*Пропонується метод знаходження скалярних оцінок, а також метод побудови інтегральних оцінок результатів науково-дослідної діяльності науковців. Останній базується на основі розрахунку метричних відстаней між точками багатовимірного простору, координати яких складаються зі скалярних оцінок. Ці методи можуть бути використані при здійсненні комплексного оцінювання науковців, вищих навчальних закладів*

*Ключові слова: індекс цитування, оцінка науково-дослідної діяльності, бібліометричні показники, інтегральна оцінка*

*Предлагается метод нахождения скалярных оценок, а также метод построения интегральных оценок результатов научно-исследовательской деятельности научных работников. Последний основан на расчете метрических расстояний между точками многомерного пространства, координаты которых составлены из скалярных оценок. Эти методы могут быть использованы при осуществлении комплексного оценивания научных работников, высших учебных заведений*

*Ключевые слова: индекс цитирования, оценка научно-исследовательской деятельности, библиометрические показатели, интегральная оценка*

UDC 005.8
DOI: 10.15587/1729-4061.2017.103651

# EVALUATION METHODS OF THE RESULTS OF SCIENTIFIC RESEARCH ACTIVITY OF SCIENTISTS BASED ON THE ANALYSIS OF PUBLICATION CITATIONS

**A. Biloshchytskyi**
Doctor of Technical Sciences, Professor
Department of Technology Management
Taras Shevchenko National University of Kyiv
Volodymyrska str., 60, Kyiv, Ukraine, 01033
E-mail: bao1978@gmail.com

**A. Kuchansky**
PhD, Associate Professor
Department of Cybersecurity and computer engineering*
E-mail: kuczanski@gmail.com

**Yu. Andrashko**
Lecturer**
E-mail: yurii.andrashko@uzhnu.edu.ua

**S. Biloshchytska**
PhD, Associate Professor
Department of Information Technology Designing and Applied Mathematics*
E-mail: bsvetlana2007@ukr.net

**O. Kuzka**
PhD, Associate Professor**
E-mail: oleksandr.kuzka@uzhnu.edu.ua

**O. Terentyev**
Doctor of Technical Sciences, Associate Professor
State «Research Institute of building production»
Lobanovskoho ave., 51, Kyiv, Ukraine, 03037
E-mail: terentyev79@ukr.net

*Kyiv National University of Construction and Architecture
Povitroflotsky ave., 31, Kyiv, Ukraine, 03037
**Department of System Analysis and Optimization Theory
Uzhhorod National University
Narodna sq., 3, Uzhhorod, Ukraine, 88000

## 1. Introduction

Dynamic development of scientific environment of any country is an extremely important factor that enhances its prestige, economic development, development of new technologies in various fields of human activity. An important task that has been tackled by researchers over the past decades is to create mechanisms to effectively manage the development of scientific environment. This can be done by engaging private organizations, financial support from the state authorities of different levels, expansion of international cooperation within the frameworks of certain scientific and educational projects (Horizon 2020, Erasmus+). That is why an important task for private companies interested in the development of science-intensive technologies, as well as for foreign partners, is creating effective criteria for the evaluation of results of the research activities of scientists, higher educational institutions, as well as structural subdivisions of these educational organizations.

Evaluation of the results of research activities provides a possibility to validate the process of scientific research against the goals that were set at the stage of planning and, if necessary, to adjust the course of these studies. One of the components of the evaluation of scientific-research work is the assessment of the main outcomes of this work: scientific publications. A criterion of significance of publications may include using the results of these articles in other scientific research. That is why the evaluation of results of the research work can be conducted by finding various bibliographic indicators of publication citations.

Evaluation of scientific-research work in general can be based on the personal assessments of scientists who are engaged in such studies. Generally accepted criteria of evaluation of the results of scientific-research work of scientists are the indicators of publication citations that were published by these scientists. These indicators are typically scalar magnitudes. An approach to building such magnitudes has a number of advantages, but there are drawbacks as well. Among such shortcomings is the loss of part of the input data and the existence of such boundary cases when a parameter does not change its value when the number of citations and publications does increase. Such situation arises for many known indexes that calculate citations: the Hirsch index, the index I-10, the g-index. Here is an example of such situation. Assume that a scientist published n′ papers, which later became fundamental in a certain direction of research, and then he finished his career. These publications are widely used in other studies and each of them was quoted $d_i$ times. If $d_i > n'$, then traditional bibliometric indexes will equal n′, that is, the outcome of research work of this particular scholar is not very successful and important, though this assessment is not adequate. Therefore, new methods must avoid such cases when scientific-research work is carried out, new papers are published and subsequently cited, but the evaluation of results of scientific-research work does not change.

