UDC 005.8
DOI: 10.15587/1729-4061.2021.233655

DEVELOPMENT OF THE SET MODELS AND A METHOD TO FORM INFORMATION SPACES OF SCIENTIFIC ACTIVITY SUBJECTS FOR THE STEADY DEVELOPMENT OF HIGHER EDUCATION ESTABLISHMENTS

Andrii Biloshchytskyi**
Doctor of Technical Sciences, Professor*
Alexander Kuchansky
Corresponding author
PhD, Associate Professor**
E-mail: kuczanski@gmail.com
Yurii Andrashko
PhD, Associate Pofessor
Department of System Analysis and Optimization Theory
Uzhhorod National University
Narodna sq., 3, Uzhhorod, Ukraine, 88000
Serik Omirbayev
Doctor of Economical Sciences, Professor
Rectorate*
Aidos Mukhatayev
PhD, Associate Professor
Department of Strategy and Corporative Management*
Adil Faizullin
Master of Technical Sciences
Department of Quality Assurance*
Sapar Toxanov
Doctoral Student
D. Serikbayev East Kazakhstan Technical University
A. K. Protozanov str., 69, Ust-Kamenogorsk,
Republic of Kazakhstan, 070004
Center of Competence and Excellence*
*Astana IT University
Mangilik Yel ave., EXPO Business Center, Block C.1,
Nur-Sultan, Republic of Kazakhstan, 010000
**Department of Information Systems and Technologies
Taras Shevchenko National University of Kyiv
Volodymyrska str., 60, Kyiv, Ukraine, 01033

This paper describes the basic conceptual apparatus required to form information spaces for scientific activity subjects. Set models have been built to identify collective and individual scientific activity subjects, including information on the subjects' publication citations, their abstracts, as well as their indicators in scientometric databases, etc. A conceptual scheme of interaction between collective and individual scientific activity subjects has been described, taking into consideration the dynamics of their productivity.

A method has been proposed to form the information spaces for the collective and individual scientific activity subjects such as higher education establishments and scientists. The method involves a series of stages to identify and construct citation and scientific cooperation networks, to form subject scientific spaces, and, based on them, to devise methods in order to quantify productivity. The results of methods application form the components of the relevant information spaces of scientific activity subjects. The spaces to be built could be used to solve the task of selecting subjects for the implementation of joint scientific and educational projects. In addition, these spaces could be applied to form the organizational and functional framework of the collective scientific activity subjects, including their structural units, which would contribute to ensuring their stable development.

Creating the information spaces of scientific activity subjects underlies resolving those issues that would stimulate investment in research and innovation, strengthen cooperation between universities, improve the efficiency and productivity of the scientific enterprise. It has been confirmed experimentally that the potential of a collective subject of scientific activity, including individual subjects, the rate of change of identifiers of whom is positive, would have a non-negative potential. A rate of change in the normalized indicators of identifiers of individual and collective scientific activity subjects has been calculated for the period from January 2019 to December 2020 for three higher education establishments

Keywords: information space, scientific activity subject, higher education establishment, set model

Received date: 07.04.2021
Accepted date: 08.06.2021
Published date: 29.06.2021

***How to Cite**: Biloshchytskyi, A., Kuchansky, A., Andrashko, Y., Omirbayev, S., Mukhatayev, A., Faizullin, A., Toxanov, S. (2021). Development of the set models and a method to form information spaces of scientific activity subjects for the steady development of higher education establishments. Eastern-European Journal of Enterprise Technologies, 3 (2 (111)), 6–14. doi: https://doi.org/10.15587/1729-4061.2021.233655*

1. Introduction

The development of models and methods of scientific research management has an important role in order to improve scientific activity, stimulate basic and applied research, promote innovations. In this regard, it is important to devise procedures for ensuring the formation of rational composition and structure of scientific activity subjects (higher education establishments and their structural units, as well as scientists). Such procedures should be based on objective indicators of the scientific activity performance, taking into consideration all elements in the information space of scientific activity subjects.

Each subject of scientific activity exists and evolves within its information space where information on its development is accumulated and stored. The information space of a scientific activity subject maps data on the results of scientific activity, performance assessments, citation of scientific publications, etc. The output of scientific activity subjects depends on the understanding of the principles underlying the construction and functioning of their information space. In addition, launching scientific or educational projects that involve scientific activity subjects ensures the synergistic interaction of their information spaces, the transition of values from one information space to another.

Devising a method that forms the information spaces of scientific activity subjects, as well as studying relations among information spaces, ensures the steady evolution of higher education establishments. In addition, such studies are important for the development of strategic decisions to improve the scientific potential in general. The relevance of the formation of information spaces of scientific activity subjects is also defined by the strategic goals of the European Research Area (ERA) [1], which is currently the dominant concept for the creation of a single market for research, technology, and innovation in EU. ERA implies the implementation of four strategic goals [2]:

1. Prioritize investments and reforms in R&D to support digital transformation.
2. Improve access to research and innovation for researchers.
3. Ensure the development of research products by the market and the competitive leadership of Europe in the field of technology.
4. Make progress in the free circulation of knowledge, researchers, and technologies by strengthening cooperation.

One can assume that the specified ERA tasks combine key landmarks to which the scientific community should be directed in any state. Constructing information spaces of scientific activity subjects is at the heart of solving problems that would stimulate investment in research and innovation, strengthen cooperation between universities, improve the efficiency and productivity of research activities in general.

