



УДК 004.91

DOI: <https://doi.org/10.17721/AIT.2021.1.05>

Г. М. Гнатієнко, orcid.org/0000-0002-0465-5018,
Київський національний університет імені Тараса Шевченка, Київ, Україна,
О. В. Олецький, orcid.org/0000-0002-0553-5915,
Національний університет “Києво-Могилянська Академія”, Київ, Україна

Порівняння методик ранжування науковців на основі індексів Гірша та PageRank: суперечливі ситуації

Описано результати експериментального порівняння методик оцінювання та ранжування науковців і дослідників на основі їхньої публікаційної активності. Порівнюються наукометричні індикатори на основі індексів Гірша та PageRank-подібних алгоритмів; використовується матриця посилань між статтями, написаними різними авторами. Наведено приклад ситуації, коли застосування різних методик приводить до протилежних результатів – наприклад, коли автори статей, які отримали найкращі оцінки за PageRank, мають найнижчі (тобто найгірші) індекси Гірша. Описана ситуація характеризується тим, що деякий автор отримав ключовий результат, на який посилаються всі інші автори, але самих статей у нього мало, а також із можливими посиланнями в роботі такого автора, статті якого мають високий рівень цитування. Робиться спроба зіставлення результатів із можливими результатами експертного оцінювання, яке має особливе значення за побудови автоматизованих систем, що базуються на комбінуванні різних методик оцінювання. Розглянуто можливе застосування методу аналізу ієрархій. Для побудови матриць попарних порівнянь використано транзитивні шкали з параметром, який задає відношення переваги між градаціями у параметризованому вигляді.

Ключові слова: наукометрія, індекс Гірша, PageRank, метод аналізу ієрархій.

Для цитування (for citation): Г. М. Гнатієнко, О. В. Олецький. “Порівняння методик ранжування науковців на основі індексів Гірша та PageRank: суперечливі ситуації,” *Сучасні інформаційні технології*, vol. 1, p. 39–44, 2021.

ВСТУП

Проблеми оцінювання результативності наукової роботи протягом багатьох століть привертати увагу вчених усього світу. Нині особливо інтенсивно розвиваються методи наукометрії, які використовують, зокрема, для оцінювання та ранжування науковців і дослідників, наукових та освітніх інституцій, наукових журналів відповідно до їхніх здобутків, зокрема і на основі наявних статей. У цьому разі за основу беруть такі показники: кількість статей, кількість посилань на них тощо. На підставі цього для кожного дослідника можна розрахувати його індекс Гірша, або h-індекс [1], який належить до так званих author-level metrics [2].

Індекс Гірша прийнято розглядати як один з основних наукометричних показників для порівняння науковців між собою. Він має позитивні риси, але, як і інші наукометричні показники, містить і суттєві недоліки. Існує ряд його модифікацій, а також пропозицій щодо альтернативних показників. Слід зазначити, що в дослідженні різних методик ранжування та порівняння на

основі наукометричних показників домінують евристичні підходи, у яких суттєву роль при зіставленні та ранжуванні досягнень науковців відіграють суб'єктивні чинники – введені для аналізу евристики. Кожний показник має свої переваги та недоліки, і не можна однозначно сказати, який із цих показників приводить до кращих результатів.

Цікавими є ситуації, коли застосування різних методик приводить до протилежних результатів. У цій роботі наведено один із таких прикладів, отриманий у ході експериментального порівняння методик ранжування на основі індексу Гірша та на основі алгоритму PageRank [3, 4]. Здійснено також зіставлення цих результатів із можливими результатами експертного оцінювання на основі побудови матриць попарних порівнянь.

Метою роботи є проведення дослідження ранжування результативності науковців за їхніми науковими працями та посиланнями на них на основі застосування різних методик оцінювання та зіставлення одержаних результатів.

