

УДК 004.738.5

Н.О. Ткаченко

ДВНЗ «Донецький національний технічний університет», м. Донецьк
кафедра автоматики і телекомунікацій, науково-технічна бібліотека ДонНТУ
E-mail: nsot@library.dgtu.donetsk.ua

РОЗРОБКА КРИТЕРІЇВ ДЛЯ ОЦІНКИ ІНФОРМАЦІЙНИХ НАУКОВИХ РЕСУРСІВ В ІНТЕРНЕТ

У статті проаналізовані сучасні методи оцінки інформаційних ресурсів, враховуючи різні критерії. Наведені переваги та недоліки існуючих систем оцінки. Сформована порівняльна характеристика українських університетів у відповідності з існуючими системами оцінки діяльності організації. Надано список українських журналів у світових наукометричних базах даних, що використовують різні критерії для оцінки журналу. Зроблено аналіз актуальних методів багатокритеріальної оптимізації з вибором оптимального для оцінки інформаційних ресурсів.

Ключові слова: інформаційні ресурси, наукометричні показники, багатокритеріальна оптимізація.

Успішна діяльність університетів та наукових організацій безпосередньо залежить від наукових досягнень у різноманітних галузях знань. Тому висвітлення результатів своїх досліджень на конференціях та семінарах, обмін з колегами своїми науковими досягненнями виступають одними з головних завдань багатьох інституцій. Але у зв'язку зі зростаючим обсягом інформації, виникла проблема коректної оцінки наукового внеску окремого вченого або інституції в цілому, що започаткувала розвиток нової наукової дисципліни - наукометрія. Окремі критерії, що були розроблені в рамках цієї дисципліни, дозволили наочно оцінити внесок окремих інституцій в розвиток науки.

З розвитком всесвітнього павутиння перед науковим суспільством відкрилися нові можливості - спілкування з колегами в режимі онлайн, оприлюднення результатів своїх досліджень в найкоротші строки. Оцінивши значні переваги Інтернет, університети створюють власні веб-сторінки, репозитарії та електронні версії журналів, а окремі вчені, навіть, створюють власні сторінки в Інтернет [1]. Але обсяг доступної в Інтернет інформації, що зростає по годині, ускладнює її вірне оцінювання. Саме тому актуальним виступило завдання оцінки наукових ресурсів організацій у всесвітньому павутинні.

На сьогоднішній день існують різноманітні критерії оцінки інформаційних ресурсів, які умовно можна розділити на (рис. 1):

- 1) системи оцінки сайту організації (система Google Page Rank, система Yandex ТІЦ, Webometric);
- 2) критерії оцінки діяльності автора (індекс Хірша, Е-індекс, Google Scholar Citations);
- 3) наукометричні бази даних (Scopus, Web of Science, РІНЦ).

Розглянемо більш детально кожен із них.

- 1) аналіз сайту інституції за допомогою системи Google Page Rank.

PageRank – алгоритм ранжування сайтів, що використовується компанією Google [1]. Показник PageRank – це простий числовий бал від 1 до 10.

В основі алгоритму лежить аналіз веб-сторінок за допомогою зовнішніх посилань. Остаточна формула для розрахунку показника PageRank має наступний вид:

$$WPR(n) = (1 - d) + d \sum_{m \in B(n)} WPR(m) W_{(m,n)}^{in} W_{(m,n)}^{out} \quad (1)$$

де d – коефіцієнт затухання, $W_{(m,n)}^{in}$ - вага вхідних посилань, $W_{(m,n)}^{out}$ - вага вихідних посилань. Конкретна вага кожної сторінки призначається експертами з залежності від релевантності інформації на цій сторінці.

