S. D. Shtovba, Dr. Sc. (Eng.), Associate Professor; M. V. Petrychko

IDENTIFICATION OF THE DEGREE OF SIMILARITY OF EDUCATIONAL SPECIALTIES ON THE BASE OF THE ANALYSIS OF THE NAHEQA EXPERTS PROFILES

Academic community continues to discuss the expediency of combining specialties, belonging of various specialties to one branch of knowledge, etc. For substantiation of such solutions, it is important to have valid assessments of the similarity level of the specialties. Such assessments are necessary for correct organization of the competition for distribution of the state-budget places, merging of universities, formation of interdisciplinary educational programs, establishment of the correspondence of the teacher to education components, selection of the reviewers for the expert examination of the dissertations, manuscripts of the papers, grant applications. The paper suggests the method of identification of the similarity level of education specialties according to the register of NAHEQA experts. The method is based on the assumption that the ability of NAHEQA expert to assess the quality of education programs on different specialties is the indicator of the similarity of these specialties. The level of similarity is evaluated by Jaccard index. On the base NAHEQA registers data base of 4245 experts is formed, 2414 of them are experts in one specialty, 1361 are experts in two specialties and 470 are experts in three specialties. Identification of the similarity level of specialties is carried out. Greater part of 7260 pairs of specialties have almost zero similarity. Average level of similarity is determined for 52 pairs of specialties and 21 pair has high level of similarity. It should be noted that 9 out of 21 pairs with high similarity form specialties in different branches of knowledge. Thus, it was established that common branch profile is not identical to real similarity of the specialties. It is shown on the specific examples how to use the obtained results for substantiation of the managerial decisions regarding the launching of inter-disciplinary educational programs, merging specialties, shifting specialties into other branches of knowledge and for the formation of the groups of specialties for wide contest.

***Key words:** identification, management of education activity, similar specialties, branch of knowledge, interdisciplinary education programs, data analysis, Jaccard index, stickiness index, NAHEQA (National Agency for Higher Education Quality Assurance).*

Introduction

Training of the students, undergraduates and post graduates in Ukraine is realized on the base of two-level system of sciences classification. In this system the first hierarchical level is formed by branches and the second – by specialties. Two-level system was initiated in 2015 and since that time underwent four corrections in 2016, 2017, 2021 and 2022. In 2024 new updating of the system of sciences classification is expected. Discussions are kept in academic community regarding the expedience of merging specialties, belonging of different specialties to common branch, etc. For the substantiation of similar decisions, it is important to have valid assessments of the degree of simarity of the specialties. Such assessments are needed for the correct organization of the competitive tender for the distribution of state-funded places. In recent years state-funded places on certain specialties are combined and the contest is conducted within the groups. For instance, in 2023 political science, journalism, law and 3 specialties from the branch of knowledge «29 – international relations» were united in group Б37д , group Б37д unites applied mathematics and 6 specialties from the branch of knowledge «12 – information technologies», group Б38д unites 15 engineering specialties from seven branches of knowledge. Such unions are made subjectively without taking into account the level of real similarity of the specialties. Another example of the decision-making problems of the ministry level, the correct solution of which requires taking into account the similarity of

specialties is the consolidation of the universities. It is logical that it is expedient to consolidate the universities which perform educational activity mainly at similar or related specialties.

Besides ministerial tasks, listed above, of educational activity management, the knowledge of the similarity level of the specialties is necessary for managerial decision making on the university level. The example of such tasks is the formation of interdisciplinary education programs, establishment of the correspondence of the teacher to educational components, selection of the peer-reviewers to perform an expertise of the dissertations, manuscripts of the articles, grant applications, etc.

Objective of the paper is the identification of the current level of the specialties similarity on the base of statistic analysis of the public data.

Literature survey and the idea of the research

In the law «On higher education» the definition of the term «similar specialties» is missing. Word combination "similar specialties" in the law «On higher education» is mentioned only once in the definition of the branch of knowledge: *«branch of knowledge is harmonized with the International standard classification of the education as wide subject area of education and science, that includes the group of the similar specialties».* In accordance with this definition similar specialties must belong to the same branch of knowledge. Similarity must be not only formal but also content-related. The attempt to assess the conceptual similarity of the specialties by means of the analysis of the normative documents was unsuccessful. Key normative documents, regarding the content of the specialties are educational standards. Nowadays they elaborated at the bachelor and master levels almost for all specialties. The content of the specialty in these standards is outlined by the subject area and competences of the specialty. The analysis, carried out, showed that even within the limits of one branch of science specialties, where there are similar special competences or similar fragments of the subject areas are missing. Accordingly, it is impossible to identify the level of similarity of specialties according to education standards.

