A METHOD TO EVALUATE THE SCIENTIFIC ACTIVITY QUALITY OF HEIS BASED ON A SCIENTOMETRIC SUBJECTS PRESENTATION MODEL

A. Biloshchytskyi
Doctor of Technical Sciences, Professor*
E-mail: bao1978@gmail.com

O. Myronov
Postgraduate student
Department of IT**
E-mail: allexx@mail.ua

R. Reznik
Postgraduate student
Department of IT**
E-mail: nibros88@gmail.com

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

Yu. Andrashko
Lecturer
Department of System Analysis and Optimization Theory
Uzhhorod National University
Narodna sq., 3, Uzhhorod, Ukraine, 88000
E-mail: yurii.andrashko@uzhnu.edu.ua

S. Paliy
PhD, Associate Professor*
E-mail: paliy@fit.knu.ua

S. Biloshchytska
PhD, Associate Professor
Department of Information Technology Designing and
Applied Mathematics**
E-mail: bsvetlana2007@ukr.net
*Department of Network and Internet Technologies
Taras Shevchenko National University of Kyiv
Volodymyrska str., 60, Kyiv, Ukraine, 01033
**Kyiv National University of Construction and Architecture
Povitroflotskyi ave., 31, Kyiv, Ukraine, 03037

Запропоновано метод інтегральної оцінки якості наукової діяльності вищих навальних закладів та науково-дослідних структурних підрозділів цих вищих навчальних закладів. В основу метода покладено комплекс розроблених метрик перетворення якісних показників наукової діяльності у кількісні. Передбачається, що кожен числовий еквівалент якісного показника відображає рівень задоволення деяким критеріям та вимогам, які пред'являються до відповідних вищих навчальних закладів

Ключові слова: оцінка якості наукової діяльності, наукометричні суб'єкти, інтегральна оцінка, аналіз наукової діяльності, інтерпретація звітної діяльності

Предлагается разработать метод интегральной оценки качества научной деятельности высших учебных заведений и научно-исследовательских структурных подразделений этих высших учебных заведений. В основу метода положен комплекс разработанных метрик преобразования качественных показателей научной деятельности в количественные. Предполагается, что каждый числовой эквивалент качественного показателя отображает уровень удовлетворения некоторым критериям и требованиям, предъявляемым к соответствующим высшим учебным заведениям

Ключевые слова: оценка качества научной деятельности, наукометрические субъекты, интегральная оценка, анализ научной деятельности, интерпретация отчетной деятельности

1. Introduction

The era of the active introduction of information systems in the activities of units, institutions and organizations in order to implement measures to improve their efficiency establishes new requirements for improving and developing the outdated models and methods of operating information systems. Optimization of scientific, managerial, informational and educational activities of higher education institutions (HEIs) consists in reforming the underlying principles and developing new concepts of displaying and evaluating projects and products of scientific and educational activities. Such projects and products are perceived as scientometric entities in relation to higher educational establishments or other structural research units. The efficiency of an institution of higher learning entails not only the qualitative level of training students but also the volume and level of research activities carried out on the basis of the higher education institution.

The analysis of the reporting activity of universities has revealed that the rating of research activities is determined without regard to the scientific capacity of the universities. That is, due to the greater number of researchers, professors and students, and, consequently, the larger amount of funding, one category of higher education institutions will always receive a higher rating of research activities, and vice versa. Therefore, there is a problem of handling the issue of comparing universities with one another, at least in relation to the research field of activity. Thus, it is necessary to reduce the reporting indicators of educational activities of universities to the level at which quantitative efficiency values can be contrasted and compared across different HEI categories. This will allow a qualitative assessment of the scientific activity of each institution that, to various degrees, deals with science.

The rapid involvement of information technology in the scientific and educational process is a solid foundation for creating and updating modern scientometric databases, which at present are quite numerous. Their diversity, however, suggests that there is currently no single model for creating, maintaining, using, and, most importantly, evaluating the scientific activity of scientometric subjects. It is the assessment of the scientific capacity of a particular subject that plays a decisive role in comparing the "contribution to science".