That is why it is a relevant task to develop new, or to modify existing, methods of evaluation of the results of scientific-research activities of scientists, which would be devoid of the indicated shortcomings.

## 2. Literature review and problem statement

The task to evaluate results of the scientific-research work is a traditional subtask in managing educational and scientific-research institutions. Article [1] proposes a project-vector approach to the management of higher educational institutions. The functioning of higher educational institutions is regarded as a set of projects, each of which is represented by a vector in the space of scientific, educational and administrative activities. To determine the coordinates of these vectors, it is required to perform evaluation of results of the appropriate activity. Paper [1] substantiates the need to construct a universal appraisal of results of scientific-research activity in the management of higher educational institutions; however, no recommendations on how to achieve it were described.

Article [2] describes an approach to the evaluation of functioning of an educational establishment by building a parametric model. Such model requires tracking of changes in the results of various activities, including scientific, as well as prediction for the next period. Paper [3] describes an approach to the management of an educational institution based on predicting the state transition, which are represented by a Markov chain. Determining the probabilities of transitions needs a comprehensive assessment of the outcomes of the scientific-research and teaching activity of an institution and the prediction of its values over the next period. Article [4] conducted an analysis of the features and synthesized the structure of interaction between main subjects as knowledge carriers. Paper [5] defined conditions of effective work for the development of a project that can be used for the purposes of management and evaluation of activities of groups of researchers.

Article [6] examined indicators that are required for a comprehensive assessment of activities in higher education institutions, and reported the analysis of these indicators. Paper [7] gives an overview of the features of the methods for evaluating the results of performance, in particular scientific, of higher educational establishments in Ukraine. However, these publications failed to construct a universal method of evaluation.

The results of scientific-research activity of scientists can be assessed based on the indicators of citing the articles that were published by the given scientists. Paper [8] proposed an overview of scientometric databases and the ways of obtaining the main indicators of citation. The most common bibliometric indicator at present is the Hirsch index. The principle that underlies it is described in article [9]. Hirsch index is calculated as follows: a scientist receives the index h in the case that the scientist published at least h articles, each of which is cited at least h times. Paper [10] proposes applying the so-called g-index. This index represents the largest number g, which corresponds to the number of articles that were quoted in total not less than $g^2$ times. Author of article [10] draws a conclusion about the insufficiency of a simple scalar evaluation and about the need to use a complex assessment that consists of the e- and h-indices, but he does not generalize to a larger number of indicators. In addition, the method considered in paper [10], does not lead to a complete solution to the problem on the loss of information about citation.

Article [11] points out the principal shortcomings of the h- and g-indices, which include the loss of information about citation of the most popular publications of an author, and proposes employing the e-index to overcome these deficiencies. Paper [12] proposes several modifications to calculating the h-index, including taking into account the self-citation. Article [13] examined a correlation between the Hirsch index and the g-index, taking into account different samples of scientists and scientific collections that published results of scientific-research activity. In [14], authors describe the use of a generalized integral to calculate certain bibliometric indices and propose methods of establishing a functional dependence between the number of publications and the number of citations.

According to the authors, the task to evaluate the results of information search has a number of common characteristics with the problem on evaluating results of scientific-research activity of scientists. These shared characteristics imply that the pages on the Internet are connected via hyperlinks, while scientific publications are linked through citation. To solve the problem on evaluation of results of information search in the Internet, the so-called PR index is applied, which is described in article [15]. Paper [16] described applying the Monte Carlo method to speed up finding the PR index. Study [17] considered a modification to the Monte-Carlo method to find the PR index in a dynamic network whose structure is constantly changing. The results of this study open up the possibility of using methods that are based on the same principles for processing large amounts of data on the citation of scientific publications.

When constructing methods, the present study employs the approach of finding similarities among multidimensional vectors. Article [18] described formulas for finding the metric distances by different techniques, as well as the method of the indexation of vectors based on k-nearest neighbors.

## 3. The aim and objectives of the study

The goal of present study is to construct methods for the evaluation of results of scientific-research activity of scientists based on the analysis of citation of the papers published by these scientists.

To accomplish the goal, the following tasks have been set:

– construction of the method of calculating a scalar evaluation of the results of scientific-research activity of scientists;

– construction of the method that makes it possible to calculate integral evaluations of the results of scientific-research activity of scientists based on scalar evaluations of publication citations.