© Гнатієнко Г. М., Олецький О. В., 2021



ПОСТАНОВКА ЗАДАЧІ

Нехай задано такі величини:

A – множина авторів;

S – множина статей;

$G(a)$ – множина статей, написаних автором a ;

$Q(a) = |G(a)|$ – кількість статей, написаних автором;

$K(s)$ – кількість посилань на статтю s ;

$C(a)$ – кількість посилань на всі статті автора a ;

$H(a)$ – індекс Гірша автора a .

Необхідно дослідити публікаційну активність авторів та їхні показники цитованості, застосовуючи різні підходи. Зокрема, для здійснення цього є сенс застосувати h -індекс (індекс Гірша) й алгоритми PageRank.

ОГЛЯД ДОСЛІДЖЕНЬ ЗА ТЕМАТИКОЮ

У наукових дослідженнях за ступенем складності прийнято розрізняти структуровані, слабкоструктуровані та неструктуровані проблеми. Для підвищення рівня структурованості проблеми застосовують різні підходи та прийоми. Зокрема, практика упорядкування об'єктів, які досліджуються, сприяє структуризації проблеми. Відтак це розкриває можливості для поглибленого вивчення та застосування досконаліших та обгрунтованіших математичних методів.

Однак у задачах, де присутня суб'єктивна складова, завжди існують ситуації, які можуть привести до протилежних результатів. У цій статті розглянуто ситуацію, яка демонструє можливості одержання діаметрально протилежних результатів у разі застосування різних евристик. Прийняття рішень на основі застосування такого виду аналізу має бути ретельно зваженим і всебічно дослідженим.

Тому підвищення рівня структурованості проблеми полягає не тільки в упорядкуванні альтернативи, але й у дослідженні та порівнянні евристик, які застосовують для досягнення цієї мети. Зокрема, перспективним також є використання комплексних підходів з урахуванням зіставлення евристик і врахування їхніх особливостей.

Проблема, яка розглядається у цій статті, є міждисциплінарною. Для її розв'язання слід залучити інструментарій кількох наукових напрямів:

- дослідження наукового простору і, зокрема, наукометричні дослідження;

- експертні технології у цілому і задачі ранжування альтернатив і визначення їхньої ваги, зокрема;

- застосування методів багатокритеріальної оптимізації тощо.

Науковим простором є множина структурно упорядкованих об'єктів, яку можна представити

моделлю, що відображує структуру групування й упорядкування елементів цього простору [5, 6]. Епоха цифровізації, в якій ми живемо, вимагає від наукової спільноти сприйняття ідеї наукометричності та широкого застосування її в науковій діяльності [7, 8].

Досить поширеною думкою є ідея про те, що мірою корисності публікації є її цитованість [9, 10]. На основі цього постулату виокремлюють групу показників, які використовують кількість публікацій науковця та кількість цитувань цих публікацій. Серед наукометричних показників особливе місце належить індексу Гірша. Наприклад, у роботі [11] виділено 12 «гіршоподібних» наукометричних показників.

Слід зазначити, що вивчаючи інформацію, яка стосується наукових досліджень, та аналізуючи її, нині найчастіше застосовують методи математичної статистики. Для підвищення рівня досліджень потрібно розробити відповідні математичні моделі та застосовувати у цій галузі методи теорії прийняття рішень, методи розв'язання задач багатокритеріальної оптимізації, експертні технології, «м'які» обчислення тощо.

МЕТОДИ Й ІНСТРУМЕНТИ ДОСЛІДЖЕННЯ

Зазначимо, що

$$C(a) = \sum_{s \in G(a)} K(s)$$

(принаймні в найпростішому випадку). Можна розглядати не просто кількість посилань, а кількість статей, з яких є посилання на статті автора, але це окреме питання.

Автор a має індекс Гірша r (тобто $H(a) = r$), якщо він має r публікацій, на кожен з яких є не менше ніж r посилань. Причому вважають, що наявність посилань свідчить про достатньо високу якість статті та її важливість для відповідного наукового напрямку. Значну роль відіграє і авторитет автора статті.