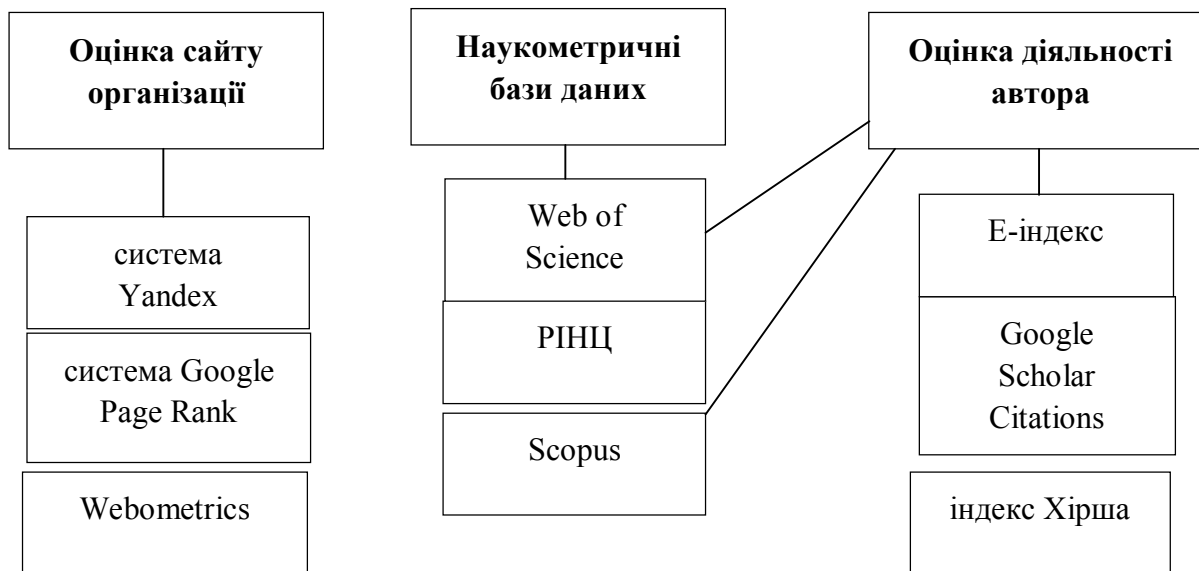


Рисунок 1 - Критерії оцінки інформаційних ресурсів

2) система аналізу сайту за допомогою Yandex.

Гідно оцінивши проблему релевантного пошуку інформації у Інтернеті система Yandex у 2000 році впроваджує свій власний показник оцінювання сайтів, відмінний від PageRank Google.

ТІЦ – тематичний індекс цитування, який використовується пошуковою системою Yandex для ранжування сайтів. В основі лежить аналіз посилань на сторінку з урахуванням ваги посилання. Методика призначення величини вагових коефіцієнтів не надається у вільному доступі та відома лише розробникам. Величина кінцевого показника ТІЦ розраховується за допомогою суми ваг посилань на сторінку (табл. 1).

3) Webometrics аналізує присутність вузів в Інтернет.

Розроблена методика оцінює сайт за чотирма параметрами: Size (S), Visibility (V), Rich Files (R), Scholar (Sc) [3]. Остаточний показник для кожного вузу розраховується за формулою з такими величинами вагових коефіцієнтів:

$$Criteria = 0,5V + 0,2S + 0,15R + 0,15Sc \quad (2)$$

Із таблиці видно, що місце університету у рейтингу залежить від обраної системи оцінки і, у деяких випадках, може різнитися у рази. Тому можна зробити висновок, що відсутність єдиної методики оцінки значно ускладнює коректну оцінку сайту організації.

Окрім наявності веб-сторінки в Інтернет, де можуть висвітлюватися результати діяльності автора або університету, важливим виступає показник зацікавленості колег у цих наукових досягненнях методом цитування наукових публікацій.

Питаннями оцінки якості наукових досліджень вже давно займаються бібліотечні співробітники та спеціалісти з наукометрії, а найпотужніші організації світу підтримують розвиток наукометричних баз даних. Деякі країни вже застосовують власні національні

наукометричні показники, але розробка загальної системи критеріїв оцінки інформаційних ресурсів все ще виступає актуальним завданням у світі [4].