## 4. Methods of constructing a scalar estimation of the results of scientific-research activity of scientists

Assume that $A=\{a_1, a_2,..., a_n\}$ is a set of scientists, n is the number of scientists, and $P=\{p_1, p_2,..., p_m\}$ is the set of articles published by the given scientists, m is the number of publications. Set $U \subset A \times P$ is a set that assigns a relation of authorship between the scientists and the articles published by the given scientists. Set $C \subset P \times P$ assigns publication citations. We shall introduce the notion of evaluation of the results of scientific-research activity.

A scalar evaluation of the results of scientific-research activity of a scientist is a certain functional representation Q:

$$Q: A \to R, \tag{1}$$

where R is the set of real numbers.

A vector evaluation of the results of scientific-research activity of scientists is a certain functional representation $Q_v$:

$$Q_v: A \to R^v, \tag{2}$$

where $R^v$ is the set of v-dimensional real vectors.

*Statement of the problem on the evaluation of results of scientific-research activity of scientists.* We shall denote a set of all publications of a scientist $a_i$, $i=\overline{1,n}$ through

$$P(a_i)=\left\{p_j \in P \,\middle|\, (a_i, p_j) \in U\right\}, \quad i=\overline{1,n}, \quad j=\overline{1,m}, \tag{3}$$

where set $U \subset A \times P$ represents the authorship of scientist $a_i$ for publications $p_j$.

We shall define a set of publications cited by every scientist $a_i$, $i=\overline{1,n}$ – $\overline{C}(a_i)$, and the set of publications that cite the articles of the scientist – $C(a_i)$, in the following form:

$$\overline{C}(a_i)=\left\{p_j \in P \,\middle|\, (p_y, p_j) \in C, p_y \in P(a_i), y=\overline{1,m}\right\},$$
$$i=\overline{1,n}, \quad j=\overline{1,m}, \tag{4}$$

$$C(a_i)=\left\{p_j \in P \,\middle|\, (p_j, p_y) \in C, p_y \in P(a_i), y=\overline{1,m}\right\},$$
$$i=\overline{1,n}, \quad j=\overline{1,m}. \tag{5}$$

Similarly, for each article $p_j$, we shall consider the set of its authors

$$A(p_j)=\left\{a_i \in A \,\middle|\, (a_i, p_j) \in U\right\},$$
$$i=\overline{1,n}, \quad j=\overline{1,m}, \tag{6}$$

as well as the set of articles cited by the given article $p_j$ – $\overline{C}(p_j)$ and the set of articles that cite article $p_j$ – $C(p_j)$

$$\overline{C}(p_j)=\left\{p_j \in P \,\middle|\, (p_j, p_y) \in C, y=\overline{1,m}\right\}, \quad j=\overline{1,m}, \tag{7}$$

$$C(p_j)=\left\{p_j \in P \,\middle|\, (p_y, p_j) \in C, y=\overline{1,m}\right\}, \quad j=\overline{1,m}. \tag{8}$$

The problem on the evaluation of results of scientific-research activity of scientists is to find for each scientist $a_i$, $i=\overline{1,n}$, based on the assigned information regarding the citation of his/her publications, certain evaluation $q_i$, which can be represented in the form of functional Q:

$$q_i = Q\left(P(a_i), C(a_i)\right), \quad i=\overline{1,n}, \tag{9}$$

where $P(a_i)$ is the set of articles of scientist $a_i$, $C(a_i)$ is the set of all articles that cite the articles of scientist $a_i$, $q_i$ is the scalar evaluation of results of scientific-research activity of scientist $a_i$, maximal values of functional $Q(P(a_i),C(a_i))$ correspond to the best results of scientific-research work of appropriate scientist $a_i$, $i=\overline{1,n}$ from the point of view of achieving maximal effectiveness.

Let us consider assigning the known evaluations of scientific activity in terms of sets A, P, C, U. Without limiting the generality, we shall assume that the publications and scientists in the assigned sets are arranged in order of the non-growth in citation, that is,

$$\|C(p_1)\| \ge \|C(p_2)\| \ge \ldots \ge \|C(p_m)\|, \tag{10}$$

$$\|C(a_1)\| \ge \|C(a_2)\| \ge \ldots \ge \|C(a_n)\|, \tag{11}$$

where $\|C\|$ is the rule of the set, which is defined as the number of elements in this set.