The importance of this study is that the implementation of measures highlighted in relation to the development and improvement of scientometric databases will increase the quantitative and qualitative indicators of each subject of scientific activity. Therefore, the information culture of scientists should be formed and scientific research should be developed with the use of scientometric databases, which should become the main requirement in the conditions of rapid informatization of the empirical activity of individuals, universities and other scientometric subjects. The development and modernization of the system of evaluating and indexing scientific works should inspire scientists to create a quality product that can become the basis for further thorough research and new developments in scientific directions. This will make scientists interested, at least from the point of view that first and foremost the number of scientific works by an author and the number of references to respective works will determine the author's rating among other scholars. Consequently, the quality of scientific activity of a separate institution, on the basis of which relevant research is carried out, will grow.

2. Literature review and problem statement

We will assume that the scientific activity quality (SAQ) of scientometric subjects, in particular HEIs, is a certain numerical equivalent that reflects the aggregate level of complying with the relevant requirements and criteria that are set for HEIs and/or research units by the Ministry of Education and Science (MES). It is assumed that each institution must meet the requirements imposed on it. The fulfillment of the requirements should take place on the specified terms and at the appropriate qualitative level.

The task of evaluating the quality of scientific activity implies that higher education institutions are the highest level in the hierarchy of scientometric subjects (authors – works – journals – HEIs). Then it becomes apparent that in order to achieve quality of scientific work, higher education institutions must calculate data for entities of a lower level. The lower-level entities include articles published by the institution, magazines (if the institution is a publisher), as well as authors of relevant works.

Studies [1, 2] are similar in structure and content; they address the issues of detecting plagiarism in scientific articles or dissertation studies. The problem of plagiarism and its detection is one of the components of evaluating scientific activities effectively. However, the studies do not contain recommendations for authors to minimize plagiarism while performing work, and they do not suggest motivation for writing high-quality scientific works. Article [3] offers a method for finding scalar estimates, as well as a method for constructing integral estimates of the results of scientific research activities of scholars. The study also suggests methods that can be used in integrated assessment of scientists and higher education institutions. However, the peculiarities of applying constructed estimates of research activities of scientists for assessing higher education institutions are not described in detail. Paper [4] highlights the issues of parametric formalization of the educational system and the transition to a multifactor quality assessment. However, the concept of "quality" contains optimization of processes occurring in institutions, bypassing the definition given in the beginning of the section. Studies [5, 6] have determined that the most common characteristics of scientific work productivity are the number of works, the number of citations, and the generalized indicator of scientific performance – the h-index (Hirsch Index).

However, these studies do not address the "scientific capacities" of a particular scientific organization or institution and do not take into account the rank of works. It is assumed that all works are equal: conference proceedings, articles in domestic journals, articles in international collections, etc. No account is taken of the status of journals. In addition, the h-index, which accompanies these articles as the proposed reliable estimate, may completely differ in various scientometric databases, since this index depends, firstly, on the coverage area of the web space and secondly, on the selected database. Therefore, it is hardly possible to consider the h-index a universal indicator. Article [7] examines the principles of managing innovative projects and programs in the context of a crisis, which is an important topic at present. The formal model of conceptualization of the best practices and knowledge in project management is thoroughly reviewed and described. Given the fact that research and development (R&D) management can also be considered a project and a set of programs, there are no methods and suggestions for adapting this model to each field of activity, in particular, the scientific one. Examples of the proposed model implementation in practice are not provided at all.

Study [8] reviews and compares universities in the United States of America and Europe. The research is based on a university-wide representation in a set of institutional constituents: the size and composition of the faculties, the placement and structure of funding. However, the study focuses on correlations of the institutional components of higher education institutions with the distribution of funding and the relationship between the composition of universities and regional GDP per capita. Article [9] analyzes the total number of citations of individual scholars as well as aggregate "production" and quantitative determination of the impact of scientific works. Studies are recommended as bench-

marks for theoretical models of career growth. Work [10] offers a new algorithm and validation method to eliminate the ambiguity of authors' identification in large volumes of bibliographic data. However, the research covers and implies only the citation index of the Web of Science database.

The analysis of published data has revealed that attention to the issues of studying the scientific activity of universities, namely, the interpretation of the reporting parameters in order to increase the efficiency of scientific activity, has not been sufficiently researched. In view of this, solving the problem is a promising objective. By solving this and accompanying problems, it is possible to achieve significant progress in sciences.