Таблиця 1

Порівняльна характеристика показників PageRank Google, ТІЦ та Webometrics

Назва університету	PageRank Google	ТІЦ	Webometrics
Київський національний університет імені Т.Шевченка	1	1	2
Київський політехнічний університет	2	2	1
Національний університет «Києво-могілянська академія»	2	6	24
Національний університет «Львівська політехніка»	2	3	5
Донецький національний технічний університет	3	4	3
Донецький національний університет	3	5	9

На сьогоднішній день у науковому суспільстві прийнято враховувати наступні наукометричні показники:

- індекс Хірша або h-індекс – наукометричний показник, запропонований американським фізиком Хорхе Хіршем у 2005 році для чисельної оцінки наукової активності вченого. В основі розрахунку лежать дві характеристики: загальна кількість публікацій вченого та кількість цитувань його робіт. Вчений має індекс h за умови, що h з його N статей цитуються мінімум h раз кожна, а решта статей (N-h) цитуються не більше, ніж h раз кожна. Для відстеження авторами своєї наукової активності в мережі Інтернет відома компанія Google запровадила додатковий сервіс під назвою Google Scholar Citations, що базується на двох показниках (h-індекс та i10-індекс). Для його використання необхідно лише створити свій профіль у системі Google.

- E-індекс. Критика індексу Хірша призвела до виникнення додаткових наукометричних показників. За словами професора Готта, h-індекс не враховує індивідуальний вклад самого вченого або наявність прихованого цитування. Саме тому, у 2010 році він пропонує використання, так званого, E-індексу. Однак, навіть цей показник залишається не досконалим, бо не враховує усіх параметрів, наприклад підвищення цитування своїх робіт методом самоцитування.

Для відстеження цитування наукових публікацій у виданнях або оцінки наукової активності організації науковці використовують міжнародні наукометричні бази даних, серед яких провідне місце займають дві: Scopus та Web of Knowledge.

1) Scopus – реферативна база даних видавництва Elsevier. База індексує видання, опубліковані після 1996 р. Цитованість кожної статті підраховується методом аналізу у статті списку використаної літератури Scopus використовує два показники наукової активності:

- h-індекс – для оцінки активності окремого вченого або організації;
- SJR (SCImago Journal Rank) – показник компанії SCImago для оцінки наукових журналів.

2) Web of Knowledge, на відміну від Scopus, у своїй практиці широко використовує такий показник, як імпакт-фактор журналу. Розрахунок імпакт-фактора заснований на трьохлітньому періоді.

3) Питаннями наукометрії займаються і вчені колишнього СРСР, зокрема компанія «Наукова електронна бібліотека (ELIBRARY.ru)» у 2005 році вводить такий показник, як російський індекс наукового цитування – РІНЦ для аналізу публікацій російських вчених.

Використання різних критеріїв оцінки журналу (імпакт-фактор, РІНЦ, SJR) у вищезазначених базах призвело до виникнення власних рейтингів (табл. 2) та відсутності єдиного критерію оцінки, що ускладнює однозначне визначення авторитетності журналу. Тому розробка єдиної системи оцінки окремого журналу досі залишається актуальною.

Таблиця 2

Список українських журналів у наукометричних базах даних

Назва журналу	Science Journal Rankings	Імпакт-фактор	РІНЦ
Symmetry, Integrability and Geometry: Methods and Applications	0,278 (3)	1.071 (1)	не індексується
Ukrainian Journal of Physical Optics	0.851 (1)	0,341 (6)	не індексується
Condensed Matter Physics	0,436 (2)	0,811 (2)	0,409 (3)
Сверхтвердые материалы	не індексується	0.785 (3)	не індексується
Biopolymers and cell	0,138 (4)	не індексується	0,195 (4)
Фізика низьких температур	не індексується	0,73 (4)	0,529 (1)
Теоретическая и экспериментальная химия	не індексується	0,509 (5)	0,450 (2)