We shall consider basic indices for calculating the citations, which can be utilized for the evaluation of results of scientific activity of researchers:

$$h(a_i)=\max_{y=\overline{1,m}} \min\left\{y, \left\|C(p_y)\right\|\right\}, \quad p_y \in P(a_i), \tag{12}$$

where $h(a_i)$ is the Hirsch h-index,

$$g(a_i)=\max_{y=\overline{1,m}} \min\left\{y, \left\lfloor \sqrt{\sum_{x=1}^{y} \left\|C(p_x)\right\|} \right\rfloor\right\}, \quad p_y \in P(a_i), \tag{13}$$

where $g(a_i)$ is the g-index,

$$e(a_i)=\sqrt{\sum_{y=1}^{h(a_i)}\left\|C(p_y)\right\|-\left(h(a_i)\right)^2}, \qquad (14)$$

where $e(a_i)$ is the e-index, $h(a_i)$ is the Hirsch h-index,

$$i_{10}(a_i)=\left\|C^{10}(a_i)\right\|, \qquad (15)$$

where $i_{10}(a_i)$ is the index I-10; set $C^{10}(a_i)$ is the set of articles cited not less than 10 times, that is,

$$C^{10}(a_i)=\left\{p_y\in P(a_i)\,\middle|\,\left\|C(P_y)\ge 10\right\|\right\}.$$

The main shortcoming is that each of the aforementioned indexes loses part of the citation information, in particular:

– h-index loses information beyond the core of Hirsch (h-core): it does not take into account information about publications cited less than h times, as well as publication citations less than h times;

– g-index loses information beyond the g-core depending on the ratio of citations to the number of publications of the author. If

$$g<\left\|P(a_i)\right\|,$$

it does not take into account information about publications cited less than g times. If

$$g=\left\|P(a_i)\right\|,$$

it does not take into account information about the citation of all publications;

– e-index loses information on the citation of publications cited less than h times;

– index I-10 loses information about publications cited less than 10 times.

Let us consider a method for the calculation of evaluation of the results of scientific-research activity of scientists, which does not lose information about any publication and any citation of an author. Similar PR method is used by the search engine Google as one of the parameters for ranking Web pages in order to arrange search results on the Internet that match the user's request [14]. We generalized the idea of calculating evaluations by the PR method and modified it to evaluate scientific-research activity of scientists. We shall term the modified method PR-q. According to it, a scalar evaluation of the results of scientific-research activity of scientist $a_i$, $i=\overline{1,n}$ is calculated by formula:

$$q_i=\sum_{z=1}^{n}\beta_{iz}\xi_z q_z,\quad i=\overline{1,n}, \qquad (16)$$

where $q_i$ is the evaluation of the scientific-research activity of scientist $a_i$; $\beta_{iz}$ is the coefficient, which is determined by the number of publication citations of scientist $a_i$ in the articles of scientist $a_z$; $\xi_z$ is the coefficient that enables the existence of a nontrivial solution to a system of linear algebraic equations (16); $q_z$ is the evaluation of scientific-research activity of scientist $a_z$.

As a result of applying formula (16), we shall construct a uniform system of linear algebraic equations in the form:

$$Bq=0, \qquad (17)$$

where B is the matrix of coefficients of the given system in the form:

$$B=\begin{pmatrix}1-\beta_{11}\xi_1 & -\beta_{12}\xi_2 & -\beta_{13}\xi_3 & \dots & -\beta_{1n}\xi_n\\ -\beta_{21}\xi_1 & 1-\beta_{22}\xi_2 & -\beta_{23}\xi_3 & \dots & -\beta_{2n}\xi_n\\ -\beta_{31}\xi_1 & -\beta_{32}\xi_2 & 1-\beta_{33}\xi_3 & \dots & -\beta_{3n}\xi_n\\ \vdots & \vdots & \vdots & \ddots & \vdots\\ -\beta_{n1}\xi_1 & -\beta_{n2}\xi_2 & -\beta_{n3}\xi_3 & \dots & 1-\beta_{nn}\xi_n\end{pmatrix}, \qquad (18)$$

q is the vector-column of unknown evaluations:

$$q=\begin{pmatrix}q_1\\ q_2\\ \vdots\\ q_n\end{pmatrix}. \qquad (19)$$

To ensure that there is non-trivial (identically not equal to zero) solution to system (16), it is required that matrix B is degenerate, that is, $|B|=0$.

*The first method for determining the coefficients of system (16).* Coefficients of system (16) can be determined by formulas:

$$\beta_{iz}=\left\|C(a_i)\cap\overline{C}(a_z)\right\|, \qquad (20)$$

$$\xi_z=\left\|\overline{C}(a_z)\right\|^{-1},\quad i=\overline{1,n},\quad z=\overline{1,n}, \qquad (21)$$

where $\beta_{iz}$ is the number of articles by scientist $a_z$, in which the given scientist cites the publications of scientist $a_i$; $\xi_z$ is the magnitude, inverse to the total number of publications by scientist $a_z$.