In scientometrics there exists several approaches to the assessment of the level of interdisciplinarity of scientific research, on the base of which the similarity of specialties can be identified. Most frequently used approach for the similarity identification is citation analysis [1]. However, recently, methods of statistic analysis of the text have been used, in particular, in [2] the authors suggested to use the linguistic approach to study the organization and evolution of scientific branches, used in Web of Science. Three approaches to the determination of the similarity of the specialties are compared in the research: on the base of expert classification, on the citation base, on the base of linguistic similarity. Similarity was proposed to evaluate by the discrepancies matrix (*dissimilarity*): the smaller the value, the greater similarity, the greater the value, the smaller the similarity. Experiments were carried out for the papers from Web of Science, marked by three-level system of science classification. The idea of similarity evaluation by citation is that similarity between the specialties increases if in the paper on one specialty there are references to the paper on other specialty. In case of linguistic similarity, the likeness is determined by comparison of the frequency of the words emergence in the context of specialties. It was determined that the citations-based approaches and linguistic analysis provide similar likeness assessment and greatly differ from the expert assessment of the similarity. Expert assessment gives the idealized notion concerning the similarity of the specialty, at the same time, citation analysis enables to reveal social similarity of specialties and linguistic approach enables to reveal the cognitive similarity. For instance, in [3] the interdisciplinarity level is identified by the number of citations from the paper of certain researcher on the papers in the journals from the

Web of Science, where each of the journals is referred to one of the subject categories of Web of Science. In [4] the identification is performed by means of diversity indices by the references distribution among specialties, taking into account the distances between disciplines. Less common approach is identification of interdisciplinarity by the analysis of the graphs of collaboration over joint projects and co-authorships [5], which is determined by the researchers affiliation to scientific specialties. Stirling diversity index is used as the mathematical model. Some studies use simultaneously citation analysis and linguistic analysis for determining the interdisciplinarity level [6, 7].

All above-mentioned approaches require large information resources and are labor consuming. Besides, the methods based on citation analysis are rather inertial. It is impossible to evaluate immediately interdisciplinarity of new work as it requires some time for its citation. Unlike the mentioned approaches, rapid method of scientific specialties similarity evaluation, based on text analysis is suggested in the article [8]. The method is based on the usage of the services of information resources of Dimensions system, it contains more than 140 mil categorized scientific publications. Each publication in Dimensions system is referred to one or several specialties according to the system of sciences classification ANZSRC-2020. Categorization of the publications in Dimensions is performed on the base of machine learning, mainly on the base of the analysis of the content, names, abstract and key words. Calculation of the specialties similarity in [8] is carried out according to Jaccard-index as a relation of the number of common publications of two specialties to total number of publications by these two specialties.

It is impossible to use directly the approach, described above for the assessment of the cognition of the domestic specialties due to such two reasons. The first reason – lack of the base of scientific publications with the corresponding categorization of each publication to the domestic specialties. The second reason is the lack of the correspondence between the domestic specialties and specialties of ANZSRC-2020, that makes it impossible the transfer of the results from [8] on Ukrainian specialties. At the same time, method [8] can be applied for the assessment of the similarity of Ukrainian specialties not on the base of the categorized publication, but on the base of other information resources. It was revealed that there exists the register of NAHEQA experts, each of the experts is referred to one, two or three specialties. NAHEQA expert has necessary knowledge to assess correctly the quality of education programs of different specialties. The correspondence of NAHEQA expert to several specialties is considered to be the indicator of their similarity. At such assumption, the method from [8], can be used for the identification of the similarity level of the specialties, but instead of the base of scientific publications the base of NAHEQA experts will be used.