2. Literature review and problem statement

The first component in the stage of constructing a method to build information spaces of scientific activity subjects is the analysis of networks of scientific cooperation and citation networks, building which makes it possible to calculate important characteristics for scientific activity subjects that could underlie determining their identifiers. The main tasks that are addressed within the framework of the analysis of scientific networks are the examination of the principles underlying the creation and functioning of networks of scientific cooperation [3] and citation networks [4]. The tasks of analyzing topics [5], construction of scientific networks based on common keywords are also solved. Paper [6] found that the research into cooperation networks should be performed at the micro level since links at the macro level do not properly reflect the peculiarities of cooperation between the scientific activity subjects. Work [7] reported the first attempts to use the analysis of scientific networks of cooperation to reflect the microstructure of interaction between scientists. Study [8] investigates the development of a network of scientific cooperation in the dynamics in three aspects: current topological characteristics, forecasting two-regime scaling analytically and based on numerical simulations. The cited study shows that internal links play a crucial role in determining the scaling behavior and network topology. Works [9, 10] describe the practical results from investigating the structure of networks of scientific cooperation in the field of scientometrics.

A powerful tool for scientometric evaluation of the performance of individual scientific activity subjects is the network of publications citations. Paper [11] describes the use of a link ranking method to evaluate the performance of scientific activity based on the construction of a citation network.

A second component is the formation of subject scientific spaces that make it possible to group scientific activity subjects based on a common area. Paper [12] describes a method to form subject scientific spaces based on clustering the citation graph. For this task, work [13] chose a method of n-gram analysis to establish the similarity between the abstracts of scientific publications, which makes it possible to obtain a smaller number of clusters.

Assessment of the productivity of scientific activity subjects is a third component in the formation of information spaces. Evaluation results make it possible to identify subjects, rank them taking into consideration the area-specific scientific space. Works [14, 15] describe a multicriterial method for assessing the performance of scientific activity subjects, which may be the basis for building their identifiers. Paper [16] describes the use of network services for information and analytical support of scientific and pedagogical research. The approach to an independent evaluation of the quality of electronic publications and publication activity of scientific activity subjects through the analysis of the values of scientometric indicators of open electronic bibliometric systems is also given. The authors conclude that scientometric bases, which are now international and widely used, have a series of significant limitations. To ensure a more adequate representation of the nature and degree of scientific activity by scientists, their personal contribution to the advancement of certain branches of science, further development of the technical component of scientometric bases and methodological evaluation base is required. The results of the application of the method that forms the information spaces of scientific activity subjects could be the basis for finding a solution to the multi-criteria problem of choosing potential scientific partners for cooperation [17, 18], the development of competent principles for building scientific partnerships [19, 20], the introduction of innovative scientific projects [21, 22], etc.

The above works addressed some aspects of the stages of formation of information spaces of scientific activity subjects, which are focused primarily on the state of subjects at a particular point in time. The considered methods do not fully reflect the development of productivity of scientific activity subjects and require a systematic approach to investigating their connections. That requires the study of changes in the state of the subjects in the dynamics, taking into account the results of performance evaluation in the previous time, etc.

3. The aim and objectives of the study

The purpose of this research is to devise a method to form the information spaces of scientific activity subjects in order to ensure the steady evolution of higher education establishments.

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

– to describe the basic conceptual apparatus underlying the formation of information spaces of scientific activity subjects;

– to build multiple identification models and a conceptual scheme of interaction of individual and collective scientific activity subjects;

– to determine stages in the construction of information spaces of scientific activity subjects.

4. The study materials and methods

Ensuring the steady evolution of higher education establishments depends on the understanding of the principles of formation and interaction of information spaces of scientific activity subjects, which is part of these institutions of higher education.

To obtain results on the construction of set models of identification of collective and individual scientific activity subjects, we used systems analysis and a theory of sets. Systems analysis makes it possible to decompose a collective scientific activity subject and describe a set of its parameters, as well as interaction among them, which contribute to its steady evolution. A theory of sets is used to describe models of identification of scientific activity subjects, making it possible to formalize key indicators that are important for the steady development of relevant subjects.

Network theory was applied to describe the network of citations and the network of scientific cooperation. Elements of cluster analysis were used to build subject-specific scientific spaces. For cluster analysis taking into consideration additional parameters of the results of scientific activity of subjects, we employed methods of probabilistic latent semantic analysis and n-gram analysis.

The hypothesis of this study is that the development of individual scientific activity subjects exerts an impact on the steady development of collective scientific activity subjects.

5. Set models of identification of scientific activity subjects and a method that forms information spaces

5. 1. Basic conceptual apparatus underlying the formation of information spaces of scientific activity subjects

Space is a set of ordered objects or entities, often with a multilevel structure, whose place in this structure is determined on the basis of their identifiers.

Information space is a space whose components are information objects that are organized into a multilevel structure with identifiers, which are determined according to certain circumstances and rules and are the result of coordinated activities of the society.

The result of scientific activity as an intellectual and creative process is to acquire new fundamental and/or applied knowledge. This process involves a series of mandatory or situational actions. Namely: analysis and generalization of facts, generation of hypotheses and new concepts, movement of scientific knowledge, based on known facts and concepts, in the direction determined by the research hypothesis, etc. Given the global nature of the scientific community, scientific activity cannot be considered through the prism of one scientist. It should be understood that the process of scientific activity is, for the most part, collective, it is actively or indirectly influenced from the outside by a global network of scientific cooperation. Moreover, this influence has feedback.

Thus, the organization of scientific activities is a process aimed at forming, restoring, or improving these links to improve efficiency. Scientific activity is impossible without an educational one. These are two complementary processes that could be ensured effectively only in those organizational systems that combine, according to certain rules a team, motivated to coordinate to achieve the goal of people. Accumulated results of this activity in the form of information objects form an informational educational and scientific space.