We shall prove the existence of a nontrivial solution to system (16). To do this, let us find the sum of the rows of matrix (18) by formula

$$1-\sum_{i=1}^{n}\beta_{iz}\xi_z=1-\sum_{i=1}^{n}\frac{\left\|C(a_i)\cap\overline{C}(a_z)\right\|}{\left\|\overline{C}(a_z)\right\|}=1-\frac{\left\|\overline{C}(a_z)\right\|}{\left\|\overline{C}(a_z)\right\|}=0,\ z=\overline{1,n}, \qquad (22)$$

that is, the rows of matrix (18) are linear dependent.

Therefore, we proved that matrix (18) is a degenerate matrix, and hence there is a non-trivial solution to (16).

*The second method to determine the coefficients of system (16).* In addition to procedure (20), (21), coefficients of system (16) can be calculated as follows:

$$\beta_{iz}=\sum_{s=1}^{m}\frac{\left\|C(a_i)\cap\{p_s\}\right\|}{\left\|A(p_s)\right\|}, \qquad (23)$$

$$\xi_z=\left(\sum_{s=1}^{n}\frac{1}{\left\|A(p_s)\right\|}\right)^{-1},\quad i=\overline{1,n},\quad z=\overline{1,n}, \qquad (24)$$

where coefficients $\xi_z$ and $\beta_{iz}$ take into account in the evaluation the number of co-authors of each article $p_s$.

If there exists a non-trivial solution to system (16), then there is a multitude of solutions, proportional to the given one. Thus, after finding evaluations using the PR-q method, it is advisable to normalize these evaluations according to one of the formulas:

$$q'(a_i)=\frac{q_i}{\sum_{k=1}^{n}q_k},\quad i=\overline{1,n}, \qquad (25)$$

$$q''(a_i)=\frac{q_i}{\max\limits_{k=\overline{1,n}} q_k},\quad i=\overline{1,n}, \tag{26}$$

where $q_i$ is the evaluation of results of scientific-research activity of scientist $a_i$, which is calculated by the PR-q method, $q'(a_i)$ is the evaluation $q_i$ normalized by the sum, $q''(a_i)$ is the evaluation $q_i$ normalized by the maximum.

It should be noted that in the case when we consider large enough number of scientists, then evaluations $q'(a_i)$ will take the values close to zero. Evaluations

$$q''(a_i)\in[0,1]$$

regardless of the number of scientists for $i=\overline{1,n}$.

We shall determine computational complexity for finding an evaluation using the PR-q method:

1. *Employing the iteration Gauss-Seidel Method to solve a system of linear algebraic equations (16).* At each iteration, we calculate approximate value of n evaluations. Each evaluation of results of scientific-research activity of scientist $a_i$, $i=\overline{1,n}$, requires computing a sum of terms whose number is equal to the number of scientists who in their publications cited $a_i$. To calculate all the coefficients of $\beta_{iz}$, $i=\overline{1,n}$, $z=\overline{1,n}$, it is required to perform a number of elementary operations that is proportional to the number of citations performed by scientist $a_z$. Therefore, the total complexity of computation has an order of O(n). Proportionality coefficient between a number of scientists and the number of elementary operations depends on the number of citations of a single article, and the number of publications of a scientist. Given the average number of citations (a scientist cites about 20 publications in one article) and publications of one scientist (25), in order to calculate one approximation of one evaluation, it is necessary to perform 25×25×20×20=25,000 elementary arithmetic operations on average. To find one approximation of evaluations of all the scientists in Ukraine, it is necessary to execute $1.75\times10^9$ arithmetic operations. The number of iterations of the method to achieve the assigned accuracy is a magnitude, which is difficult to assess, but typically, 10–15 iterations suffice. Therefore, the total complexity will have an order of $10^{10}$.

2. *Applying the Monte-Carlo method to solving system (16).* The Monte Carlo method significantly reduces computational complexity, because in order to calculate evaluation of one scientist, it is necessary to recalculate the evaluations only in the vicinity, received as a result of a random walk based on the results of citation. Evaluations of the rest of the scientists, who did not make it to the vicinity, will be considered constant and will remain without recalculation. To calculate one approximation of one evaluation, it is necessary to perform 1000 elementary arithmetic operations on average. Article [17] reported efficiency of calculating PR when executing only 2 walks, that is, the vicinity has capacity proportional to the square of the number of citations of one publication. Given the average number of citations, to find one evaluation, it is possible to consider a vicinity of $20^2$ authors. That is, calculation of one evaluation will require $10^6$ elementary arithmetic operations. It should be noted that if data are represented in the form of incidence lists, then when calculating an evaluation using the PR-q method, the number of memory access operations and the number of comparisons is proportional to the number of arithmetic operations.