Доступність цих баз лише у вигляді передплачених ресурсів загостило проблему їх вільного використання. Постійне підвищення цін на наукові журнали та скорочення фінансування наукових організацій викликало обурення вчених по всьому світу. Це призвело до зародження руху відкритого доступу, який на сьогоднішній день активно розвивається двома шляхами: журнали відкритого доступу та репозитарії [5]. І якщо створення відкритих журналів ще залишається для багатьох університетів проблемою, то ідею організації власного репозитарію втілили в життя більше 1500 організацій по всьому світу [6]. Таким чином, перед науковцями повстало нове завдання щодо врахування наукового внеску репозитаріїв у діяльність організації. Саме тому при оцінці діяльності автора (h-індекс) компанія Google почала аналізувати присутність його робіт в архівах відкритого доступу, а іспанська лабораторія Cybermetrics Lab навіть заснувала світовий рейтинг репозитаріїв.

Відсутність єдиної методики оцінювання інформаційних ресурсів призвело до проблеми однозначності цієї оцінки. Відповідно до цього актуальним виступило завдання розробки єдиного критерію оцінки ресурсів на основі існуючих, що досягається шляхом об'єднання у своєму складі декількох критеріїв. Таким чином загальна формула розрахунку нових критеріїв для різних видів ресурсів матиме наступний вигляд:

а) оцінка автора:

$$A = \sum_{i=1}^s \alpha_i a_i(X) = \alpha_1 a_1(X) + \alpha_2 a_2(X) + \alpha_3 a_3(X) + \alpha_4 a_4(X) \quad (3)$$

де α_i - ваговий множник, $a_1(X)$ – h-індекс автора в GoogleScholar, $a_2(X)$ - i10-індекс автора в GoogleScholar, $a_3(X)$ - h-індекс автора в Scopus, $a_4(X)$ - кількість робіт автора в GoogleScholar. Оскільки рух відкритого доступу набуває стрімкого розвитку основний акцент при виборі вагових коефіцієнтів повинен робитися на роботи у GoogleScholar. Враховуючи це пропонуються наступні вагові коефіцієнти: $\alpha_1 = 0,4$, $\alpha_2 = 0,1$, $\alpha_3 = 0,3$, $\alpha_4 = 0,2$, але їх остаточний вибір може здійснюватися окремою організацією відповідно до своїх уподобань.

b) оцінка сайту університету:

$$U = \sum_{i=1}^s \beta_i u_i(X) = \beta_1 u_1(X) + \beta_2 u_2(X) + \beta_3 u_3(X) \quad (4)$$

де β_i - ваговий множник, $u_1(X)$ – місце університету у рейтингу Webometrics, $u_2(X)$ - оцінка сайту системою Google; $u_3(X)$ - оцінка сайту системою Yandex.

c) оцінка журналу:

$$G = \sum_{i=1}^s \gamma_i g_i(X) = \gamma_1 g_1(X) + \gamma_2 g_2(X) + \gamma_3 g_3(X) \quad (5)$$

де γ_i - ваговий множник, $g_1(X)$ – присутність журналу у директорії відкритого доступу DOAJ, $g_2(X)$ - SJR журналу, $g_3(X)$ - імпакт-фактор журналу, $g_4(X)$ - місце у національній наукометричній базі даних (наприклад, РІНЦ). Вибір вагових коефіцієнтів здійснюється з урахуванням впливовості світових наукометричних баз даних, тобто з акцентом на показники журналу у базах даних Scopus та Web of Science. Але можливість безкоштовного використання показника SJR дозволяє зробити вибір на його користь, як найбільш впливового. Результати “what – if” аналізу для різних комбінацій вагових коефіцієнтів приведені на рисунках 2 і 3 для рейтингу вибраних українських університетів та наукових журналів відповідно.