Неявно враховується і кількість статей, оскільки очевидно, що

$$\forall a: H(a) \leq Q(a). \quad (1)$$

Проте тут є принципове обмеження, яке виникає у разі досягнення рівності, у яку може перетворитися нерівність (1):

$$H(a) \leq Q(a). \quad (2)$$

Для ситуації, яка виражена рівністю (2), подальше збільшення кількості посилань, хоч воно само по собі є позитивним фактором, на зростання індексу Гірша вже не впливає. Якість відповідних статей, так само як і якість посилань на них, враховується недостатньо.



Алгоритм PageRank [3, 4] був успішно застосований у Google для ранжування вебсторінок. Але його можна використати для ранжування вузлів довольного графа на основі аналізу зв'язків між ними.

Базовий варіант алгоритму можна охарактеризувати так.

Нехай є n вузлів (вершин графа); граф вважаємо орієнтованим. Нехай граф задано матрицею суміжностей M , елементи якої отримують за таким принципом:

$$M_{ij} = \begin{cases} 1, & \text{якщо } i\text{-й вузол посилається на } j\text{-й,} \\ 0 & \text{– в іншому випадку,} \end{cases} \quad (3)$$

$i, j = 1, \dots, n.$

На основі матриці M з елементами вигляду (3) будується матриця перехідних імовірностей P (тут ми дотримуємося стандартної імовірнісної термінології, пов'язаної з випадковим блуканням і відповідним марковським ланцюгом, хоча з тим же успіхом можна розглядати і неймовірнісні інтерпретації):

$$P_{ij} = (1 - \alpha) / n + \alpha \times r_{ij},$$

$$r_{ij} = M_{ij} / s_i, \quad i, j = 1, \dots, n,$$

де s_i – кількість вихідних дуг з i -ї вершини;

α – деяке задане значення;

$0 < \alpha < 1$ – параметр алгоритму; зокрема при ранжуванні веб-сторінок часто приймається $\alpha = 0,85$.

Можна сказати, що сформульований таким чином класичний алгоритм PageRank враховує не тільки кількість посилань на вузол, але й вагу цих посилань: посилання від вузлів із більш високим рангом мають більшу вагу.

У нашій задачі вузлами графа виступають статті (це відмінність, зокрема, від роботи [2], в якій вузли відповідають авторам). Відповідно,

$M_{ij} = 1$, якщо стаття s_i посилається на статтю s_j .

Тоді, очевидно, що

$$K(s_j) = \sum_{i=1}^n M_{ij}. \quad (4)$$

Описана методика призначена для ранжування статей на основі значень (4). Проте зрозуміло, що її з очевидними модифікаціями можна використати і для оцінювання та ранжування авторів цих статей. Так, оцінка автора $V(a)$ може бути отримана на основі оцінок його статей $v(s_1), v(s_2), \dots$ наприклад, за однією з формул:

$$V(a) = \max_{s \in G(a)} v(s) \quad (5)$$

(оцінка найкращої статті автора) або

$$V(a) = \sum_{s \in G(a)} v(s) / |G(a)|. \quad (6)$$

(середня оцінка за всіма статтями автора).

Крім формул (5) та (6) можна використати і комбіновану оцінку автора на основі обчислення зваженої суми

$$V(a) = \sum_{s \in G(a)} \rho(s) v(s), \quad (7)$$

тобто, коли різним статтям авторів будуть приписуватися різні ваги $\rho(s)$ залежно від їхніх рангів. Утім, приклад, наведений у статті, носитиме досить базовий характер, і для цих цілей конкретна формула вигляду (5), (6) чи (7) не матиме особливого значення.

ПОРІВНЯННЯ ПІДХОДІВ ТА ЕКСПЕРТНЕ ОЦІНЮВАННЯ

Розглянемо конкретний приклад, який ілюструє описані вище підходи. Цей приклад показує ситуацію суперечності, коли автор статті, найкращої за критерієм PageRank, отримує найменший індекс Гірша.