Information space of scientific activity subjects is an information space whose components are information objects that are organized into a certain structure with identifiers. Identifiers are determined by the results of educational and scientific activities of certain subjects of this activity. That is, the information space of scientific activity subjects includes scientific activity subjects and a retrospective representation of the identifiers of these subjects, determined by the results of their scientific activity. Information spaces of various scientific activity subjects could intersect and be segmented when included in subject scientific spaces.

Scientific activity subjects to be considered here are the higher education establishments and their structural subunits that are directly engaged in educational and scientific activities. The sub-units include institutes, faculties, departments, scientific and teaching staff of these units, groups of scientists, united by the implementation of some projects, as well as individual scientists. One of the mandatory components of the activities of such subjects is scientific activity. In this context, there is a task to quantify the output of this activity. Individual scientific activity subjects are scientists. Collective scientific activity subjects are the scientific organizations and higher education establishments, which are formed from individual scientific activity subjects. Each scientific activity subject is represented by a model of identification and information space.

The productivity of scientific activity subjects is a quantitative indicator of the effectiveness of scientific activity subjects. In the process of calculating this indicator, both devised and known indexes of scientific activity could be used, depending on the needs of calculation and its application. The productivity of scientific activity subjects is often considered as the ratio of labor results relative to the time of fixation of scientific indicators or in comparison with other scientific activity subjects. Therefore, to calculate this indicator, it is necessary to clearly explore the information space of the scientific activity subjects, its connections with other information spaces. This could be done by separating a part of the information space that is responsible for a certain subject area of research.

Subject-specific scientific space is a set of scientific activity subjects and parts of their information spaces, which are ordered and united according to the criterion of the joint direction of research activities. At the heart of this space are the scientific activity subject and its part of information spaces with corresponding identifiers.

The organizational and functional structure of scientific activity subjects as an organization is a set of links that establish the ordering, regulation, coordination of scientific activities of subjects to achieve the goals, in particular, improving the effectiveness of scientific research and its potential.

5. 2. Set models of identification of scientific activity subjects

Let $A=\{a_1, a_2, ..., a_n\}$ be a set of scientific activity subjects or scientists who are active representatives of the educational and scientific space, that is, they are engaged in active publication activities, n is the number of scientists. Let $Q=\{q_1, q_2, ..., q_m\}$ be a set of scientific publications published by scientists from set A, m is the number of publications.

Let the set of all pairs between the elements of the sets A and Q be assigned

$$A\times Q=\{(a,q) \mid a\in A\wedge q\in Q\} \quad (1)$$

and the binary relationship $U\subset A\times Q$, which defines the authorship of publications $q\in Q$. In addition, we determine a binary relationship that defines the citation of all publications $q\in Q$,

$$Q\times Q=\{(q_i,q_j) \mid q_i,q_j\in Q, i\neq j\}. \quad (2)$$

A set of all publications by a scientist $a\in A$ is denoted as

$$Q(a_i,t)=\{q_j\in Q \mid (a_i,q_j)\in U, t\in T\},\quad i=\overline{1,n},\quad j=\overline{1,m}. \quad (3)$$

A set of scientific publications cited by each scientific activity subject at the time t is denoted through $\overline{C}(a_i,t)$, the set of publications in which the publications of the scientific activity subject are cited at the time t – through $C(a_i)$, thus:

$$\overline{C}(a_i,t)=\begin{Bmatrix} q_j\in Q \mid (q_y,q_j)\in C, \\ q_y\in Q(a_i), \\ t\in T, y=\overline{1,m} \end{Bmatrix},\quad i=\overline{1,n},\quad j=\overline{1,m}. \quad (4)$$

$$C(a_i,t)=\begin{Bmatrix} q_j\in Q \mid (q_j,q_y)\in C, \\ q_y\in Q(a_i), \\ t\in T, y=\overline{1,m} \end{Bmatrix},\quad i=\overline{1,n},\quad j=\overline{1,m}. \quad (5)$$

For each publication $q\in Q$, determine the set of its authors

$$A(q_j)=\{a_i\in A \mid (a_i,q_j)\in U\},\quad i=\overline{1,n},\quad j=\overline{1,m}. \quad (6)$$

as well as a set of scientific publications quoted by a given publication $q\in Q$ at the time t – $\overline{C}(q_j)$, and a set of publications quoting the publication $q\in Q$ at the time t – $C(q)$:

$$\overline{C}(q_j,t)=\begin{Bmatrix} q_j\in Q \mid (q_j,q_y)\in C, \\ t\in T, y=\overline{1,m} \end{Bmatrix},\quad j=\overline{1,m}, \quad (7)$$

$$C(q_j,t)=\begin{Bmatrix} q_j\in Q \mid (q_y,q_j)\in C, \\ t\in T, y=\overline{1,m} \end{Bmatrix},\quad j=\overline{1,m}. \quad (8)$$

Each publication $q\in Q$ is matched with its abstract $q^A\in Q^A$.

Scalar assessment of the productivity of a scientific activity subject is a certain functional representation Φ:

$$\Phi: A\rightarrow R, \quad (9)$$

where R is the set of real numbers.