Initial data for identification of the similarity of specialties

Nowadays in Ukraine training of higher education applicants is carried out on 28 fields of knowledge. Each field of knowledge contains from 1 to 9 specialties. Totally there are 121 specialty, their list is presented below:

    011 – educational, pedagogical sciences;
    012 – preschool education;
    013 – primary education;
    014 – secondary education (by subject specialties);
    015 – vocational education (by specializations);
    016 – special education;
    017 – physical education and sports;
    021 – audiovisual art and production;
    022 – design;

023 – fine arts, decorative arts, restoration;
024 – choreography;
025 – music art;
026 – dramatic art;
027 – museum studies, monument studies;
028 – management of sociocultural activity;
029 – information, library science and archive studies ;
031 – religious studies ;
032 – history and archeology;
033 – philosophy;
034 – cultural study;
035 – philology;
041 – theology;
051 – economics;
052 – political science;
053 – psychology;
054 – sociology;
061 – journalism;
071 – accounting and taxation;
072 – finances, banking, insurance and equities market;
073 – management;
075 – marketing;
076 – entrepreneurship and trade;
081 – law;
091 – biology and biochemistry;
101 – environmental studies;
102 – chemistry;
103 – earth science;
104 – physics and astronomy;
105 – applied physics and nanomaterials;
106 – geography;
111 – mathematics;
112 – statistics;
113 – applied mathematics;
121 – software engineering;
122 – computer science;
123 – computer engineering;
124 – system analysis;
125 – cybersecurity and information protection;
126 – information systems and technologies;
131 – applied mechanics;
132 – materials science;
133 – industrial engineering;
134 – aeronautic, rocket and space equipment;
135 – shipbuilding;
136 – metallurgy;
141 – electric power industry, electric engineering and electromechanics;
142 – power engineering industry;
143 – nuclear power engineering;

144 – thermal power engineering;
145 – renewable sources of energy and hydropower engineering;
161 – chemical technologies and engineering ;
162 – biotechnologies and bioengineering;
163 – biomedical engineering;
171 – electronics;
172 – electronic communication and radio engineering;
173 – aircraft electronics;
174 – automation, computer-integrated technologies and robot engineering;
175 – information-measuring technologies;
176 – micro- and nano- system engineering;
181 – food technologies;
182 – consumer industry technologies;
183 – environmental technologies;
184 – mining engineering;
185 – oil and gas engineering and technologies;
186 – publishing and printing;
187 – woodworking and furniture technologies;
191 – architecture and urban planning;
192 – construction and civil engineering;
193 – geodesy and land management;
194 – hydraulic engineering construction, aquatic engineering and technologies;
201 – agricultural science;
202 – protection and quarantine of plants;
203 – gardening, vegetable and vine growing;
204 – technology of live stock products processing;
205 – forestry;
206 – park and garden economy;
207 – aquatic bioresources and aquaculture;
208 – agroengineering;
211 – veterinary medicine;
221 – dentistry;
222 – medicine;
223 – nursing;
224 – technologies of medical diagnostics and treatment;
225 – medical psychology;
226 – pharmaceutics, industrial pharmaceutics;
227 – therapy and rehabilitation;
228 – pediatrics;
229 – public health;
231 – social work;
232 – social maintenance;
241 – hotel and catering business;
242 – tourism and recreation management;
251 – state security;
252 – border security;
253 – military administration (by kinds of armed forces);
254 – provision of armed forces;
255 – weapons and military equipment;

256 – national security (by separate spheres of provision and types of activity);
257 – information security management;
261 – fire-safety;
262 – law enforcement activities;
263 – civil security;
271 – maritime transport and inland water transport;
272 – air transport;
273 – railway transport;
274 – automobile transport;
275 – transport technologies (by the type of transport);
281 – public management and administration;
291 – international relations, public communication and regional studies;
292 – мinternational economic relations;
293 – international law.

Title of each specialty consists of digital code and content word-combination. Further for the designation of the specialties its digital code will be used. First two digits in the code correspond to the number of branch and the last digit – to the order number of the specialty within the branch.