3. The aim and objectives of the study

The aim of the work is to determine the generalized quality index and the results of scientific research of an individual scholar, university department, and higher education institution. Therefore, the main objective is to develop the theoretical foundations, models and methods, mechanisms and tools that will form the basis for the concept of presenting and evaluating scientometric content. This, in turn, will provide the opportunity to standardize and unify the concepts under consideration through a set of proposed solutions.

Thus, in order to achieve this aim, it is proposed to solve consistently the following tasks:

– to develop a mathematical model for representing scientometric subjects in the web-space;

– to implement the method of integral assessment of the quality of scientific activity of higher educational establishments and/or other research units;

– to describe the practical value of the results of the research for authors of scientific works.

4. A mathematical model of representing scientometric subjects

The purpose of placing scientific works on the Internet is to create a global database on scholars, scientific works, scientific journals, and institutions for determining the rating of citation and popularity of the above-mentioned subjects.

Let us consider a mathematical model of representing scientific works as a system with an ordered set of the following elements, and let us submit it in the form of the tuple

$$N = \langle U, S, R, M, f_{u,s} \rangle, \tag{1}$$

where $U = \{u^1, u^2, ..., u^{n_u}\}$ is a set of users of a system (authors, editors, etc.), u^r, $r = \overline{1, n_u}$ is a user of the system, n_u is the number of users; $S = \{s_1, s_2, ..., s_{n_s}\}$ is a set of scientific and journalistic objects of the system; s_i, $i = \overline{1, n_s}$ is a published scientific work, n_s is the number of published scientific works; $R = \{R_1, R_2, ..., R_{n_r}\}$ is a set of applications from users of the system (input parameter), n_r is the number of applications; $M = \{M_1, M_2, ..., M_{n_m}\}$ is a set of messages to users of the system (output parameter), n_m is the number of messages; $f_{u,s}$ is the function of evaluating the published works, taking into account the number of external references to the published works S from the users $U', U' \in U$, and also the "weight" of the user $U'' \subseteq U'$, who is an author and refers to the corre-

sponding work. At once we shall note that in assessing the scientific activity of scientometric subjects, we do not take into account the quality of the reviewed scientific works. That is, the semantic content of the published studies is not considered.

The set of the scientific journalistic objects S of a system N determines the available published scientific works in this system, each s_i of which is determined by a set of attributes and metadata:

$$s_i = \{v, f, m, D\}, \tag{2}$$

where $s_i \in S$ is each published scientific work; v is the volume of the scientific article (the number of words, the number of characters, the file size, taking into account the format, images, etc.); f is the work format (*.doc, *.rtf, *.pdf, etc.); m is additional materials accompanying the published work (graphs, diagrams, etc.); and D is a set of scientific work metadata.

We shall provide the following definitions.

Under _application_ R, we will consider a request from a user to publish a scientific article s_i, and it will be defined as the tuple

$$R_i^r = \langle u^r, s_i \rangle,$$

$$R \subset U \times S, \tag{3}$$

where u_r is the user who is the source of the request; s_i is the scientific article submitted by the user for consideration.

Message M will be considered as a set of recommendations to correct/edit or approve of the submitted article, which is determined by the tuple

$$M_{i\,k}^r = \langle R_i^r, T_k \rangle, \tag{4}$$

where R_i^r is the application from the user; $T = \{T1, T2, ..., Tk\}$ is a set of messages and/or recommendations for correction or approval of a scientific article, and $k = \overline{1, n_T}$, n_T is the number of responses.

Thus, we assume that some journal J consists of n scientific journalistic objects O_y (issues): $y = \overline{1, n}$:

$$J = \{O_1, O_2, ..., O_n\}, \tag{5}$$

where $O_y \subseteq S$.

For a number of O_y issues, related to the corresponding journals J by some composition γ_o, it is possible to compare m_o images (scientific articles) each of which is a scientific journalistic object s_i; then we obtain:

$$\gamma_o : J \rightarrow S,$$

$$\gamma_o(O_y) = \{S_1^{o_y}, S_2^{o_y}, ..., S_{m_o}^{o_y}\}, \tag{6}$$

where $S_{m_o}^{o_y}$ is a set of scientific articles in the issue O_y.