## 5. Method of constructing a vector integral evaluation of the results of scientific-research activity of scientists

Assume that w scalar evaluations are assigned, which determine the results of scientific-research activity of some scientist $a_i$, $i=\overline{1,n}$. We shall denote these evaluations through $f_i^1,f_i^2,\ldots,f_i^w$. Based on these evaluations, it is possible to construct a w-dimensional vector

$$F_i=\left(f_i^1,f_i^2,\ldots,f_i^w\right),\quad i=\overline{1,n}, \tag{27}$$

where $f_i^b$ is the scalar actual evaluation of the results of scientific-research activity of scientist $a_i$, $i=\overline{1,n}$, $F_i\in R^w$.

The values $f_i^1,f_i^2,\ldots,f_i^w$ can represent an estimation of the Hirsch h-index (12), g-index (13), e-index (14), index I-10 (15) or the index, calculated by the PR-q method (16), (26). For example, vector of evaluation of the results of scientific-research activity of scientist $a_i$, which takes into account the specified evaluations, will take the form

$$F_i=\left(h(a_i),g(a_i),e(a_i),i_{10}(a_i),q''(a_i)\right),\quad i=\overline{1,n}, \tag{28}$$

that is, the measurability of the assigned vector is determined by the number of indices that are taken into account during assessment.

Without loss of generality, let us consider some point in a w-dimensional space

$$F^*=\left(f^{1*},f^{2*},\ldots,f^{w*}\right), \tag{29}$$

where $F^*\in R^w$, R is the set of real numbers; $f^{b*}$are the evaluations of results of -research activity of scientists that are the best in terms of achieving maximal efficiency or effectiveness in line with an appropriate index, $b=\overline{1,w}$, that is, they satisfy condition

$$f^{b*}\ge\max_{i=\overline{1,n}}\left(f_i^b\right),\quad b=\overline{1,w}. \tag{30}$$

Such a point in a w-dimensional space for whose components condition (30) is satisfied, is termed ideal. In order to evaluate results of scientific-research activity of scientist $a_i$, it is necessary to find the degree of closeness between the ideal point and point $F_i\in R^w$. The degree of closeness in the case of quantitative evaluation is based on a specific metric. Proximity between two vectors is determined based on some metric distance $\rho(F_i, F^*)$ between these vectors. For the given problem, these distances can be calculated by formulas:

$$\rho^1\left(F_i,F^*\right)=\sqrt{\sum_{b=1}^{w}\left(f_i^b-f^{b*}\right)^2},\quad i=\overline{1,n}, \tag{31}$$

where $\rho^1(F_i, F^*)$ is the Euclidean distance,

$$\rho^2\left(F_i,F^*\right)=\sum_{b=1}^{w}\left|f_i^b-f^{b*}\right|,\quad i=\overline{1,n}, \tag{32}$$

where $\rho^2(F_i, F^*)$ is the city metric,

$$\rho^3\left(F_i,F^*\right)=\left(\sum_{b=1}^{w}\left|f_i^b-f^{b*}\right|^v\right)^{\frac{1}{v}},\quad i=\overline{1,n}, \tag{33}$$

where $\rho^3(F_i, F^*)$ is the Minkowski distance for period v>2, $v\in N$, N is the set of natural numbers.

The described method for the evaluation of results of scientific-research activity of scientists, which takes into account metric distances, makes it possible to construct for a particular scientist $a_i$ the integrated evaluations $\rho^u(F_i, F^*)$, $u=\overline{1,3}$. These evaluations are determined by the closeness of the vector of evaluations, obtained using different methods, to the so-called ideal point. In turn, the components of the ideal point represent the best evaluations by each of the methods (12)–(15), (26).

It should be noted that in addition to the evaluations $\rho^u(F_i, F^*)$, $u=\overline{1,3}$, other metric distances can be employed as an integrated assessment of the results of scientific-research activity of scientists. These metric distances must satisfy the criteria: symmetry, sustainability in the self-similarity, inseparability and the triangle rule.

## 6. Discussion of results of examining the evaluation of scientific research activity of scientists

As a result of present research, we constructed evaluation techniques of the results of scientific research activity of scientists and scientific areas. We proposed a method to calculate the evaluation of scientific research activity of scientists. This method, in contrast to the known indexes (h-index, g-index, e-index, index I-10), is characterized by the fact that it takes into account all the information about citation of authors, without losing any information about the publications. By using this method, we calculate a scalar evaluation of results of scientific research activity of scientists based on the solution to a system of linear algebraic equations, the matrix of coefficients represents degrees of publication citations among different scientists. A special feature of the method is that the similar matrix of coefficients of the system of linear algebraic equations is very sparse. For example, assume that a certain article contains about 20 citations. It is known that the number of publications indexed by Google Scientist over 2016 alone exceeded 6.35 million. By simple calculations we receive that every column of the matrix of coefficients constructed in accordance with the described method, only 0.0003 % elements will differ from zero. To solve such systems of linear algebraic equations, it is advisable to use the Gauss-Seidel numerical method. Given that the problem on finding the PR rating is effectively solved by the methods of simulation modeling [17], then in order to find the PR-q rating, it is possible to apply the method of Monte Carlo or other stochastic methods.