Base of experts will be formed according to two public registers of NAHEQA – «Register of the experts among scientific-pedagogical staff, researchers » and «Register of the experts among the higher education students » valid on the 1st of February 2024. After uniting the registers the data will be cleared – all specialties will be reduced to the current list, for instance, specialty 151 will be replaced by 174, 152 – by 175, the duplicates will be eliminated. In case of the discrepancy of the digital code and content title the specialty to the expert will be denoted by the content title in the register. As a result, the formed base comprises 4245 experts, among them 2414 are experts in one specialty, 1361 are experts in two specialties and 470 are experts in three specialties. Distribution of the experts into forty most popular specialties in presented in Fig. 1. There are few experts in specialties 257, 145, 225, 252, 256, 027, 112, 135, 224 – from 0 to five persons, that is why, the validity of the statistical conclusions for these specialties will be low. Number of experts for specialties 143, 173, 187, 021, 194, 203, 207, 026, 134 and 272 will be from 6 to 9, that is why, these statistical conclusions must be used very carefully.
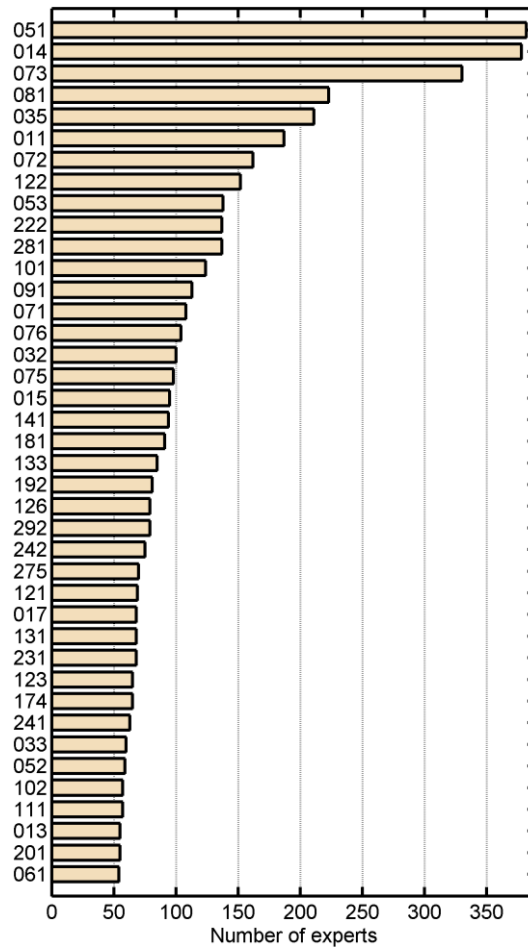
Fig. 1. Fragment of NAHEQA experts distribution according to specialties

Identification of the specialties similarity

Index of pair of specialties similarity (*A*, *B*) is proposed to calculate in the following manner:

$$J(A, B) = \frac{N_{A \cap B}}{N_A + N_B - N_{A \cap B}}, \qquad (1)$$

where N_A – is a number of experts for specialty *A*; N_B - is a number of experts for specialty *B*; $N_{A \cap B}$ - is a number of specialties who are simultaneously experts both on specialty *A*, and specialty *B*.

Formula (1) is equivalent to the similarity formula of the specialties in [8], except that as initial data the base of the categorized experts is used, whereas in [8] – the base of the categorized scientific publications.

The application of the formula (1) will be illustrated by the example of the similarity calculation of specialties 113 and 124. There are 42 experts for specialty 113 in the base, for specialty 124 – 27 experts. Experts simultaneously in both specialties are 6 persons. Accordingly, the similarity level of specialties 113 and 124 is:

$$J(113, 124) = \frac{6}{42 + 27 - 6} = 0.095.$$

Ranking distribution of the pairs of specialties by Jaccard index (1) is presented in Fig. 2. High similarity have 21 pair of specialties, for these specialties Jaccard index exceeds 0.1. For 52 pairs of specialties the similarity is average, Jaccard index for these specialties is in the