Besides, it is necessary to ensure that the following conditions are met for the correct presentation format:

$$\left\{ \begin{array}{l} v_{s_i} \leq v_{\max}, v_{s_i} \in s_i^{o_y} \\ p_{i\min} \leq p_i \leq p_{i\max}, p_i \in D \subset s_i^{o_y} \end{array} \right\}, \forall s_i \in O_y, \tag{7}$$

where v_{s_i} is the volume of the article $s_i^{o_y}$ (the number of words, the number of characters, the file size, taking into ac-

count the format, images, etc.); $v_{\max}$ is the maximum allowed size of the article $s_i^{o_y}$; p_i is the metadata attribute of the article D, which displays the number of internal references in the article $s_i^{o_y}$; $p_{i\min}$ is the minimum number of internal references in the article $s_i^{o_y}$; $p_{i\max}$ is the maximum number of internal references in the article $s_i^{o_y}$.

The function $f_{u,s}$ is based on calculating the product of the ratio of the total size ("weight") of a specific citation to the total volume of the text of the scientific article and the "weight" of the author referring to the corresponding article. As a resource for assessing the quality and results of the scientific activity of the scientometric entities, there is the totality of the scientometric subjects E:

$$E = \langle S, J, I \rangle, \tag{8}$$

where E is a tuple of scientometric subjects; S is the set of articles to be evaluated; J is the set of journals to be evaluated; I is the set of institutions to be evaluated.

Let us consider the set of components of the scientific content DB as an aggregate of sets of data, metadata, and documents:

$$DB = \bigcup_{i=1}^{n_E} D(C_i), \tag{9}$$

where n_E is the number of levels of the hierarchy of scientometric subjects that should be evaluated (in this case, there are three of them: articles, journals, and institutions); $D_i(C_i)$ is a set of data and metadata of the scientometric subjects that correspond to the current level of the C_i hierarchy (article, journal, and institution):

$$f_{u,s} : E \rightarrow \bigcup_{i=1}^{n_E} C_i,$$

$$C_i = \{S \vee J \vee I\}. \tag{10}$$

Based on the calculation formulae (9) and (10), the mathematical model of the system, which is the basis for evaluating the scientific activity of scientometric subjects, will take the form:

$$DB = \left\langle \bigcup_{i=1}^{n_S} D^S(C_i), \bigcup_{i=1}^{n_J} D^J(C_i), \bigcup_{i=1}^{n_I} D^I(C_i) \right\rangle, \tag{11}$$

where $D^S(C_i)$ is a set of data and metadata of the scientometric entities reflecting the level of the hierarchy "published work" S, n_S is the total number of published works in the system; $D^J(C_i)$ is a set of data and metadata of the scientometric entities that represent the level of the hierarchy "journal" J, n_J is the total number of journals in the system; $D^S(C_i)$ is a set of data and metadata of scientometric subjects reflecting the level of the hierarchy "institution" I, n_I is the total number of institutions in the system.

It is clear that the system is dynamic, that is, a system that changes in time. The variable in this system is the number of scientific articles (in the case of one magazine of a university or organization); therefore, we modify the model as follows:

$$DB = \left\langle \bigcup_{t=t_0}^{t_1} \left(\bigcup_{i=1}^{n_S} D^S(C_i) \right), \left(\bigcup_{i=1}^{n_J} D^J(C_i) \right), \left(\bigcup_{i=1}^{n_I} D^I(C_i) \right) \right\rangle, \tag{12}$$

where t_o is the time when the first reference to the published work is made in the system; t_1 is the current moment of time; $t \in [t_0; t_1]$ is the period of time within which there is an assessment of a higher education institution for J=const, I=const.