The individual scientific activity subject at the time t is represented by a set of identifiers that determine it:

$$\aleph^I(a_i,t)=\left\langle \begin{matrix} Q(a_i,t),\overline{C}(a_i,t),C(a_i,t), \\ Q^A(a_i,t),\Phi(a_i,t),M(a_i,t) \end{matrix} \right\rangle, \quad (10)$$

where $Q(a_i,t)$) is the set of publications by an individual scientific activity subject a_i at the time t;

$\overline{C}(a_i,t)$ – the set of scientific publications cited by the scientific activity subject a_i at the time t;

$C(a_i,t)$ – the set of publications that quote the publications by the scientific activity subject a_i at the time t;

$Q^A(a_i,t)$ – the set of abstracts of the publications by the scientific activity subject a_i at the time t;

$\Phi(a_i,t)$ – the evaluation of the performance of the scientific activity subject a_i at the time t, for example, the Hirsch index;

$M(a_i,t)$ – the number of international projects in which the scientific activity subject a_i participates at the time t.

In addition, the model of identification of an individual scientific activity subject could include the impact factor of his/her publications and other parameters.

The collective scientific activity subject is represented by a set of identifiers that determine it:

$$\aleph^C(a_i,t)=\left\langle \begin{matrix} \Phi_1(a_i,t),\Phi_2(a_i,t),\Phi_3(a_i,t), \\ \Phi_4(a_i,t),\Phi_5(a_i,t) \end{matrix} \right\rangle, \quad (11)$$

where $\Phi_1(a_i,t)$ is the normalized assessment of the international activity by the scientific activity subject a_i at the time t, for example, the number of internships of employees of a collective scientific activity subject abroad, the number of projects with foreign funding, etc.;

$\Phi_2(a_i, t)$ is the normalized assessment of the cohort of higher education applicants of the collective scientific activity subject, a_i, at the time t, if the subject renders educational services, for example, as a higher education establishment (HEE). The assessment of HEE should take into consideration the educational component, the cohort of students: the average EIT score of those enrolled, passing minimum, etc., which are important estimates that determine the potential of HEE;

$\Phi_3(a_i, t)$ is the normalized assessment of the scientific or scientific-pedagogical composition of the collective scientific activity subject a_i at the time t; it can be an average quantitative assessment of the productivity of employees over a certain period of time;

$\Phi_4(a_i,t)$ is the normalized quantitative estimation of research activity of the collective scientific activity subject at the time t;

$\Phi_5(a_i,t)$ is the normalized assessment of available resource support at the time t, in particular the material and technical support of the collective scientific activity subject.

These identifiers are not limiting when setting a scientific activity subject. However, knowing the main components of the calculated parameters, it is possible to calculate the tasks that would improve the effectiveness of the organization of scientific activities in the information educational and scientific space of Ukraine.

Identifiers of the scientific activity subject at the current time t_x and retrospective values at the moments of time $t_{x-\Delta}$, $\Delta=\overline{0,x}$, form the information space of a scientific activity subject. In this case, the values of identifiers are cumulative. The representation of identifiers of individual scientific activity subjects a_i is given by a multilevel time series:

$$\left(\aleph^I\left(a_i,t_0\right),\aleph^I\left(a_i,t_1\right),\ldots,\aleph^I\left(a_i,t_{x-1}\right),\aleph^I\left(a_i,t_x\right)\right). \quad (12)$$

The representation of values of identifiers of collective scientific activity subjects a_i is given by a multilevel time series:

$$\left(\aleph^C\left(a_i,t_0\right),\aleph^C\left(a_i,t_1\right),\ldots,\aleph^C\left(a_i,t_{x-1}\right),\aleph^C\left(a_i,t_x\right)\right). \quad (13)$$

The potential of the development of scientific activity subjects is determined by the rate of change in the indicators of identifiers of these subjects, which is calculated as a percentage comparison of the current value of each identifier with the value of one of the previous periods. For collective scientific activity subjects:

$$S^C\left(a_i,t_x,t_b\right)=\frac{1}{5}\sum_{j=1}^{5}\frac{\Phi_j\left(a_i,t_x\right)-\Phi_j\left(a_i,t_{x-b}\right)}{\Phi_j\left(a_i,t_{x-b}\right)}\cdot 100\,\%, \quad (14)$$

where $S^C(a_i, t_x, t_b)$ is the rate of change in the indicators of identifiers of collective scientific activity subjects a_i at the time t_x relative to the time t_b, that is, the potential of collective scientific activity subjects a_i.

Similarly, for individual scientific activity subjects, a rate of change in the indicators is determined from the following formulas:

$$Q^N\left(a_i,t_x\right)=\frac{\left|Q\left(a_i,t_x\right)\right|-\min\limits_{j=\overline{1,n}}\left(\left|Q\left(a_j,t_x\right)\right|\right)}{\max\limits_{j=\overline{1,n}}\left(\left|Q\left(a_j,t_x\right)\right|\right)-\min\limits_{j=\overline{1,n}}\left(\left|Q\left(a_j,t_x\right)\right|\right)}, \quad (15)$$

$$C^N\left(a_i,t_x\right)=\frac{\left|C\left(a_i,t_x\right)\right|-\min\limits_{j=\overline{1,n}}\left(\left|C\left(a_j,t_x\right)\right|\right)}{\max\limits_{j=\overline{1,n}}\left(\left|C\left(a_j,t_x\right)\right|\right)-\min\limits_{j=\overline{1,n}}\left(\left|C\left(a_j,t_x\right)\right|\right)}, \quad (16)$$

$$M^N\left(a_i,t_x\right)=\frac{\left|M\left(a_i,t_x\right)\right|-\min\limits_{j=\overline{1,n}}\left(\left|M\left(a_j,t_x\right)\right|\right)}{\max\limits_{j=\overline{1,n}}\left(\left|M\left(a_j,t_x\right)\right|\right)-\min\limits_{j=\overline{1,n}}\left(\left|M\left(a_j,t_x\right)\right|\right)}, \quad (17)$$