We proposed an integrated method to evaluate the results of activity of scientists, a characteristic feature of which is the construction of vectors whose components are the scalar evaluations of the results of scientific activity. These vectors are arranged in a multidimensional metric space. A so-called ideal point is also constructed, composed of scalar evaluations, the best in terms of achieving maximal effectiveness of activity. A metric distance between the ideal point or the point, which is assigned by a vector of evaluation of scientific research activity of a scientist, determines the cumulative assessment of the given scientist. A condition for the application of the method is the availability of sufficient information about publication citations of scientists. The advantage of the method is that the evaluation of results of scientist's output is calculated comprehensively by the integrated method taking into account the assessments of other indices. In addition, another benefit is assigning a metric space that allows expanding the range of calculated values of evaluations by employing different formulas for metric distances. The disadvantage of the integral method is the problem of selection and correction of the ideal point. Another shortcoming is that the components of the constructed vectors of evaluation of the results of scientific and research activity possess components that are clearly correlated. This is linked to the fact that the calculation of these components is based on the same data on the citation of scientific publications.

The proposed methods are flexible enough since they make it possible to correct the resulting evaluation by choosing different formulas to calculate the metric distances and the coefficients, which assign the ratio of citations among the publications of scientists.

Described research is continuation of the studies into construction of methods to evaluate the results of scientific research activity of scientists, appraisal of higher educational institutions and structural departments of these higher educational organizations. Some of the results of these studies were also examined in articles [6–8].

It is planned to formalize the procedure of clustering to identify the areas where group of scientists collaborate closely. Splitting the scientists by the areas of their activities will make it possible to assess a contribution of each of them into development of appropriate field. In addition, clustering will allow us to evaluate a direction as a separate object, by analyzing the history of its development. Moreover, breakdown by directions will make it possible to examine publication citations of authors who belong in a specific direction. Another task is the evaluation of dynamics of development of a particular scientific direction based on the results of clustering the scientists. It is planned to develop a method to calculate a short-term forecast of growth of change in the integrated assessments of a scientific direction. It is assumed that the method to be developed could be used to identify promising research areas that are formed in the scientific community, which will enable operational management of various aspects of particular scientific trends.

## 7. Conclusions

1. Most known indices for the evaluation of scientific-research activity of scientists, such as the h-index, g-index, e-index, index I-10, etc., do not account for full information on citation. We therefore proposed such method to calculate the evaluation of scientific research activity of scientists, which would not lose information about any citation of the author and of any publication. This method determines a scalar evaluation of the results of scientific activity and is based on determining a number of coefficients. The coefficients define the citation of one scientist by the publications of other scientists. The resulting estimates are found by solving a system of linear algebraic equations that are built based on the calculated coefficients.

2. There are a number of known indices for a scalar estimation of the results of scientific research activity of scientists. These indices generally provide the answer to the question: how many citations of a particular author did other scientists quote over a fixed period. This makes it possible to determine effectiveness of activity of the given author. However, most of the known evaluation approaches have their own peculiarities of calculation and the disad-

vantages that are associated with the loss of some of the information. That is why it is not recommended to rely solely on one of them. For the purpose of a comprehensive evaluation of scientific research activity of scientists, we proposed a method of the vector evaluation of results and the construction of the so-called integrated assessment. Underlying this method is the construction of vectors from scalar evaluations for each scientist in a multidimensional metric space. The number of calculated scalar evaluations determines measurability of the space. In addition, the basis of the method is the construction of an ideal point, which consists of scalar evaluations, the best in terms of achieving maximal effectiveness of activity. The evaluation of each scientist is calculated as a metric distance from the ideal point to the vector of scalar evaluations of the given scientist.