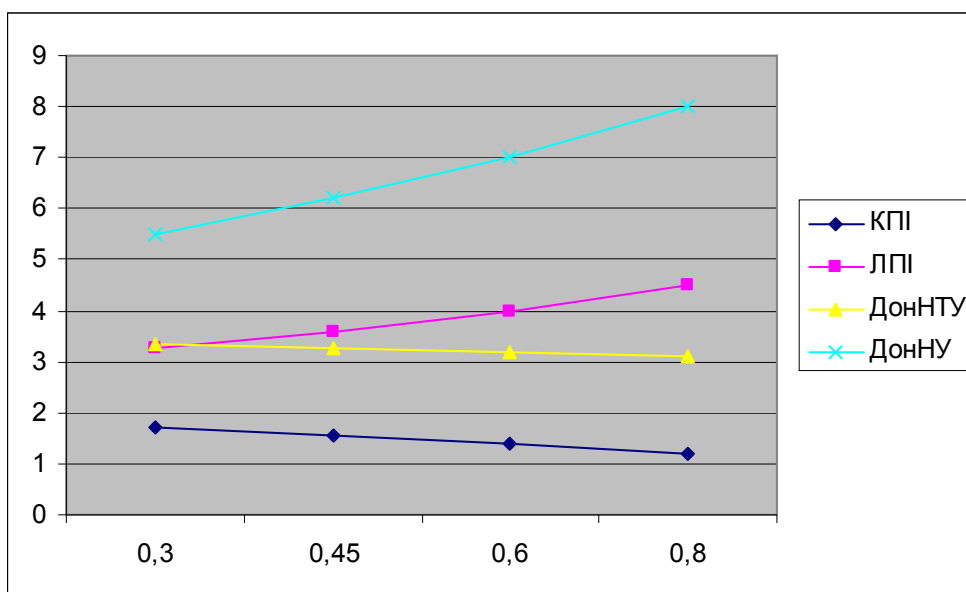


Рисунок 2 – Порівняльний рейтинг університетів

На рисунку приведена залежність показника університету від коефіцієнта β_1 , решта вагових множників пропорційно доповнюють значення β_1 до 1. Перевага віддається системі Webometrics, що більш повноцінно оцінює веб-сайт університету.

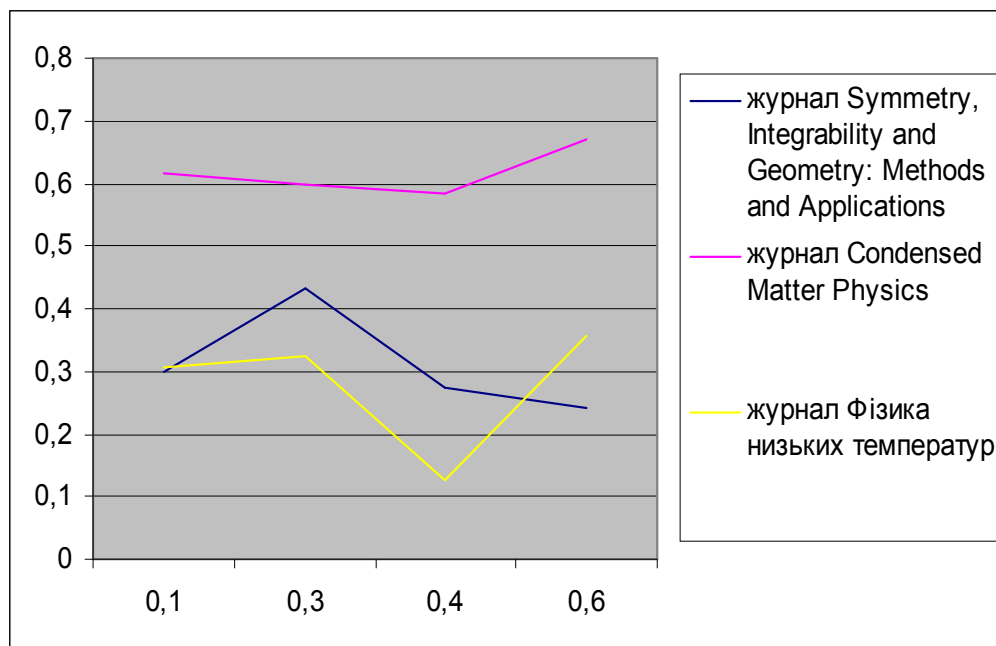


Рисунок 3 – Порівняльний рейтинг журналів

На рисунку приведена залежність показника журналу від коефіцієнта g_2 , решта вагових множників пропорційно доповнюють значення g_2 до 1.