$$\Phi^N\left(a_i,t_x\right)=\frac{\Phi\left(a_i,t_x\right)-\min\limits_{j=\overline{1,n}}\left(\Phi\left(a_j,t_x\right)\right)}{\max\limits_{j=\overline{1,n}}\left(\Phi\left(a_j,t_x\right)\right)-\min\limits_{j=\overline{1,n}}\left(\Phi\left(a_j,t_x\right)\right)}, \quad (18)$$

$$S^I\left(a_i,t_x,t_b\right)=$$
$$=\frac{1}{4}\cdot\left(\begin{array}{l}\dfrac{Q^N\left(a_i,t_x\right)-Q^N\left(a_i,t_{x-b}\right)}{Q^N\left(a_i,t_{x-b}\right)}+\\+\dfrac{C^N\left(a_i,t_x\right)-C^N\left(a_i,t_{x-b}\right)}{C^N\left(a_i,t_{x-b}\right)}+\\+\dfrac{M^N\left(a_i,t_x\right)-M^N\left(a_i,t_{x-b}\right)}{M^N\left(a_i,t_{x-b}\right)}+\\+\dfrac{\Phi^N\left(a_i,t_x\right)-\Phi^N\left(a_i,t_{x-b}\right)}{\Phi^N\left(a_i,t_{x-b}\right)}\end{array}\right)\cdot 100\,\%, \quad (19)$$

where $S^I(a_i, t_x, t_b)$ is the rate of change in the normalized indicators of identifiers of individual scientific activity subjects a_i at the time t_x relative to the time t_b, that is, the potential of individual scientific activity subjects a_i.

Let $A^I \subset A$ be a set of individual scientific activity subjects, $A^C \subset A$ – a set of collective scientific activity subjects. Let $\in \subset A^C$, where Ω is a set that includes the individual scientific activity subjects and which determines the collective scientific activity subject a_Ω, then

$$\exists\lambda\in\Omega,\ \ \forall t_y\in T,\ \ S^I\left(\lambda,t_y,t_b\right)>0\Rightarrow S^C\left(a_\Omega,t_y,t_b\right)\geq 0, \quad (20)$$

that is, there are such individual scientific activity subjects whose potential is positive. It follows from meeting the condition (20) that the potential of a collective scientific activity subject or organization, including these individual subjects, would have a non-negative potential. This is equivalent to the fact that the rate of change in the indicators of identifiers of the collective scientific activity subject would be positive.

5. 3. Stages in the construction of information spaces of scientific activity subjects

Thus, the development of individual scientific activity subjects causes the development of collective scientific activity subjects or organizations. Moreover, parts of the information spaces of scientific activity subjects intersect with the information spaces of other scientific activity subjects, for example, joint publications in which both subjects are co-authors.

Scientific activity subjects and parts of their information spaces are included in subject scientific spaces that unite subjects that work in a common scientific area. In the case of the creation of joint projects, the information spaces of individual scientific activity subjects interact, forming a new quality, which is a catalyst for the development of collective scientific activity subjects, thereby increasing their own development potential (Fig. 1).

Systematically managed development is the key to the steady development of collective scientific activity subjects: scientific organizations and higher education establishments. To build an information space of the scientific activity subject, it is necessary to perform a series of stages:

1. Calculate all identifiers for the scientific activity subject at the current time. If possible, select identifiers that were registered previously.
2. Build citation networks and networks of scientific cooperation.
3. Build subject-specific scientific spaces. This is necessary to understand in which direction or directions of scientific activity the subjects work.
4. Devise methods of quantitative evaluation of the productivity of scientific activity subjects. This is necessary to understand the potential of scientific activity subjects. The results of applying the methods form the components of the information space.

When creating scientific projects and in the formation of the organizational and functional structure of collective scientific activity subjects, the built information spaces are a key source of information for the use of appropriate models and methods of multicritical choice.

Our study examines in more detail the specified stages. The first stage is the formation of an identification model of the scientific activity subject with the necessary set of parameters used for calculations. The construction of citation networks and scientific cooperation makes it possible to determine the place of the scientific activity subject within the educational and scientific space. A separate task is to visualize such networks [23, 24].

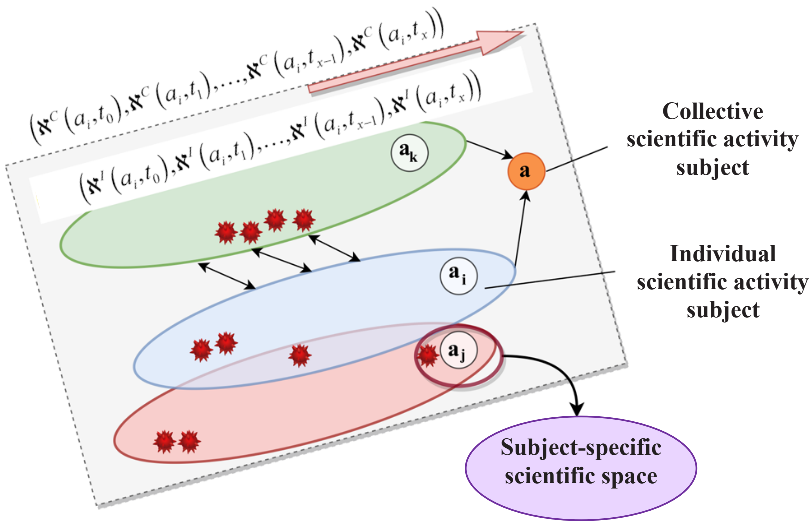

Fig. 1. Conceptual scheme of interaction among scientific activity subjects

The next step is the identification of subject scientific spaces. This can be done on the basis of clustering the scientific publications based on the proximity of abstracts using the method of locally sensitive hashing [25], by a citation graph, or using latent-semantic analysis [26]. As a result, a set of publication clusters would be obtained on the basis of which subject scientific spaces are built.