range of [0.05, 0.1]. Greater part of 7260 pairs of specialties have almost zero similarity. Top-40 most closely similar pairs of specialties are presented in in Fig3. 16 of these 40 pairs belong to different branches of knowledge. If only specialties with high similarity are considered, then 9 pairs of 21 form specialties in different branches of knowledge. Thus, common profile is not identical with real similarity of the specialties.
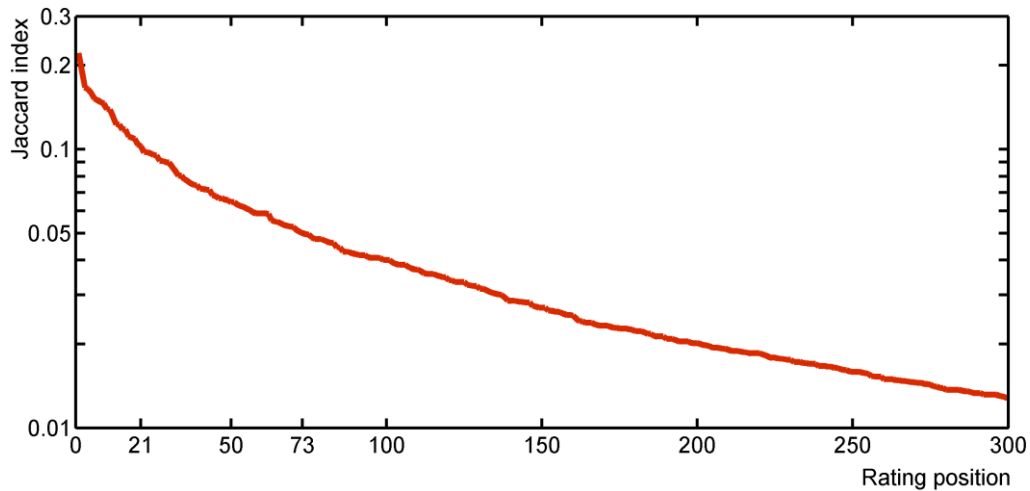

Fig. 2. Fragment of ranking similarity distribution of pair of specialties (semilogarithmic format)
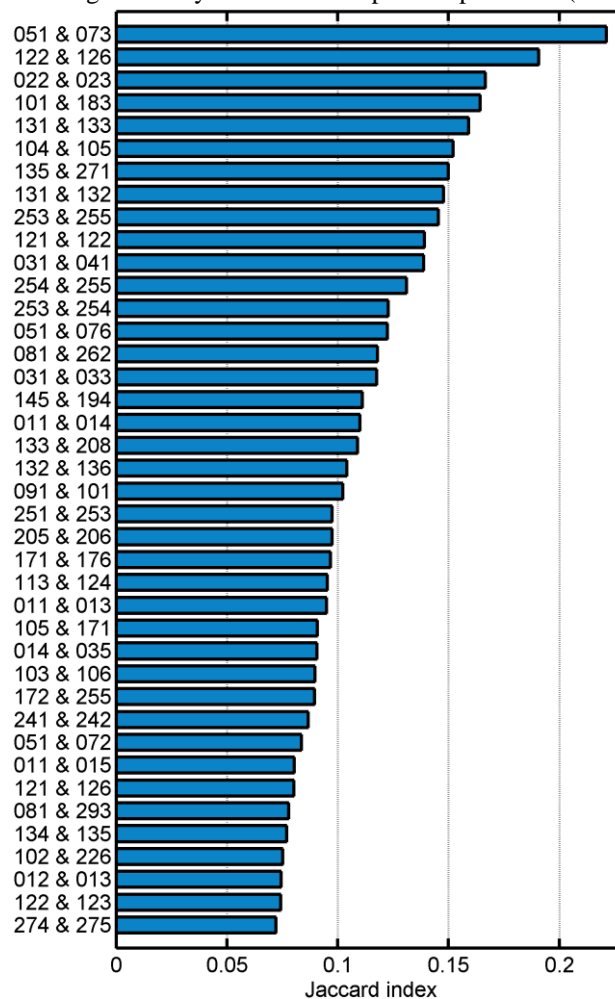

Fig. 3. Most closely similar pairs of specialties

Pairs of specialties with high level of similarity have considerable potential for formation of interdisciplinarity education programs. 21 pair with high level of similarity comprises 32