З розвитком мережі Інтернет проблема оцінки інформаційних ресурсів почала стосуватися не лише науковців та організацій, де вони працюють, а й звичайного користувача. Необмежене зростання різноманітної інформації у всесвітній мережі викликало проблему її релевантного пошуку. У зв'язку з цим відомі пошукові системи почали вдосконалювати свої сервіси, використовувати додаткові критерії. Але ця проблема залишається відкритою й донині, тому розробка загальної системи пошуку інформації залишається актуальним завданням. Для отримання найбільш релевантних даних ця система повинна всебічно оцінювати інформаційний ресурс, використовуючи розглянуті вище критерії.

При наявності декількох критеріїв, що суперечать один одному для отримання оптимального рішення на практиці використовують методи багатокритеріальної оптимізації. Відомим є ряд способів вирішення багатокритеріальних задач [7]:

- 1) оптимізація одного критерію (найбільш важливого), інші при цьому відіграють роль додаткових обмежень (метод епсилон-обмежень);
- 2) упорядкування заданої множини критеріїв і послідовна оптимізація по кожному з них (метод послідовних поступок);
- 3) метод справедливого компромісу, при якому відносний рівень зниження якості по одному або декільком критеріям не перевершує відносного рівня підвищення якості по решті критеріїв;
- 4) зведення багатьох критеріїв до одного шляхом введення апіорних (експертних) вагових коефіцієнтів для кожного з критеріїв (більш важливий критерій отримує більш високу вагу) (метод вагових множин).

Обчислення загального критерію здійснюється шляхом поєднання загальних оцінок перерахованих вище критеріїв з врахуванням внеску кожної з них. Отже доречним буде

застосування методу зважених сум [8] при якому задача багатокритеріальної оптимізації зводиться до вирішення задачі умовної оптимізації зі скалярним критерієм оптимальності: Внесок кожного з них оцінюється шляхом призначення окремому критерію свого вагового коефіцієнту:

$$\max_{x \in D_x} \varphi(X) = \max_{x \in D_x} \sum_{i=1}^s \lambda_i \phi_i(X) = \lambda_1 \phi_1(X) + \lambda_2 \phi_2(X) + \lambda_3 \phi_3(X) \quad (6)$$

де λ_i – ваговий множник, $\phi_1(X)$ – загальна оцінка автора, $\phi_2(X)$ – загальна оцінка організації, $\phi_3(X)$ – оцінка журналу.

Таким чином, можна зробити висновок, що визначення важливості кожного критерію за ваговими коефіцієнтами, а також розробка єдиної системи експертної оцінки інформаційних наукових ресурсів потребує подальших досліджень у цій області.

Висновки:

- 1) у статті проаналізовано існуючі методи оцінки інформаційних ресурсів за різними критеріями;
- 2) розглянуті можливості багатокритеріальної оптимізації для оцінки усіх видів ресурсів;
- 3) запропоновані вдосконалені критерії оцінки інформаційних ресурсів в Інтернет.