The next stage is the calculation of performance assessments of scientific activity subjects. If we assume that productivity and subject scientific spaces dynamically change over time, then it is possible to consider this change in the form of discrete-time series and calculate the nature of productivity development in the future. To evaluate the productivity of scientific activity, a method has been proposed that makes it possible to take into consideration all the citations of publications by scientific activity subjects and, at the same time, takes into consideration self-citation and cross-citation. One could also use a comprehensive method for assessing the productivity of scientific activity subjects. The result is quantitative assessments of the productivity of scientific activity subjects.

Built spaces could be used to solve the problem of choosing scientific activity subjects for the implementation of joint scientific and educational projects. This could be implemented on the basis of a multicriterial selection problem and with fuzzy output. The result is a list of partners that meet the requests by a decision-maker.

The solved problems lead to the task of formation of the organizational and functional structure of scientific activity subjects, which directly affects the change of identifiers for a higher education establishment or its structural sub-units as a scientific activity subject. As a result, there are changes in the values of identifiers of the scientific activity subject. Features of the formation of the organizational structure of higher education establishments, in particular for the task of digitalization of higher education, are described in works [27, 28].

6. Discussion of research results on the formation of information spaces of scientific activity subjects

An individual scientific activity subject is represented by a set of identifiers, which include the sets of publications, quotations, abstracts, performance assessments, and international projects in which this subject participates. To form a representation of values for the identifiers of individual scientific activity subjects and information space of the scientific activity subject, it is necessary to include their retrospective values. That imposes a certain restriction on the use of the method because at the moment there is no single source from which it is possible to obtain up-to-date data for any individual scientific activity subject at any time.

Data on publications are partially given in international scientometric databases such as Scopus, WoS, etc. In the profiles of scientists at ORCID, Publons, etc. Data on citation are given in these scientometric databases, as well as communities of publishers of academic publications such as CrossRef. Abstracts of scientific publications are much more accessible than full texts and are hosted by the websites of scientific publications. Data on international projects in which a scientific activity subject participates could be obtained from reports posted, in particular, on the National Erasmus+Office sites.

Accordingly, centralized data collection in a single system, their systematization, and comparison with an individual scientific activity subject is a prerequisite for the use of the method of forming information spaces of individual scientific activity subjects.

To form a representation of values and information space of the subject of collective scientific activity subjects, it is necessary to evaluate international activity, a cohort of applicants, scientific or scientific-pedagogical staff, research activities, and resource support. In addition to the data sources discussed above, additional data sources should also be considered. To build an information space of a scientific activity subject, it is necessary to understand in which direction or directions of scientific activity the subjects work. That is, it is necessary to build subject scientific spaces and evaluate subjects within these spaces. Methods of construction of such spaces and methods of estimation of productivity of research activity should be adapted to the task of forming information spaces of scientific activity subjects. In this context, the task of investigating the method of formation of information spaces of scientific activity subjects has yet to be explored in detail.

To verify the described method, collective scientific activity subjects at the department level were selected from Astana IT University (Nur-Sultan, Republic of Kazakhstan), Taras Shevchenko National University of Kyiv (Kyiv, Ukraine), State Higher Educational Institution "Uzhhorod National University" (Uzhhorod, Ukraine). The number of individual scientific activity subjects that are part of the collective subjects ranged from 5 to 15. The rate of change in the normalized indicators of identifiers of individual and collective scientific activity subjects for the period from January 2019 to December 2020 was calculated.

The method of formation of information spaces of scientific activity subjects was applied to the scientific activity of the collective subject: the Department of Information Systems and Technologies at the Taras Shevchenko National University of Kyiv. This collective entity consists of 11 full-time employees – individual subjects, for each of whom the rate of change in the identifiers was calculated. To determine the identifiers of subjects, the data from the scientometric base Scopus were taken as a basis. The indicators were calculated as of 1.01.2020 and 1.01.2021. The rate of change in the indicators of scientific activity by individual subjects is given in Table 1. The values are calculated from formulas (15) to (19). According to formula (14), the rate of change in the indicators of identifiers for the collective scientific activity subject (Department of Information Systems and Technologies) was calculated $S^C\left(a,t_{2021},t_{2020}\right)=17.8$ %.

Table 1

Calculation of indicators and rate of change in the indicators of scientific activity subjects

No. of individual scientific activity subject	$Q^N\left(a_i,t_x\right)$, formula (15)		$C^N\left(a_i,t_x\right)$, formula (16)		$M^N\left(a_i,t_x\right)$, formula (17)		$\Phi^N\left(a_i,t_x\right)$, formula (18)		$S^I\left(a_i,t_x,t_b\right)$, formula (19)
year	2020	2021	2020	2021	2020	2021	2020	2021	2021
1	1.00	1.00	1.00	1.00	1	1	1.00	1.00	0.00
2	0.95	0.95	1.00	1.00	1	1	1.00	1.00	0.00
3	0.62	0.53	0.55	0.53	0.5	0	0.67	0.57	0.33
4	0.19	0.16	0.02	0.01	0	0	0.11	0.14	0.12
5	0.29	0.11	0.02	0.02	0	0	0.11	0.14	0.05
7	0.10	0.11	0.12	0.12	0.5	1	0.22	0.29	-0.34
8	0.24	0.16	0.07	0.09	0.5	1	0.22	0.29	-0.28
9	0.00	0.00	0.00	0.00	0.5	0	0.00	0.00	0.25
10	0.05	0.00	0.00	0.00	0.5	0	0.00	0.00	0.50
11	0.14	0.16	0.03	0.02	0	0	0.22	0.14	0.17

For subjects No. 3–5, 9–11, the rate of change in the indicators is positive. This ensures the steady development of the collective scientific activity subject (Department of Information Systems and Technologies), for which the rate of change in the identifier indicators is 17.8 %. For other collective scientific activity subjects, the course of calculations is similar. Accordingly, in practice, we tested the fairness of the hypothesis of research and the validity of formula (20).