We consider time to be discrete with a certain period Δt={day, week, month, quarter, year, ...}:

$$t = t_0 + \kappa \Delta t, \quad 0 \le \kappa \le \frac{t_1 - t_0}{\Delta t}. \tag{13}$$

Then, the selection operator π to obtain the necessary permutation, projection, or attribute value of the resulting data set to evaluate, for example, scientific works will look as follows:

$$\pi : S \rightarrow \Re,$$

$$Z(DB) = \{\pi(s_i) \,|\, \overline{i = 1, n}, \sigma(s_i) = 1\}, \tag{14}$$

where Z(DB) is a request to evaluate of scientific articles; $\sigma(s_i)$ is the condition of choice, which is equal to 1 if condition (7) is satisfied or equal to 0 if condition (7) is not met; s_i is a scientific article to which the request has been generated.

5. Implementation of the method for assessing the quality of scientific activity of HEIs

The purpose of the suggested method consists, firstly, in developing a concept and means for assessing the scientific activity quality (SAQ) of universities and institutions, and secondly, in implementing the proposed models and methods into some information management system for scientometric content. With the help of the developed method, it will be possible to compare university ratings as well as the effectiveness of higher educational institutions against arbitrary criteria at different time intervals and time points.

However, the complexity of developing a method for estimating the SAQ of higher educational institutions consists in the impossibility of comparing higher education institutions for a number of parameters. Therefore, at this stage, it is necessary to develop appropriate metrics that will allow grouping HEIs by some characteristics and determining the corresponding normalizing coefficient or function for each group.

Let us consider the quality of scientific activity of higher educational institutions as a numerical measure of the level of compliance with the SAQ requirements set for HEIs by the Ministry of Education and Science of the Academy of Sciences. To do this, it is necessary to compile data on HEIs compliance with the relevant requirements and criteria, as well as to perform a mathematical transformation that will bring all the data to values in the interval [0; 1].

Let us consider the set of criteria

K={k_1, k_2, ..., k_i}, N=card(K).

We represent the compliance with a set of requirements and criteria for a HEI at some point (period) of time T in the form of a segment with areas of qualitative compliance with these criteria: "high quality", "satisfactory", "poor" (Fig. 1).
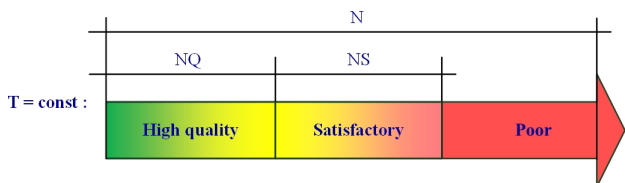
Fig. 1. Representation of the SAQ of a HEI at a certain time (period) *T*

In Fig. 1:

N is the total number of criteria and requirements that HEIs must meet;

NQ is the number of criteria compliance with which corresponds to the area "high quality";

NS is the number of criteria compliance with which corresponds to the area "satisfactorily";

"High quality": compliance with this criterion is considered to be fully in line with the requirements and capabilities of some higher education institutions;

"Satisfactory": in general, the value of the quality rate is acceptable but for some reasons cannot be considered high quality;

"Poor": the value of the quality rate does not meet the relevant criterion or requirement for a HEI.

The SAQ of a HEI is proposed to be calculated as a ratio of the number of criteria that correspond to the requirements for the total number of criteria. Moreover, the number of criteria satisfying the requirements is the sum of the number of criteria of the "high quality" range and half of the number of criteria that fall within the area "satisfactory":

$$SAQ_t = \frac{NQ(t) + \dfrac{NS(t)}{2}}{N}, \qquad (15)$$

where *t* is one of the moments of time $t_0 <= t <= t_1$.

The formula does not explicitly take into account the number of criteria whose satisfaction corresponds to the area "poor"; if necessary, this number can always be calculated by subtracting the number of criteria that correspond to the requirements of the total number of criteria and requirements: N−(NQ+NS).

In order to assess the SAQ of HEIs, it is primarily necessary to determine the list of criteria and requirements for higher education institutions and then to carry out the classification as to the compliance with the criteria (Table 1). Next, real requirements and criteria for reporting indicators of Kyiv National University of Construction and Architecture (Ukraine) will be used in this study. The list of requirements and criteria was set by the Ministry of Education and Science of Ukraine as valid for 2013–2014; however, the values of the indicators and the size of the sample differ from the original ones.