specialties, in particular, 10 specialties have high similarity with two specialties. As a result, there appears the need to determine specialties with high level of interdisciplinarity interaction on the total space of the specialty. For such objectives in [8] the stickiness index of the specialty is suggested. It is calculated as a sum of all Jaccard indices of the specialty. Top-40 specialties with the largest total level of interdisciplinarity interaction are shown in Fig. 4. The obvious leader, according to interdisciplinarity interaction is specialty 051. High stickiness index have also specialties 014, 122, 011, 101, 073 and 131.
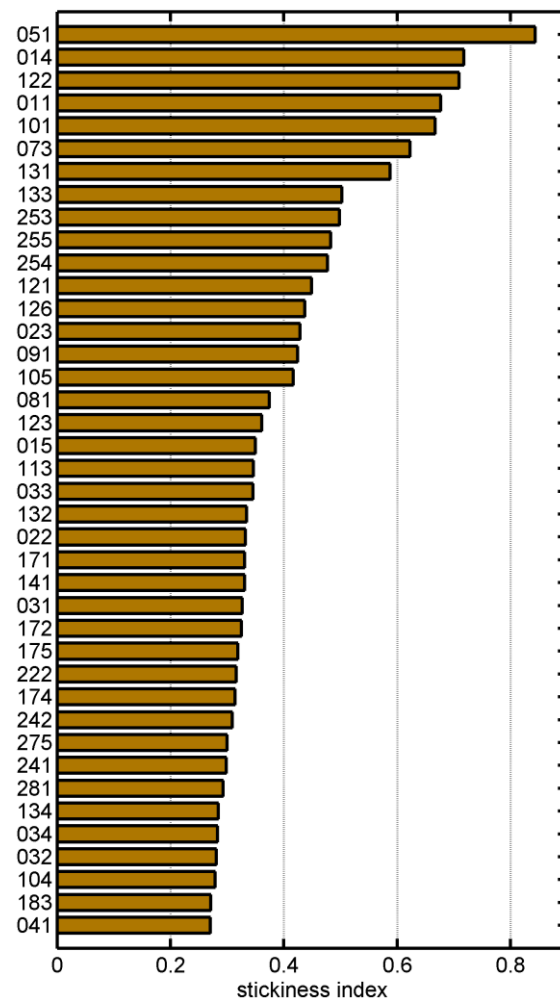

Fig. 4. Top-40 specialties according to stickiness index

Fig. 5 contains similarity diagrams for 15 specialties with maximum stickiness indices. It follows from this fig that 12 out of 15 specialties have high level of similarity with the specialties of their branch. Specialties 051, 091 and 101 are most similar with other specialties. Regarding the specialty 091, it is single in its branch, that is why, such state of things is obvious. Regarding specialty 051, besides this specialty three specialties belong to the branch 05. But the similarity within the limits of the branch for specialty 051 is too low – maximum value of Jaccard index is 0.007 for the pair (051, 054). Similarity of specialty 051 with 14 specialties of other branches exceeds 0.007. Fig. 5 shows that 5 out of 7 most similar with 051 specialties belong to branch 07. This can serve as an argument regarding the transfer of the specialty 051 in the branch 07. Concerning specialty 101 it is closely related to specialties 183 and 091. This can serve as an argument for uniting specialties 183 and 101 in one or for the formation of the common branch of knowledge for specialties 091, 101 and 183.
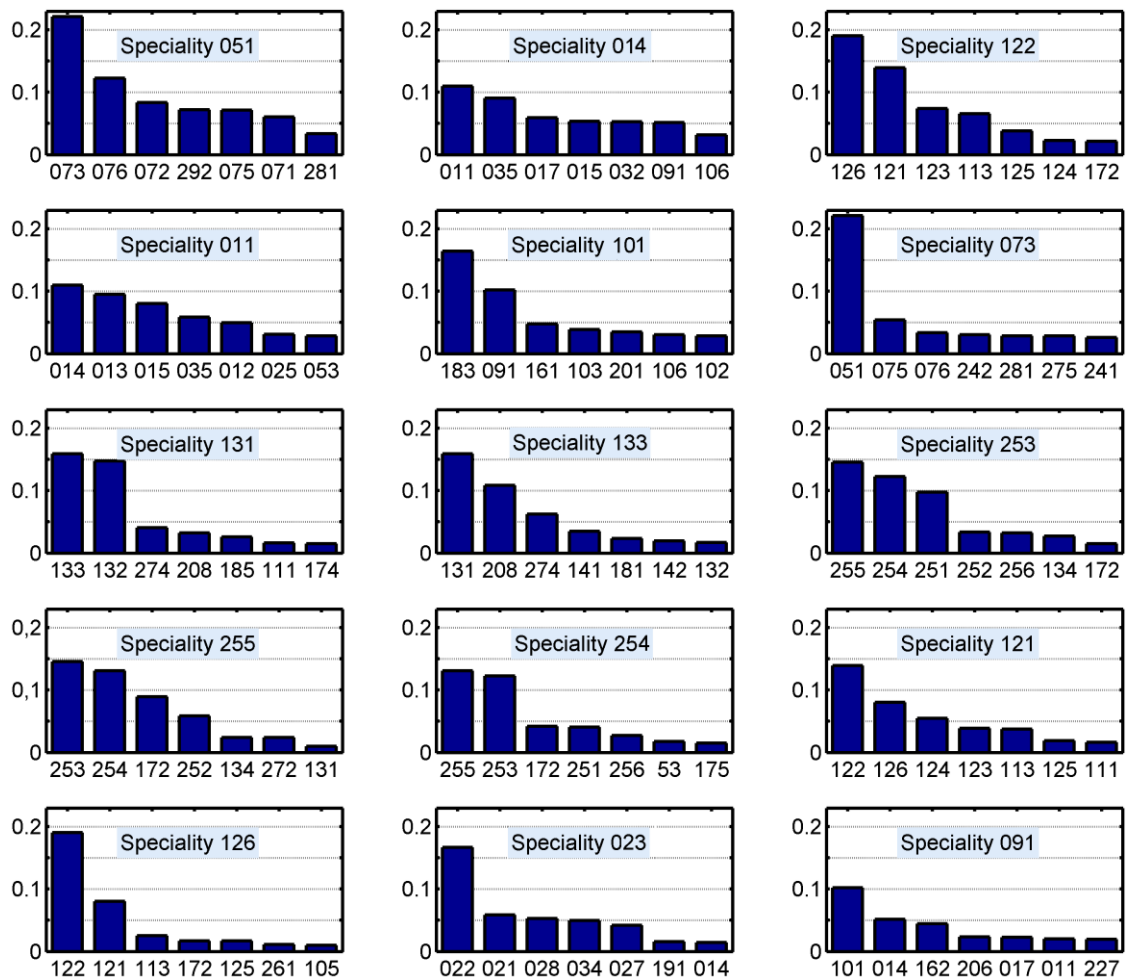
Fig. 5. Similarity diagrams of specialties from top-15 by the stickiness index