The advantage of the proposed method is a comprehensive consideration of the development of collective scientific activity subjects in dynamics. In addition, the method shows that not all individual scientific activity subjects evolve at the same rate but, to ensure the steady development of a collective subject, the presence of several subjects with positive dynamics would suffice.

7. Conclusions

1. A conceptual apparatus has been devised for the method to form information spaces of scientific activity subjects, which makes it possible to systematize information about information spaces, qualitatively describe and formalize methods of identification of subject scientific spaces, and evaluate the productivity of scientific activity subjects. Without the use of the reported conceptual apparatus, it was difficult to build a formalized model of information space and a conceptual scheme of interaction of scientific activity subjects.

2. Set models of identification of individual and collective scientific activity subjects have been built. The model of identification of a scientist as a scientific activity subject has been defined, which includes a set of publications by the scientist, a set of scientific publications cited by the scientist, a set of publications that quote the publications by the scientist, a set of abstracts of the scientist's publications, an assessment of the scientist's productivity, etc. The identification model of a higher education establishment as a scientific activity subject has been defined, which includes an assessment of the international activity of the scientific activity subject, the assessment of the cohort of applicants for higher education as a scientific activity subject, the assessment of the scientific and pedagogical staff of the higher education establishment, the assessment of research activities of the higher education establishment, the assessment of available resources, in particular, material-technical support of the higher education establishment. The devised models make it possible to build a formalized model of information space and a conceptual scheme of the interaction of scientific activity subjects. The developed models open up the possibility of further research into the qualitative formation of information spaces of scientific activity subjects and ensuring the steady functioning of higher education establishments.

3. The stages in the formation of information spaces of scientific activity subjects have been determined. The method includes a series of stages such as the identification and construction of citation networks and scientific cooperation, construction of subject-specific scientific spaces, and, based on them, methods of performance evaluation, the results of which form the components of the information space. The fairness of the hypothesis has been experimentally confirmed that the potential of a collective scientific activity subject, including individual subjects, the rate of change of identifiers of whom is positive, would have the non-negative potential. Using an example of the collective scientific activity subject (Department of Information Systems and Technologies at the Taras Shevchenko National University of Kyiv), which includes 11 individual subjects, it has been shown that five subjects have a positive rate of change in the identifier indicators. Accordingly, this explains the positive rate of change in the indicators of identifiers of the collective scientific activity subject, namely: 17.8 %.

Acknowledgment

This work was funded by the Scientific Committee of the Ministry of Education and Science of the Republic of Kazakhstan No. AP08857218 "Information technologies for assessing the scientific activities of universities, research institutes, and their sub-units".