For each criterion or requirement, there is a certain value of the indicator. Each indicator, in turn, is regulated by some norms or features, considering which, each of these parameters can be attributed to one of the three areas of complying with the needs of the scientific activity quality of a higher education institution:

$$R \rightarrow \{"high\,quality","satisfactory","poor"\}^N, \qquad (16)$$

where R is the qualitative value of each criterion k_i.

Since there is a need for a quantitative assessment of each qualitative indicator, we will involve experts for the objective assessment of the university's SAQ. Then, taking into account the calculation for each of the criteria,

$$SAQ_T(\kappa) = \frac{NQ_E(t) + \dfrac{NS_E(t)}{2}}{N_E}, \qquad (17)$$

where N_E is the total number of experts who evaluate the university's performance; NQ_E is the number of experts who evaluate criteria or requirements as "high quality"; NQ_E is the number of experts who evaluate the university's performance as "satisfactory".

Let us assume that the expert commission for the assessment of reporting will include a constant number of experts: N_E=const.

Table 1

Classification of criteria and requirements for higher education institutions

No.	Criteria (require-ments)	Input in-form-ation	Output in-form-ation	Rule	Compliance with the requirements		
					High quality	Satis-factory	Poor
1	Faculty	Faculty database	Number of professors	Enquiry to the database	−	−	+
2	Number of laureates of state prizes	Faculty database	Number of laureates	Enquiry to the database	−	+	−
n	Participa-tion in projects	List of projects	Number of projects	List of projects	+	−	−

The next stage requires the calculation of the SAQ index for each criterion (Table 2) and its qualitative interpretation, expressed in color, in accordance with (Fig. 2, 3).

Table 2

Calculation of the SAQ index for the criteria

No.	Criterion (requirement)	Value	N_E	NQ	NS	SAQ
1	The number of international grants, scientific and educational projects and programs co-performed by the higher education institution	12	100	70	15	0.78
2	The number of agreements entered into with foreign universities for the training of students in the program "Double Diploma"	0	100	0	0	0.00
3	The number of international exhibitions in science, education, and technology, in which the achievements of the higher education institution are represented	9	100	62	24	0.74
4	The number of awards (medals, diplomas) received by the higher education institution at international exhibitions in science, education, and technologies, in which the achievements of the higher education institution are represented	72	100	90	10	0.95
5	The number of awards won by prize-winners at national sports competitions (Ukrainian Championships, Universiades of Ukraine, and Ukrainian Championships among students)	12	100	80	12	0.86

For example, only 5 criteria are considered, the data of which are taken from the report of Kyiv National University of Construction and Architecture (Ukraine) for the period of 2013–2014 academic year.

Let us consider the range of values used in the SAQ method and represent it in the form of a scale (Fig. 2). The intervals of its values are bound to the criteria for evaluating on the scale "good – bad" (Table 3).
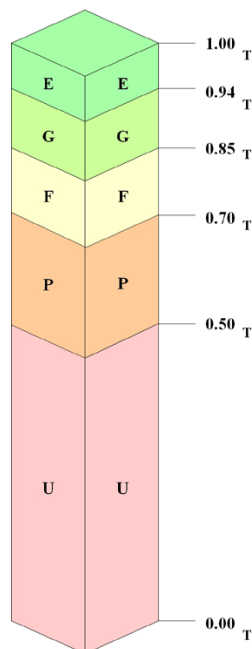

Fig. 2. The scale of SAQ

Table 3

Interpretation of the intervals of the values of the SAQ-determining method for the HEI

The scale of SAQ		
Value		Grade
From	To	
0.94	1.00	E – Excellent
0.85	0.94	G – Good
0.70	0.85	F – Fair
0.50	0.70	P – Poor
0.00	0.50	U – Unacceptable

For the accuracy and objectivity of the evaluation, we requested experts to evaluate the quality by the qualitative assessment levels. Next, we made the transformation to create three main of the five qualitative values (15):

$$NQ = \frac{E+G}{2}, \quad NS = \frac{F+P}{2}. \tag{18}$$

It is obvious that the SAQ will always be a value from 0 to 1. On the scale of grading, "1" is a perfect SAQ, and "0" is completely unacceptable. Intermediate evaluation criteria are given in Table 3.