Нехай перший автор є автором ключової, тобто найважливішої статті, результати якої використовують інші автори, і відповідно на яку є посилання з усіх інших статей. Проте він є автором лише однієї статті, і тому його індекс Гірша не може бути більшим за 1. Статті ж інших авторів мають посилання одна на іншу.

У цьому прикладі є 4 автори, 9 статей. Статті розподілено між авторами таким чином (тут нумерація починається з 0):

$$G(0) = \{0\}, G(1) = \{1,8\}, G(2) = \{2,3,4\},$$

$$G(3) = \{5,6,7\}.$$

Отже,

$$Q(0) = 1, Q(1) = 2, Q(2) = Q(3) = 3.$$

Граф, який задає структуру посилань з однієї статті на іншу, задано такою матрицею суміжності:

$$M = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 1 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 & 0 & 1 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 & 1 & 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 & 1 & 1 & 0 & 1 & 1 \\ 1 & 0 & 1 & 1 & 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

Для параметра α візьмемо одне з типових значень: $\alpha = 0,85$.

Міри важливості вузлів (статей), що розраховані за алгоритмом PageRank, мають вигляд:



0,3431 0,3207 0,0509 0,0509 0,0509
0,0518 0,0518 0,0509 0,0292

Таким чином, 0-ва стаття (вона ж єдина стаття 0-го автора) отримує перевагу за PageRank, як і слід було чекати. На цьому ж прикладі можна подивитися підвищення ваги наступної статті за рахунок того, що на неї посилається найкраща. Хоча слід зазначити, що в цьому випадку таке підвищення може не розглядатися як особливо заслужене (не виключена ситуація, що автор цієї статті міг якось домовитися з автором найкращої статті).

На цій основі можна розрахувати оцінки авторів. Так, якщо оцінку автора обчислювати як максимум оцінок окремих статей, то маємо:

$$V(0) = 0,3431; V(1) = 0,3207; V(2) = 0,0059; \\ V(3) = 0,0518.$$

Проте значення індексів Гірша авторів такі:

$$H(0) = 1, H(1) = 2, H(2) = 3, H(3) = 3.$$

Таким чином, експеримент показує, що автори двох статей, які отримали найвищі (тобто найкращі) оцінки за PageRank, мають найнижчі (тобто найгірші) індекси Гірша.

У подібних ситуаціях часто доцільно залучати до розгляду методики експертного оцінювання [12, 18 та ін.]. Зіставлення результатів роботи різних алгоритмів може виявитися корисним у ситуаціях, коли доводиться робити висновки, які методики автоматизованого алгоритмічного оцінювання заслуговують на більшу довіру або, можливо, навпаки.

Особливу увагу в цьому контексті слід звернути на метод попарних порівнянь, який часто використовують у межах методу аналізу ієрархій Сааті [13, 14 та ін.].

У прикладі, що розглядається, видаються обґрунтованими такі міркування експертів. Імовірно, що вони надали б перевагу першому (0-му) автору, оскільки на його роботу спираються всі інші автори. Наступні два автори могли б розглядатися як рівнозначні. Останній мав би отримати найгіршу оцінку – оскільки на його роботи немає жодних посилань, крім самоцитування та сумнівного посилання від 0-го автора.