Список використаної літератури

- 1) Аноприенко А.Я. Университет в современном информационном пространстве: тенденции, рейтинги и опыт развития портала магистров ДонНТУ / А. Я. Аноприенко // Наукові труди ДонНТУ. Серія «Інформатика, кібернетика та обчислювальна техніка». – 2011. – Вып. 13(185). – С. 224-235.
- 2) PageRank [Электронный ресурс]. – Режим доступа: http://en.wikipedia.org/wiki/Page_Rank#. – Заглавие с экрана.
- 3) Сайт рейтинга Ranking Web of World Universities (Испания) [Электронный ресурс]. – Режим доступа: // <http://www.webometrics.info>. – Заглавие с экрана.
- 4) Солов'яненко Денис Політика індексації видань у наукометричних базах даних Web of Science та SciVerseScopus / Денис Солов'яненко // Бібліотечний вісник. - 2012. - № 1. - С. 6-21
- 5) Воропаєва, Вікторія The ELibUkr project: new opportunities for digital data access in the Donetsk National Technical University Library / Вікторія Воропаєва // Матеріали 32-ї щорічної IATUL конференції “Libraries for An Open Environment: strategies, technologies and partnerships”, (Варшава, 29 травня – 2 червня 2011р.). – Варшава, 2011.
- 6) Ткаченко Н.А. Возможности университетского репозитария для представления научных достижений ученых ДонНТУ в мировом информационном пространстве / Н. А. Ткаченко; ред. кол.: Ж. В. Марфіна, О.В. Бикова, К.О. Шарамет // Сучасна бібліотека в інноваційному освітньому просторі: матеріали наук.-практ. конф., (м. Луганськ, 17 - 18 берез. 2011 р.). – Луганськ: Вид-во ДЗ «ЛНУ імені Тараса Шевченка», 2011. - С. 192 – 197.
- 7) Многокритериальная оптимизация. Теория, вычисления и приложения / Р. Штойер, Р. Штойер; пер. с англ. С.М. Столяровой; под ред. А.В. Лотова. - М.: Радио и связь, 1992. – 504 с.: ил.
- 8) Patrick Ngatchou Pareto Multi Objective Optimization [Электронный ресурс] / Patrick Ngatchou, Anahita Zarei, M. A. El-Sharkawi. - Режим доступа: <http://thescipub.com/html/10.3844/ajassp.2010.840.84>.

Надійшла до редакції:
30.04.2013

Рецензент:
д-р техн. наук, проф. Скобцов Ю.О.

Н.А. Ткаченко

ГВУЗ «Донецкий национальный технический университет»

Разработка критериев оценки информационных научных ресурсов в интернет. В статье проанализированы современные методы оценки информационных ресурсов, учитывая различные критерии. Приведены преимущества и недостатки существующих систем оценки. Сформирована сравнительная характеристика украинских университетов в соответствии с существующими системами оценки деятельности организации. Приведен список украинских журналов в мировых наукометрических базах данных, использующих различные критерии для оценки журнала. Сделан анализ актуальных методов многокритериальной оптимизации с выбором оптимального для оценки информационных ресурсов.

Ключевые слова: многокритериальная оптимизация, наукометрические показатели, информационные ресурсы.

N.A. Tkachenko

Donetsk National Technical University

Development of the Criteria to Evaluate Scientific Information Resources in Internet. This article describes the problem of information resources evaluation. Existing criteria for evaluating resources in the Internet were studied in detail, these resources were divided into: systems of web sites evaluation, systems of evaluation of individual authors and citation databases. The paper provides techniques of calculating criteria for each system of site rating, which include the following systems: Google PageRank, Yandex and Webometrics. Comparative characteristic of Ukrainian universities in three different systems for evaluating sites was formed. The paper reviews existing citations indicators to assess a particular author: Hirsch index, Google Scholar Citations, and E-index. Possibilities and disadvantages of specific scientometric indicators were shown. Different citation databases (Web of Knowledge, Scientific Digital Library, Scopus), which employ different criteria for evaluating journal (impact factor, author, SJR), are used for tracking citations of scientific publications in journals. Using different criteria for evaluating scientific journals leads to the problem of having a unique ranking. The analysis of Ukrainian journals in various citation databases was done. The paper describes two ways for developing an open access, especially university repositories. A technique for the formation of the world rankings repositories, based on four indicators: Size (S), Visibility (V), Rich Files (R), Scholar (Sc) was examined. In practice we use methods of multicriteria optimization to find the optimal solutions on several parameters. The paper considers the possibilities of different methods of multicriteria optimization for this problem, considering specificity of the objects the method of weighted factors was chosen. Common criteria for evaluating information resources on the Internet were proposed.

Keywords: multiobjective optimization, scientometric indices, information resources.