Since the state of achieving the SAQ by an HEI changes at any time, it is expedient to rewrite the formula in such a way that the SAQ of the HEI could be compared not only in terms of a certain moment of time but also to understand the trends of the development of higher educational institutions over time. Therefore, we use a weighted sum for each criterion:

$$SAQ_{HEI}(\kappa) = \frac{\Delta t}{t_1 - t_2} \sum_{t=t_0}^{t_1} \frac{NQ_E(t) + \frac{NS_E(t)}{2}}{N_E}. \tag{19}$$

We will calculate the total index of SAQ_{HEI}, using the sum of the criteria assigned to each area of compliance:

$$n_{NQ} = \text{Count}\left(\sum NQ(k)\right), \; n_{NS} = \text{Count}\left(\sum NS(k)\right),$$

$$SAQ_{HEI} = \sum_{k=1}^{n_k} \frac{n_{NQ} + \frac{n_{NS}}{2}}{N}, \tag{20}$$

where n_{NQ} is the number of sums of criteria classified as "high quality"; n_{NS} is the number of sums of criteria classified as "satisfactory"; and n_k is the number of criteria.

After obtaining the SAQ value for each institution and/or unit, it is possible to compare the success of the institutions. There is also an opportunity to adjust optimally the allocation of resources of an institution to address the imperfections that become apparent when using the proposed approach.

6. Discussion of the results of the research on the assessment of the scientific work quality of a HEI

Since all universities, structural units and research organizations are similar in composition and structure, it is necessary, above all, to build the categorization of HEIs into groups for roughly equal reporting values. This can be achieved by finding the similarity coefficient of the quantitative parameters of similar reporting indicators for each institution, the values of which would significantly differ from each other. This should be done because of the inequality of the "scientific capabilities" of individual universities and organizations, which depend on the level of accreditation, the level of funding of the institution, the number of scientific and pedagogical staff, etc. The coefficients of similarity of the indicators can be found in the following way:

1. To calculate the ratio, for example, of an indicator of the number of authors' articles to, for example, an indicator of the number of articles for a certain period, and, similarly, to perform the specified actions with the data of other types of comparisons within each category of universities.

2. To calculate the ratio of the number of characteristics obtained in paragraph 1 to the same characteristics obtained for other categories of universities.

It is proposed to categorize universities based on the maximum value of roughly equal reporting values within the established limits for each category. Based on the maximum values of performance indicators in the last period, it is possible to compare the quality indicators of universities and the scientific and educational activities carried out on the basis thereof among themselves. Then it is necessary to define such linguistic variables that will reflect the qualitative state of compliance with the relevant indicators. For example, the variable "scientific publishing activity" can be assigned a value excellent, good, fair, satisfactory, and unacceptable. For each value of a variable, it is necessary to set the appropriate numeric equivalents that must be justified and standardized. This can be achieved by using fuzzy logic techniques, which will provide flexibility when setting up and adjusting parameters.

The expediency of creating and using the developed method is the ability to compare the ratings of universities, comparing the effectiveness of universities in accordance with the criteria that are presented to them. Also, it is possible to trace the dynamics of the development of universities and structural research units.

However, the introduction of the proposed models in practice should take into account the fact that the number of scientometric subjects, and hence metadata, including the metadata of the scientific articles themselves, will accumulate over time. In addition, metadata may become obsolete, that is, the entropy of such a system will also increase over time. Therefore, an important task is to maintain relevance in the system of mapping and evaluating the research performance of the scientometric subjects. The growth of entropy involves the presence of uncoordinated data in the system H(N):

$$H(N) : \lim_{t \to +\infty} X_i = 1,$$

where t is the time when data and metadata are not updated in the system.

Since systems of this scale include constant access to the database, as well as the processing of some data, it is necessary to update them regularly to support the system in a relevant state: $H(N) : \lim_{t \to +\infty} X_i = 0.$

The proposed method of integral evaluation of the scientific performance quality of HEIs can be used in the development of control systems for scientometric content, or in the development of a plug-in and its integration into existing systems of reflecting scientometric content.