Матриця попарних порівнянь у такій ситуації могла б мати такий вигляд:

$$\begin{pmatrix} 1 & 2 & 2 & 3 \\ 1/2 & 1 & 1 & 2 \\ 1/2 & 1 & 1 & 2 \\ 1/3 & 1/2 & 1/2 & 1 \end{pmatrix}.$$

Для елементів цієї матриці, відповідно до стандартного підходу, значення 2 означає мінімальну перевагу одного варіанта над іншим, 3 – дещо більш виражену перевагу). Деякі автори звертали увагу на те, що застосування саме такої шкали градацій не завжди приводить до добрих результатів, зокрема через те, що розкид між максимальною та мінімальною оцінкою може

вийти надто великим, і в таких випадках може виявитися доцільним застосування так званих транзитивних шкал [14–16 та ін.]. Для транзитивних шкал вводиться деякий параметр τ , який визначає, наскільки наступний рівень градацій порівняння є більшим за попередній. Наприклад, якщо рівнозначність альтернатив оцінюється значенням 1, а мінімальний рівень переваги оцінюється як τ , то наступна градація переваг буде оцінена як τ^2 , і т. д. Але це питання заслуговує на окремий розгляд.

Зазначимо, що для оцінювання переваг між варіантами може бути застосована шкала Томаса Сааті з градаціями (1, 3, 5, 7, 9) або шкала оцінювання задач, яка використовується у методології Agile з градаціями (1, 2, 3, 5, 8, 13, 21). Проте дослідження застосування різних типів шкал не є предметом цієї статті і може бути здійснене у подальшому в окремій роботі.

Для наведеної матриці попарних порівнянь її нормалізований головний власний вектор, кожна компонента якого є оцінкою відповідного автора, дорівнює

$$0,4236 \quad 0,2270 \quad 0,2270 \quad 0,1223.$$

Таким чином, 0-й автор перемагає, із цього погляду оцінки на основі PageRank виявляються ближчими до результатів експертного оцінювання порівняно з оцінками на основі аналізу індексів Гірша.

Проілюструємо також можливе використання параметризованих транзитивних шкал. Параметризована матриця попарних порівнянь для нашого прикладу могла б мати вигляд

$$\begin{pmatrix} 1 & \tau & \tau & \tau^2 \\ \frac{1}{\tau} & 1 & 1 & \tau \\ \frac{1}{\tau} & 1 & 1 & \tau \\ \frac{1}{\tau^2} & \frac{1}{\tau} & \frac{1}{\tau} & 1 \end{pmatrix}.$$

При $\tau = 1.1$ вектор оцінок складатиметься з елементів, які мають такі значення:

$$0,2744 \quad 0,2494 \quad 0,2494 \quad 0,2268.$$



Таким чином, зі зменшенням τ розкид між оцінками суттєво зменшився. Вибір параметра τ може мати суттєве значення для застосування алгоритмів оцінювання. Деякою мірою питання про вибір τ обговорювалося в [16] для задачі автоматизованого оцінювання студентських учбових проєктів. Проте, узагалі кажучи, питання застосування транзитивних шкал і вибору параметра τ , зокрема і для аналізу наукометричних даних, заслуговує на окремих розгляд.

ПЕРСПЕКТИВИ ПОДАЛЬШИХ ДОСЛІДЖЕНЬ

Для детальнішого зіставлення евристик, які застосовуються у наукометричних дослідженнях, слід додатково дослідити різні підходи та зіставити результати не лише у двох описаних напрямках, але порівняти ширший набір інструментів, які застосовують дослідники.

Слід також запропонувати критерії вибору інструментальних засобів для наукометричних досліджень, виявити позитивні та негативні риси використовуваних евристик.

ВИСНОВКИ

Застосування єдиного критерію у наукометричних дослідженнях є зручним інструментом, але часто – нерепрезентативним та одностороннім. Цей інструмент у багатьох практичних ситуаціях не відображає належною мірою переможців серед науковців, і лише висвітлює одну грань наукових досягнень.

Наведений приклад експериментально підтверджує, що ранжування авторів на основі аналізу індексів Гірша може суттєво відрізнитися від ранжувань, отриманих на основі оцінок PageRank для статей, і на основі цього – відрізнитися для авторів відповідних статей. Справді, експеримент для наведеного прикладу показує, що автори статей, найкращих за PageRank, мають найгірші індекси Гірша.

