системи, охоплення українських видань, точність даних.

Критерій «доступність використання системи» є важливим чинником під час вибору системи, оскільки враховуються як затрати на пошук цитованості так і можливості доступу, використання та аналізування додатковими програмними засобами.

Критерій «охоплення українських видань» є найважливішим критерієм оцінювання, враховуючи те, що існує необхідність аналізу відкритих наукових архівів України, який характеризує обсяги індексування українських видань.

Критерій «точність даних» є наступним важливим критерієм оцінювання, під яким розуміється правдивість числа цитованості документа, надання повної інформації про публікації за автором, відкидання хибних результатів.

Побудуємо дерево цілей, яке складається з рівнів: альтернатив, критеріїв та цілей (рис. 1).

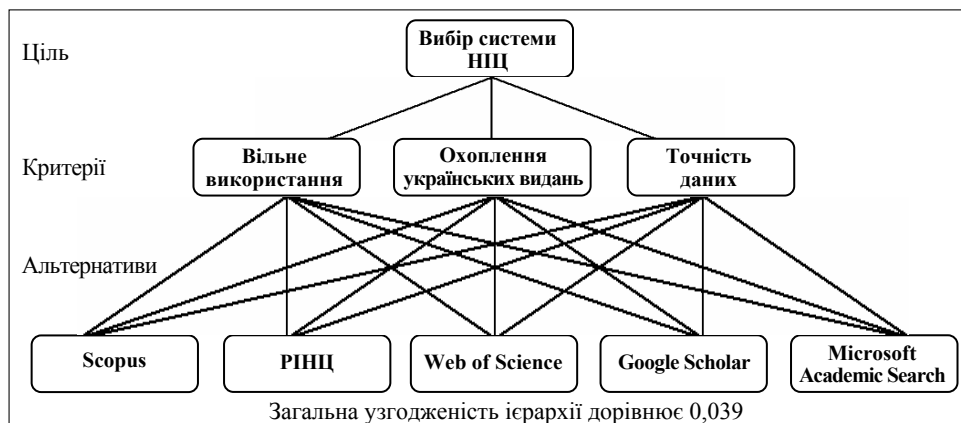


Рис. 1. Дерево цілей

Методом аналізу ієрархій проведемо попарне порівняння альтернатив за визначеними критеріями. Метою є вибір пошукової системи наукового індексу цитування для її використання у подальшому аналізі відкритих українських наукових архівів. Враховуючи мету, найбільше ваги приділяється критерію «охоплення українських видань». Результати обчислень подано у табл. 1 та діаграмі рис. 2.

Таблиця 1. Результати обчислень

Фактори (Міра)		Scopus (0,105)	РІНЦ (0,169)	WoS (0,162)	GS (0,235)	MAS (0,329)
Вузол	<b>Вільне використання</b>					
Міра	0,282					
<b>Матриця попарних порівнянь</b>	№	1	2	3	4	5
	1	1,000	0,333	2,000	0,111	0,250
	2	6,000	1,000	6,000	0,167	0,333
	3	0,500	0,167	1,000	0,111	0,125
	4	9,000	9,000	9,000	1,000	2,000
	5	9,000	3,000	8,000	1,000	1,000
<b>Характеристика матриці</b>	$\lambda_{\max} = 5,192$ $I3 = 0,048$ $BY = 0,043$					
Вузол	<b>Охоплення українських видань</b>					
Міра	0,669					
<b>Матриця попарних порівнянь</b>	№	1	2	3	4	5
	1	1,000	3,000	0,500	0,111	2,000
	2	0,333	1,000	0,500	0,125	1,000
	3	2,000	2,000	1,000	0,111	1,000
	4	9,000	8,000	9,000	1,000	9,000
	5	0,500	1,000	1,000	0,111	1,000
<b>Характеристика матриці</b>	$\lambda_{\max} = 5,286$ $I3 = 0,071$ $BY = 0,064$					
Вузол	<b>Точність даних</b>					
Міра	0,088					
<b>Матриця попарних порівнянь</b>	№	1	2	3	4	5
	1	1,000	6,000	4,000	8,000	8,000
	2	0,167	1,000	1,000	4,000	4,000
	3	0,250	1,000	1,000	6,000	6,000
	4	0,125	0,250	0,167	1,000	1,000
	5	0,125	0,250	0,167	1,000	1,000
<b>Характеристика матриці</b>	$\lambda_{\max} = 5,224$ $I3 = 0,056$ $BY = 0,050$					