References

1. Lizunov, P. Vector project management of higher educational establishment [Text] / P. Lizunov, A. Biloshchytskyi, S. Biloshchytska // Management of Development of Complex Systems. – 2011. – Issue 6. – P. 135–139.
2. Otradskaya, T. Development process models for evaluation of performance of the educational establishments [Text] / T. Otradskaya, V. Gogunsky // Eastern-European Journal of Enterprise Technologies. – 2016. – Vol. 3, Issue 3 (81). – P. 12–22. doi: 10.15587/1729-4061.2016.66562
3. Yakovenko, V. Forecasting the state of the quality management system of educational institution [Text] / V. Yakovenko, V. Gogunskii // Research systems and informational technologies. – 2009. – Issue 2. – P. 50–57.
4. Gogunskii, V. «Lifelong learning» is a new paradigm of personnel training in enterprises [Text] / V. Gogunskii, O. Kolesnikov, K. Kolesnikova, D. Lukianov // Eastern-European Journal of Enterprise Technologies. – 2016. – Vol. 4, Issue 2 (82). – P. 4–10. doi: 10.15587/1729-4061.2016.74905
5. Sherstyuk, O. The research on role differentiation as a method of forming the project team [Text] / O. Sherstyuk, T. Olekh, K. Kolesnikova // Eastern-European Journal of Enterprise Technologies. – 2016. – Vol. 2, Issue 3 (80). – P. 63–68. doi: 10.15587/1729-4061.2016.65681
6. Koroleva, T. S. Evaluation criteria for research Institutes activities [Text] / T. S. Koroleva, I. A. Vasiliev, I. O. Torozhkov // Proceedings of the St. Petersburg Scientific Research Institute of Forestry. – 2014. – Issue 2. – P. 94–111.
7. Andrashko, Yu. Performance evaluation of teaching staff and universities overview [Text] / Yu. Andrashko, A. Biloshchytskyi, A. Kuchansky, S. Biloshchytska, T. Lyashchenko // Management of Development of Complex Systems. – 2017. – Issue 29. – P. 151–159.
8. Burkov, V. Parameters citation of scientific publications in scientometric databases [Text] / V. Burkov, A. Biloshchytskyi, V. Gogunskii // Informatization of higher education. – 2013. – Issue 15. – P. 134–139.
9. Hirsch, J. E. An index to quantify an individual's scientific research output [Text] / J. E. Hirsch // Proceedings of the National Academy of Sciences. – 2005. – Vol. 102, Issue 46. – P. 16569–16572. doi: 10.1073/pnas.0507655102
10. Egghe, L. Theory and practise of the g-index [Text] / L. Egghe // Scientometrics. – 2006. – Vol. 69, Issue 1. – P. 131–152. doi: 10.1007/s11192-006-0144-7
11. Zhang, C.-T. The e-Index, Complementing the h-Index for Excess Citations [Text] / C.-T. Zhang // PLoS ONE. – 2009. – Vol. 4, Issue 5. – P. e5429. doi: 10.1371/journal.pone.0005429
12. Kosmulski, M. A new Hirsch-type index saves time and works equally well as the original h-index [Text] / M. Kosmulski // International Society for Scientometrics and Informetrics. – 2006. – P. 4–6.
13. Egghe, L. The Hirsch index and related impact measures [Text] / L. Egghe // Annual Review of Information Science and Technology. – 2010. – Vol. 44, Issue 1. – P. 65–114. doi: 10.1002/aris.2010.1440440109
14. Gagolewski, M. Monotone measures and universal integrals in a uniform framework for the scientific impact assessment problem [Text] / M. Gagolewski, R. Mesiar // Information Sciences. – 2014. – Vol. 263. – P. 166–174. doi: 10.1016/j.ins.2013.12.004
15. Page, L. The PageRank Citation Ranking: Bringing Order to the Web [Text] / L. Page, S. Brin, R. Motwani, T. Winograd // Proceedings of the 7th International World Wide Web Conference. – Brisbane, Australia, 1998. – P. 161–172.
16. Avrachenkov, K. Monte Carlo Methods in PageRank Computation: When One Iteration is Sufficient [Text] / K. Avrachenkov, N. Litvak, D. Nemirovsky, N. Osipova // SIAM Journal on Numerical Analysis. – 2007. – Vol. 45, Issue 2. – P. 890–904. doi: 10.1137/050643799
17. Liao, Q. Monte Carlo Based Incremental PageRank on Evolving Graphs [Text] / Q. Liao, S. Jiang, M. Yu, Y. Yang, T. Li // Lecture Notes in Computer Science. – 2017. – P. 356–367. doi: 10.1007/978-3-319-57454-7_28
18. Kuchanky, A. Selective pattern matching method for time-series forecasting [Text] / A. Kuchanky, A. Biloshchytskyi // Eastern-European Journal of Enterprise Technologies. – 2015. – Vol. 6, Issue 4 (78). – P. 13–18. doi: 10.15587/1729-4061.2015.54812