Проте слід зазначити, що ранжування як на основі індексів Гірша, так і на основі PageRank, має свої переваги та недоліки – як, власне, й інші методики. Тому може виявитися доцільним комбінування різних методик – зокрема, шляхом аналізу того, яким чином методики автоматизованого алгоритмічного оцінювання забезпечують дотримання певних принципів і політик оцінювання. Можна, наприклад, говорити про надання окремих методикам певних коефіцієнтів надійності, і про налаштування цих коефіцієнтів на основі застосування тих чи інших процедур машинного навчання, зокрема і навчання з підкріпленням [17, 19].

Можна зробити однозначний висновок, що дослідження результативності науковців має бути різностороннім, а застосування інструментів – комплексним. Для порівняння й адекватного комбінування різних методик може виявитися дуже корисним застосування методів експертного оцінювання; у статті наводиться відповідний приклад. У цьому контексті варто також звернути увагу на роботу [20], в якій ідеться про аналіз різних методик оцінювання дослідників, зокрема на основі використання наукометричних показників.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

- [1] J.E. Hirsch, “An index to quantify an individual's scientific research output”, *PNAS*, vol.102 (46), pp.16569–16572, 2005.
- [2] J.D. West, M.C. Jensen, R.C. Dandrea, G.J. Gordon, C.T. Bergstrom. “Author-level Eigenfactor metrics: Evaluating the influence of authors, institutions, and countries within the social science research network community”, *Journal of the American Society for Information Science and Technology*, vol.64(4), pp.787–801, 2013.
- [3] S. Brin, L. Page. “The anatomy of a large-scale hypertextual web search engine”, in *Proc. WWW*, 1998, pp.107–117.
- [4] К.Д. Маннинг, П. Рагхаван, Х. Шютце, *Введение в информационный поиск*, Москва: ООО «И.Д. Вильямс», 2011.
- [5] В.Г. Кремень, В.Ю. Биков, “Категорії «простір» і «середовище»: особливості модельного подання та освітнього застосування”, *Теорія і практика управління соціальними системами*, no.2, pp.3–16, 2013.
- [6] О.Ю. Кучанський, “Складові аналізу наукових мереж”, *Управління розвитком складних систем*, no.41, pp.115–126, 2020.
- [7] А.В. Кільченко, “Вітчизняний та зарубіжний досвід використанні інформаційно-цифрових технологій для оцінювання результативності науково-педагогічних досліджень”, in *Звітна науково-практична конференція Інституту інформаційних технологій і засобів навчання НАПН України : матеріали науково-практичної конференції*, 2021, pp.48–54.
- [8] В.Ю. Биков, О.М. Спірін, А.О. Білощицький та ін., “Відкриті цифрові системи в оцінюванні результатів науково-педагогічних досліджень”, *Інформаційні технології і засоби навчання*, vol.75(1), pp.294–315, 2020.
- [9] В.М. Горючий, “Критерії якості наукових досліджень у контексті забезпечення національних інтересів”, *Вісник Національної академії наук України*, no.6, pp.74–80, 2015.
- [10] О.М. Спірін, С.М. Іванова, А.В. Кільченко, Т.Д. Новицька, “Використання наукометричних баз даних і систем вебаналітики для моніторингу електронних наукових фахових видань”, *Інформаційні технології в освіті*, vol.45(4), pp.18–30, 2020.
- [11] С.Д. Штовба, Е.В. Штовба, “Обзор наукометрических показателей для оценки публикационной деятельности ученого”, *Управление большими системами. Спец. вып. 44: Наукометрия и экспертиза в управлении наукой*, pp.262–278, 2013.
- [12] В.Є. Снитюк, Г.М. Гнатієнко, *Експертні технології прийняття рішень*, Київ: Маклаут, 2008.



[13] Т. Саати, *Принятие решений. Метод анализа иерархий*, Москва: Радио и связь, 1993.

[14] И. Г. Черноруцкий. *Методы принятия решений*, Санкт-Петербург: БХВ-Петербург, 2005.