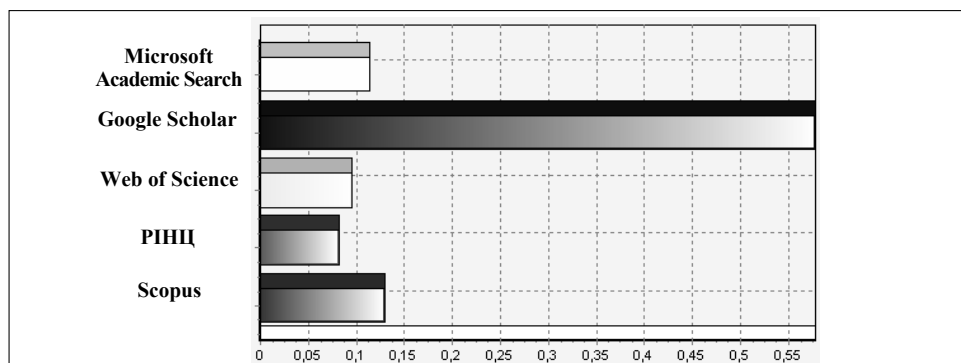


Рис. 2. Діаграма результату: Scopus — 0,13; Scopus — 0,082; Web of Science — 0,096; Google Scholar — 0,578; Microsoft Academic Search — 0,114

## АНАЛІЗ ОТРИМАНИХ РЕЗУЛЬТАТІВ

Досліджуючи пошукові системи наукового індексу цитованості, сформовано їх порівняльну таблицю (табл. 2). Досить мала кількість українських видань індексується системами, зокрема: Web of Knowledge включає 24 українських видань до БД, п'ять із них мають встановлений імпаکت-фактор; Scopus — 36 українських видань, у тому числі 19 назв журналів, які активно індексуються фахівцями Ельзевір, 16 назв журналів, індексацію яких, станом на 2009 р., припинено, та один збірник матеріалів конференції [14]. Google Scholar працює як пошукова система Google, а також можна направляти пошукового робота на домен, тим самим охоплюючи значні наукові ресурси. Microsoft Academic Search на сьогодні ще перебуває у формі тестування, але вже можна знайти українських вчених у її БД. Microsoft Academic Search відрізняється сучасними візуальними інструментами.

Фільтрування надлишкової інформації дозволяє отримати лише необхідні дані за запитом, оскільки відкидає дані про авторів з однаковими прізвищами, що перевіряються приналежністю до галузі науки, місцем праці, посадою тощо. Найкраще реалізовано в системі Scopus, а тому вважається, що дані цієї системи є найбільш правдивими. Негативним фактором є те, що ускладнюється реєстрація видань для індексування і більшість українських видань ігноруються. Кожна система надає можливості розширеного пошуку (роком видання, напрямом науки тощо).

**Таблиця 2.** Порівняльна характеристика п'яти пошукових систем наукового індексу цитованості

№	Категорія	Scopus	РІНЦ	Web of Science	Google Scholar	Microsoft Academic Search
1	Індексування наукових україномовних публікацій	–	+–	–	+	+
2	Можливостей автоматизації запитів	+–	+	+–	+	+
3	Фільтрування надлишкової інформації	+	+–	+–	–	–
4	Правдивість числа цитованості наукового документа	+	+–	+–	+–	+–
5	Можливість інтеграції з інформаційними системами	–	–	–	+	+–
6	Відкритість (доступність)	–	+	–	+	+
7	Можливості розширеного пошуку	+	+	+	+	+
8	Активність індексування наукових документів	+–	+–	+–	+–	+–

У табл. 2, під «+» розуміємо високий рівень виконання відповідної категорії; під «+ –» — часткове виконання або наявність обмежень; під «–» — відсутність можливостей виконання аналізу наукових публікацій або лише мінімальне виконання (при задоволенні специфічних умов виконання та із використанням додаткових засобів).