Fig. 6 contains the similarity diagrams for 6 specialties, branch belonging of which provokes numerous discussions. It follows from the diagram that specialty 017 may either remain in its current branch 01, or be united with the specialty 227 or be the part of the branch 22. It is expedient to transfer specialty 041 to the branch 03, and, probably, unite with the specialty 031, taking into account their very high similarity. Specialty 106 must remain in the branch 10, there are no reasons to transfer it to branch 05. It is worth mentioning, that specialties 106 and 242 have high potential for the formation of the interdisciplinarity education programs. Specialty 113 is very similar with the specialties of the branch 12, this may serve as an argument in favor of common wide contest with specialties of the branch Б37д. Regarding the profile, the similarity of the specialty 113 to branch 12 is higher, than to current branch 11. But the similarity of specialties 113 and 111 within the branch 11 is also significant. Specialty 123 has moderate similarity with the specialties both of its branch and with the specialties of branch 17, mainly with the specialty 174. That is why, specialty 123 can be referred both to the branch 12, and branch 17. But its similarity with the specialties of the branch 12 is higher. It is expedient to transfer specialty 174 into branch 12 because the total similarity with the specialties of this branch is the highest. Although, in this case, the arguments are not so strong as concerning the migration of the specialty 041 into the branch 03.
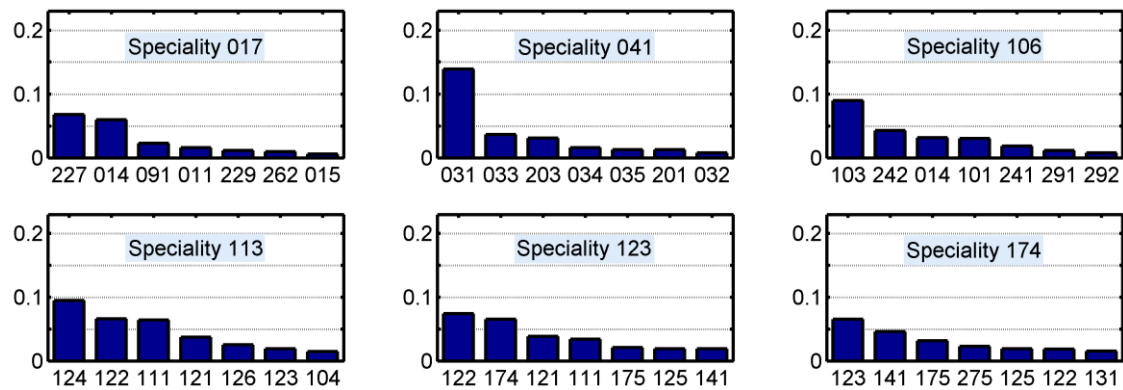
Fig. 6.Similarity diagram of certain problematic specialties