[15] O. V. Oletsky, E.V. Ivohin, "Formalizing the Procedure for the Formation of a Dynamic Equilibrium of Alternatives in a Multi-Agent Environment in Decision-Making by Majority of Votes", *Cybern Syst Anal*, vol.57, pp.47-56, 2021.

[16] А. В. Олецкий, М.Ф. Махно, "О повышении уровня адекватности в результатах процесса оценивания учебных проектов на основе параметрической релаксации метода парных сравнений", *Проблемы управления и информатики*, no.1, pp.122–133, 2021.

[17] С. Рассел., П.Норвиг, *Искусственный интеллект: современный подход*, Москва: Изд. дом «Вильямс», 2006.

[18] M. Antonewich, A. Didyk, and V. Snytyuk, "Optimization of Functions of Two Variables by Deformed Stars Method," in *Proc. ATIT*, 2019, pp. 475–480.

[19] С. М. Николенко, А.Л. Тулупьев. *Самообучающиеся системы*, Москва: МЦНМО, 2009.

[20] Н. Hnatiienko, V. Snytyuk, N. Tmienova, O. Voloshyn, "Determining the effectiveness of scientific research of universities staff" in Selected Papers of the 7th International conference "Information Technology and Interactions" (IT&I-2020). *CEUR Workshop Proceedings*, 2883, 2021, pp. 164–176.

[21] M. Shell. (2002) IEEEtran homepage on CTAN. [Online]. Available: <http://www.ctan.org/tex-archive/macros/latex/contrib/supported/IEEEtran/>

Стаття надійшла до редколегії

03.10.2021

Comparing methods of ranking scientists on the base of Hirsch index and PageRank: contradictory situations

Experiments aimed at comparing different methods of estimating and ranking scientists and researchers on the base of their publication activity are reported. Scientometric indicators based on h-index and PageRank are being compared. For such a comparison, a graph of citations represented by a matrix was applied. An example when different methods lead to opposite results was described. For example, authors having the best PageRank-based estimations may have the least h-indices. Such a situation is possible when a high-cited author managed to obtain a key result cited by all the other authors but this author has few papers. A comparison with methods of expert estimations was carried out, which appears to be very useful for building automated systems combining various methods of algorithmic estimating and ranking. The Analytic Hierarchy Process was applied. For building pairwise comparison matrices, transitive scales with a parameter representing how much times the next level of advantage is bigger than the previous one were harnessed.

Keywords: scientometrics, h-index, PageRank, Analytic Hierarchy Process.



Григорій Гнатієнко. Закінчив факультет кібернетики Київського національного університету імені Тараса Шевченка, канд. техн. наук. Місце роботи: факультет інформаційних технологій Київського національного університету імені Тараса Шевченка, Київ, Україна. Посада: заступник декана з наукової роботи. Наукові інтереси: оброблення експертної інформації, прийняття рішень, інформаційні технології.

Hryhorii Hnatiienko. PhD. Education: Faculty of Cybernetics, Taras Shevchenko National University of Kyiv. Affiliation: Faculty of Information Technologies of Taras Shevchenko National University of Kyiv. Kyiv, Ukraine. Position: Deputy Dean for Research. Research interests: processing of expert information, decision making, information technologies.



Олексій Олецкий. Закінчив факультет кібернетики Київського національного університету імені Тараса Шевченка, канд. техн. наук. Місце роботи: факультет інформатики, Національний університет «Києво-Могилянська академія», Київ, Україна. Посада: доцент кафедри мультимедійних систем. Наукові інтереси: прийняття рішень, штучний інтелект.

Oleksiy Olets'kyi. PhD, Associate Professor. Education: Faculty of Cybernetics, Taras Shevchenko National University of Kyiv. Affiliation: Faculty of Informatics, National University of Kyiv-Mohyla Academy. Kyiv, Ukraine. Position: Associate Professor. Research interests – decision making, artificial intelligence.