## ВИСНОВКИ

Проаналізовано найбільш поширені пошукові системи наукового індексу цитування, наявність українських видань та прізвищ науковців у цих базах даних, основні можливості запитів та розширеного пошуку. Індеси цитування, які взято із закордонних підрахунків, потребують поправок. Враховуючи те, що в Україні не має власної системи, яка б індексувала українські наукові видання, визначала показник посилань та відповідала власним потребам і стандартам, існує необхідність вибору із існуючих систем. Вхідження лише декількох видань до переліку індексування пошуковими системами числа цитованості не може адекватно оцінювати український науковий потенціал. Постає проблема вибору такої системи, враховуючи усі переваги та недоліки. Вибір — це лише перший крок до формування правил аналізу відкритих українських наукових архівів.

Визначені критерії порівняння дозволили вибрати кращу систему методом аналізу ієрархій, але оскільки було визначено і її недоліки, то вона потребує подальшої адаптації та вдосконалення, що вирішуватиметься у подальшій роботі.

## ЛІТЕРАТУРА

1. *Індекс* научного цитування. — <http://ria-stk.ru/mmq/adetail.php?ID=47489>.
2. *Усачёв А.С.* Российский индекс научного цитирования. — <http://dlib.eastview.com/browse/doc/21773739>.
3. *Писляков В.В.* Зачем создавать национальные индексы цитирования? // Научные и технические библиотеки. — 2007. — №2. — С. 65–71.
4. *Scopus*. — <http://ntb.pstu.edu/index.php?id=59&L=0>.
5. *Jacso P.* As we may search — Comparison of major features of the Web of Science, Scopus and Google Scholar citation-based and citation-enhanced databases // Current Science. — 2005. — **89**, № 9, 10. — P. 1537–1547.
6. *PIHЦ*. — <http://ntb.pstu.edu/index.php?id=61&L=0>.
7. *Додонов А.Г., Ланде Д.В., Путятин В.Г.* Современные поисковые технологии — проблемы и некоторые пути их решения // Реєстрація, зберігання в обробка даних. — 2010. — **12**, №3. — С. 36–35.
8. *Google Scholar*. — <http://www.abc.chemistry.bsu.by/intro/part10/04.html>.
9. *Bar-ilan J.* Which h-index? — A comparison of WoS, Scopus and Google Scholar // Scientometrics. — 2008. — **74**, № 2. — P. 257–271.
10. *Рекомендации* относительно поиска журналов, которые входят в международные научно-метрические базы данных (SCOPUS, Web of Science). — <http://ru.snu.edu.ua/index.php?mode=392>.
11. *Falagas M.E.* Comparison of PubMed, Scopus, Web of Science, and Google Scholar: strengths and weaknesses / Falagas M.E., Pitsouni E.I., Malietzis G.A., Pappas G. // FASEB J. — 2008. — P. 338–342.
12. *Microsoft Academic Search: get another h-index calculation!* — [http://blogs.warwick.ac.uk/libresearch/entry/microsoft\\_academic\\_search/](http://blogs.warwick.ac.uk/libresearch/entry/microsoft_academic_search/).
13. *What is Microsoft Academic Search?!* — <http://academic.research.microsoft.com/About/Help.htm>.
14. *Костенко Л.Й., Соловяненко Д.В.* Звіт про роботу семінару «Бібліотека та наукометрія: світовий досвід, українська перспектива»: (матеріали міжнародної наукової конференції «Бібліотечно-інформаційний комплекс у контексті розвитку суспільства знань»). — [http://www.nbuv.gov.ua/new/09\\_kiev/seminar\\_zvit.html](http://www.nbuv.gov.ua/new/09_kiev/seminar_zvit.html).

Надійшла 27.02.2012