Conclusions

Method of similarity level identification of the educational specialties by the public information – by the register of NAHEQA experts is proposed. The suggested method of identification is based on the assumption that the ability of NAHEQA expert to evaluate the quality of educational programs by various specialties is the indicator of similarity of these specialties. Mathematically, similarity level is evaluated by Jaccard index. The base of experts by means of uniting and clearing data from the registers of NAHEQA is formed. Identification of the similarity level of specialties is performed, on specific examples is shown how to use the obtained results for the substantiation of the managerial decisions regarding the uniting of specialties, specialties placement into other branches of knowledge, formation of the groups of specialties for wide competition as well as for launching the interdisciplinarity education programs. Limitations regarding the usage of the obtained results for specialties where the number of NAHEQA experts is not large, is noted.

REFERENCES

1. Approaches to understanding and measuring interdisciplinary scientific research (IDR): A review of the literature / C. S. Wagner, J. D. Roessner, K. Bobb [et al.] // Journal of Informetrics. – 2011. – Volume 5, Issue 1. – P. 14 – 26. https://doi.org/10.1016/j.joi.2010.06.004.

2. Using text analysis to quantify the similarity and evolution of scientific disciplines [Electronic resource] / L. Dias, M. Gerlach, J. Scharloth, E. G. Altmann // Royal Society Open Science. – 2018. – № 5 (1). – Access mode : https://doi.org/10.1098/rsos.171545.

3. Measuring researcher interdisciplinarity / A. L. Porter, A. S. Cohen, J. D. Roessner [et al.] // Scientometrics. – 2007. – № 72 (1). – P. 117 – 147. https://doi.org/10.1007/s11192-007-1700-5.

4. Van Noorden R. Interdisciplinary research by the numbers. / R. Van Noorden // Nature Publishing Group. – 2015. – № 525. – P. 306 – 307. https://doi.org/10.1038/525306a.

5. Karlovčec M. Interdisciplinarity of scientific fields and its evolution based on graph of project collaboration and co-authoring / M. Karlovčec, D. Mladenić // Scientometrics. – 2015. – Volume 102 (1). – P. 433 – 454. https://doi.org/10.1007/s11192-014-1355-y.

6. Braam R. R. Mapping of science by combined co-citation and word analysis. II: Dynamical aspects / R. R. Braam, H. F. Moed, A. F. J. van Raan // Journal of the American Society for Information Science. – 1991. – №42 (4). – P. 252 – 266. https://doi.org/10.1002/(SICI)1097-4571(199105)42:4<252::AID-ASI2>3.0.CO;2-G.

7. Using network science and text analytics to produce surveys in a scientific topic / F. N. Silva, D. R. Amancio, M. Bardosova, [et al.] // Journal of Informetrics. – 2016. – № 10 (2). – P. 487 – 502. https://doi.org/10.1016/j.joi.2016.03.008.

8. Shtovba S. Jaccard index-based assessing the similarity of research fields in dimensions / S. Shtovba, M. Petrychko // In CEUR Workshop Proceedings. CEUR-WS. – 2019. – Vol. 2533. – P. 117 – 128.

Shtovba Sergiy – Doctor of Science (Eng.), Professor with the Department of Information Technologies, e-mail: s.shtovba@donnu.edu.ua.
Vasyl Stus Donetsk National University.

Petrychko Mykola – Post Graduate with the Department of Computer Control Systems.
Vinnytsia National Technical University.