7. Conclusion

1. The proposed mathematical model of representing scientometric subjects in the web-space creates a general picture of the scientometric information environment. It helps distinguish between entities that are subjects of scientometrics, determine the connections between these entities, and form a scientific and methodological concept of mapping scientometric subjects on the Internet. This, in turn, can be an impetus for authors of scientific articles and researchers to move to a higher quality level of writing scientific and journalistic works. It is essential because, in terms of the desire to be "higher in the ranking" of popular studies, authors tend to write high-quality and trend works.

2. The implementation of the principles set forth in this study on evaluating the scientific activity quality of HEIs in the information system will enable institutions and structural research units, in addition to obtaining statistical data, to respond in a timely manner to imperfect indicators, thereby stimulating the improvement of the quality of their own scientific and educational activities. Moreover, the provision of open access to the reporting information of all universities will also lead to a competition, which will also affect the growth of the efficiency and quality of educational activities of universities in the best possible way.

3. The results of the research can be considered as an incentive for authors to pursue modern and technological scientific developments and to present them in scientific works, which, in turn, will serve as an integral part in assessing the quality of scientific work of a university with the help of the proposed method. This research was aimed at solving the contradictions in evaluating scientometric subjects that consist of heterogeneous categories of higher education institutions, types of published scientific works, categories of journals in which these works are published, etc. Such contradictions make it impossible to compare and objectively evaluate existing and newly created scientific works and hence all other scientometric subjects. This is due to the fact that a published scientific work is a key entity, and it should be the basis in the evaluation and indexation of others.

References

1. Biloshchytskyi, A. Conceptual model of automatic system of near duplicates detection in electronic documents [Text] / A. Biloshchytskyi, A. Kuchansky, S. Biloshchytska, A. Dubnytska // 2017 14th International Conference The Experience of Designing and Application of CAD Systems in Microelectronics (CADSM). – 2017. doi: 10.1109/cadsm.2017.7916155

2. Lizunov, P. Detection of near dublicates in tables based on the locality-sensitive hashing method and the nearest neighbor method [Text] / P. Lizunov, A. Biloshchytskyi, A. Kuchansky, S. Biloshchytska, L. Chala // Eastern-European Journal of Enterprise Technologies. – 2016. – Vol. 6, Issue 4 (84). – P. 4–10. doi: 10.15587/1729-4061.2016.86243

3. Biloshchytskyi, A. Evaluation methods of the results of scientific research activity of scientists based on the analysis of publication citations [Text] / A. Biloshchytskyi, A. Kuchansky, Y. Andrashko, S. Biloshchytska, O. Kuzka, O. Terentyev // Eastern-European Journal of Enterprise Technologies. – 2017. – Vol. 3, Issue 2. – P. 4–10. doi: 10.15587/1729-4061.2017.103651

4. 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

5. 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

6. Hirsch, J. E. Does the h index have predictive power? [Text] / J. E. Hirsch // Proceedings of the National Academy of Sciences. – 2007. – Vol. 104, Issue 49. – P. 19193–19198. doi: 10.1073/pnas.0707962104

7. Bushuyev, S. D. Convergence of knowledge in project management [Text] / S. D. Bushuyev, D. A. Bushuyev, V. B. Rogozina, O. V. Mikhieieva // 2015 IEEE 8th International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS). – 2015. doi: 10.1109/idaacs.2015.7341355

8. Wolszczak-Derlacz, J. An evaluation and explanation of (in)efficiency in higher education institutions in Europe and the U.S. with the application of two-stage semi-parametric DEA [Text] / J. Wolszczak-Derlacz // Research Policy. – 2017. – Vol. 46, Issue 9. – P. 1595–1605. doi: 10.1016/j.respol.2017.07.010

9. Petersen, A. M. Statistical regularities in the rank-citation profile of scientists [Text] / A. M. Petersen, H. E. Stanley, S. Succi // Scientific Reports. – 2011. – Vol. 1, Issue 1. doi: 10.1038/srep00181

10. Schulz, C. Exploiting citation networks for large-scale author name disambiguation [Text] / C. Schulz, A. Mazloumian, A. M. Petersen, O. Penner, D. Helbing // EPJ Data Science. – 2014. – Vol. 3, Issue 1. doi: 10.1140/epjds/s13688-014-0011-3