References

1. Communication from the commission to the council, the european parliament, the economic and social committee and the committee of the regions. Towards a European research area (2000). Commission of the European Communities. Available at: https://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2000:0006:FIN:EN:PDF
2. European research area (ERA). Available at: https://ec.europa.eu/info/research-and-innovation/strategy/era_en
3. Glänzel, W. (2012). Bibliometric methods for detecting and analysing emerging research topics. El Profesional de La Informacion, 21 (2), 194–201. doi: https://doi.org/10.3145/epi.2012.mar.11
4. Egghe, L., Rousseau, R. (1993). Evolution of information production processes and its relation to the Lorenz dominance order. Information Processing & Management, 29 (4), 499–513. doi: https://doi.org/10.1016/0306-4573(93)90045-f
5. Lizunov, P., Biloshchytskyi, A., Kuchansky, A., Andrashko, Y., Biloshchytska, S. (2020). The use of probabilistic latent semantic analysis to identify scientific subject spaces and to evaluate the completeness of covering the results of dissertation studies. Eastern-European Journal of Enterprise Technologies, 4 (4 (106)), 21–28. doi: https://doi.org/10.15587/1729-4061.2020.209886
6. Kremen, V., Bykov, V. (2013). Category "space" and "environment": features model submissions and educational application. Teoriya i praktyka upravlinnia sotsialnymy systemamy, 2, 3–16.
7. Otte, E., Rousseau, R. (2002). Social network analysis: a powerful strategy, also for the information sciences. Journal of Information Science, 28 (6), 441–453. doi: https://doi.org/10.1177/016555150202800601
8. Barabási, A. L., Jeong, H., Néda, Z., Ravasz, E., Schubert, A., Vicsek, T. (2002). Evolution of the social network of scientific collaborations. Physica A: Statistical Mechanics and Its Applications, 311 (3-4), 590–614. doi: https://doi.org/10.1016/s0378-4371(02)00736-7
9. Hou, H., Kretschmer, H., Liu, Z. (2008). The structure of scientific collaboration networks in Scientometrics. Scientometrics, 75 (2), 189–202. doi: https://doi.org/10.1007/s11192-007-1771-3
10. Newman, M. E. J. (2001). The structure of scientific collaboration networks. Proceedings of the National Academy of Sciences, 98 (2), 404–409. doi: https://doi.org/10.1073/pnas.98.2.404
11. Biloshchytskyi, A., Kuchansky, A., Andrashko, Yu., Biloshchytska, S., (2020). Use of the link ranking method to evaluate scientific activities of scientific space subjects. Scientific Journal of Astana IT University, 1, 12–20. doi: https://doi.org/10.37943/aitu.2020.1.63600
12. Biloshchytskyi, A., Kuchansky, A., Andrashko, Yu., Biloshchytska, S., Kuzka, O., Shabala, Ye., Lyashchenko, T. (2017). A method for the identification of scientists' research areas based on a cluster analysis of scientific publications. Eastern-European Journal of Enterprise Technologies, 5 (2 (87)), 4–11. doi: https://doi.org/10.15587/1729-4061.2017.112323
13. Lizunov, P., Biloshchytskyi, A., Kuchansky, A., Andrashko, Y., Biloshchytska, S. (2019). Improvement of the method for scientific publications clustering based on n-gram analysis and fuzzy method for selecting research partners. Eastern-European Journal of Enterprise Technologies, 4 (4 (100)), 6–14. doi: https://doi.org/10.15587/1729-4061.2019.175139
14. Kuchansky, A., Andrashko, Yu., Biloshchytskyi, A., Danchenko, E., Ilarionov, O., Vatskel, I., Honcharenko, T. (2018). The method for evaluation of educational environment subjects' performance based on the calculation of volumes of m-simplexes. Eastern-European Journal of Enterprise Technologies, 2 (4 (92)), 15–25. doi: https://doi.org/10.15587/1729-4061.2018.126287
15. Hnatiienko, H., Snytyuk, V., Tmienova, N., Voloshyn, O. (2020). Determining the effectiveness of scientific research of universities staff. CEUR Workshop Proceedings, 2833, 164–176. Available at: http://ceur-ws.org/Vol-2833/Paper_15.pdf
16. Bykov, V. Yu., Spirin, O. M., Soroko, N. V. (2015). Elektronni bibliometrychni systemy yak zasib informatsiyno-analitychnoi pidtrymky naukovo-pedahohichnykh doslidzhen. Informatsiyno-komunikatsiyni tekhnolohiyi v suchasniy osviti: dosvid, problemy, perspektyvy, 1, 91–100.
17. Garcez, M. P., Sbragia, R., Kruglianskas, I. (2014). Factors for selecting partners in innovation projects – qualitative evidences from non-equity bilateral alliances in the Brazilian petrochemical leader. Review of Administration and Innovation - RAI, 11 (2), 241. doi: https://doi.org/10.5773/rai.v11i2.1292
18. Feng, W. D., Chen, J., Zhao, C. J. (2000). Partners selection process and optimization model for virtual corporations based on genetic algorithms. Journal of Tsinghua University (Science and Technology), 40, 120–124.
19. Lukianov, D., Bespanskaya-Paulenka, K., Gogunskii, V., Kolesnikov, O., Moskaliuk, A., Dmitrenko, K. (2017). Development of the markov model of a project as a system of role communications in a team. Eastern-European Journal of Enterprise Technologies, 3 (3 (87)), 21–28. doi: https://doi.org/10.15587/1729-4061.2017.103231
20. Korzh, R., Peleshchyshyn, A., Syerov, Y., Fedushko, S. (2016). University's Information Image as a Result of University Web Communities' Activities. Advances in Intelligent Systems and Computing, 115–127. doi: https://doi.org/10.1007/978-3-319-45991-2_8

21. Kolomiiets, A., Morozov, V. (2021). Investigation of Optimization Models in Decisions Making on Integration of Innovative Projects. Lecture Notes in Computational Intelligence and Decision Making, 51–64. doi: https://doi.org/10.1007/978-3-030-54215-3_4
22. Morozov V., Kalnichenko O., Mezentseva, O. (2020). The method of interaction modeling on basis of deep learning the neural networks in complex IT-projects. International Journal of Computing, 19 (1), 88–96. doi: https://doi.org/10.47839/ijc.19.1.1697
23. Fruchterman, T. M. J., Reingold, E. M. (1991). Graph drawing by force-directed placement. Software: Practice and Experience, 21 (11), 1129–1164. doi: https://doi.org/10.1002/spe.4380211102
24. Hu, Y. (2005). Efficient, high-quality force-directed graph drawing. The Mathematica Journal, 10 (1), 37–71. Available at: http://asus.myds.me:6543/paper/ktall/37%20-%201984%20-%20Efficient,%20High-Quality%20Force-Directed%20Graph%20Drawing.pdf
25. Huang, Q., Feng, J., Zhang, Y., Fang, Q., Ng, W. (2015). Query-aware locality-sensitive hashing for approximate nearest neighbor search. Proceedings of the VLDB Endowment, 9 (1), 1–12. doi: https://doi.org/10.14778/2850469.2850470
26. Altszyler, E., Ribeiro, S., Sigman, M., Fern ndez Slezak, D. (2017). The interpretation of dream meaning: Resolving ambiguity using Latent Semantic Analysis in a small corpus of text. Consciousness and Cognition, 56, 178–187. doi: https://doi.org/10.1016/j.concog.2017.09.004
27. Tymchenko, D., Korogod, N., Novorodovska, T. (2020). Technology transfer office model. Scientific Journal of Astana IT University, 3, 83–90. doi: https://doi.org/10.37943/aitu.2020.73.19.008
28. Kropachev, P., Imanov, M., Borisevich, Y., Dhomane, I. (2020). Information technologies and the future of education in the Republic of Kazakhstan. Scientific Journal of Astana IT University, 1, 30–38. doi: https://doi.org/10.37943/aitu.2020.1.63639