

Ontological Approach to Modeling the Current Labor Market Needs for Automated Workshop Control in Higher Education

Elena Boldyreva^[0000-0001-6357-4448] and Vadim Kholoshnia^[0000-0001-6485-6340]

ITMO University. 197101, St.Peterburg, Russia
eaboldyreva@itmo.ru
vdkholoshnia@gmail.com

Abstract. This paper is devoted to the question of current labor market needs modeling for automated workshop control in knowledge-intensive areas. On the basis of the ontological model of labor market needs, the content of the workshop is systematically processed for certain profiles in knowledge-intensive areas. An approach to the e-workshop project control based on the pedagogical design model ADDIE is described. The approach is implemented on the basis of an information system that takes into account the regular updating of the technological and tool base in professional areas. The paper proposes a method for forming and modifying the list of main learning tasks - the kernel of the workshop, based on the analysis of labor market needs. This method involves systematic monitoring of vacancies in this profile, processing their descriptions in order to highlight the requirements of employers. The methodology is implemented in the framework of the information system for workshops designing. Vacancy data is obtained from the API hh.ru, lists of the applicants requirements and responsibilities are extracted from their descriptions. The resulting lists of requirements are processed using natural language processing and cluster analysis technologies.

Keywords: workshops design · knowledge-intensive areas · labor market needs · machine learning · natural language processing · unsupervised text clustering

1 Introduction

During the period of high competition in the labor market, the main task of educational institutions is to prepare graduates who are competitive in the labor market. Constant updating of the goals, learning outcomes and content of educational programs becomes not just a requirement of regulations and educational standards, but a necessary condition for ensuring high-quality training of students. At the level of Federal state standards, measures are being taken to

strengthen the influence of employers on the learning process in higher education institutions: involving them in the development of educational standards, creating basic departments at enterprises, and ensuring the practical orientation of learning process [1–4].

Such actions are particularly relevant when it comes to training specialists in knowledge-intensive areas. A feature of knowledge-intensive areas is the constant updating of the technological and tool base, and, as a result, constantly changing the requirements of employers.

This article is devoted to the formation and further systematic revision of the workshop kernel for knowledge-intensive areas. The workshop kernel in the context of this paper is a list of the main educational tasks that correspond to the labor functions (professional tasks that are constantly faced by active specialists in the labor market) and the competencies of the graduate. The method of forming the workshop kernel is proposed on the basis of an information system that constantly monitors and analyzes the requirements of the professional area - professional standards and vacancies descriptions in online-recruitment systems.

2 Overview of existing solutions

At the moment, the direction related to information support for the implementation of the educational process is actively developing. Modeling of the curriculum based on the ontological approach is described in the paper [5]. Graph models and structural approaches to curriculum development are presented in the paper [6]. Examples of works that perform modeling and expert analysis of educational programs are also [7, 8]. However, these works do not imply a direct impact of the professional areas requirements on the educational trajectory formation.

In order to monitor the connection between the needs of employers and the implementation of the educational process, concepts of various intelligent systems are emerging. The following papers describe an automated system for monitoring and analyzing the compliance of personnel needs based on the nomenclature of universities [9], a qualification-oriented expert system for managing the educational process [10], and an intelligent system for supporting the educational programs formation based on neural network models of language, taking into account the professional areas requirements [11]. These expert systems interact with the labor market needs (its normative documents) and suggest making recommendations for modifying the normative documents of the educational process. The paper [11] also describes the interaction with vacancies descriptions, but does not specify how this interaction is carried out, the emphasis is on the analysis of professional standards.

It should be noted that the normative documents of the educational and professional areas are inert in relation to the rapidly changing requirements for the practical part implementation of the educational process. Therefore, the main drawback of systems that focus on the normative documents is that all recommendations are made for educational programs, but no specific recommendations are given for the implementation of these. Such recommendations must neces-

sarily include the tools and technologies updates in the professional areas, if the purpose of training is to train highly qualified specialists of this profile.

3 Approach to the formation of the actual workshop kernel based on the analysis of the labor market needs of knowledge-intensive areas

This article describes an approach to solving this problem, namely the system of automated workshop design and workshop project control [12] based on the professional area needs and the opinions of employers in knowledge-intensive areas. On the basis of the approach, an information system for the electronic workshop design is developed and implements the model of pedagogical design ADDIE [13, 14] at the stages of analysis and design. A characteristic feature of the approach is its application in knowledge-intensive subject areas that have annual update of technologies and tools.

The approach describes the process of creating and modifying the workshop content for the education profile of students. The model of this process is shown in Figure 1.

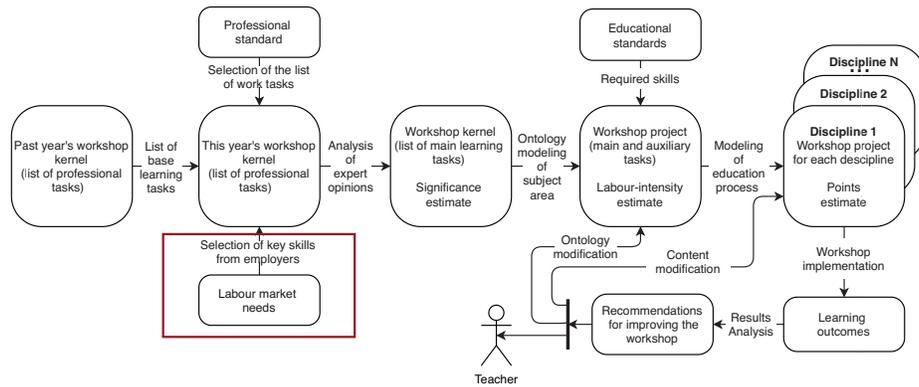


Fig. 1. Process of workshop content creating for a specific educational profile for the current year

The approach to the e-workshop project control for knowledge-intensive areas based on the opinions of employers involves the following steps:

1. Formation the workshop kernel for a current education profile.
 During the first design, the information system forms the the workshop kernel list of the main learning tasks for a specific profile of graduates. The main learning tasks at this stage are the adaptation of labor tasks that the student must be ready to solve in order to be competitive in the labor market. The collection of labor tasks is carried out by analyzing professional standards,

normative documentation of the profile and corresponding vacancies from online-recruitment systems. The obtained data are processed using semantic analysis technologies [15], clustered by subject, and the most popular labor tasks and technologies (tools) that need to be solved are identified. If the workshop design is not the first time, then the list of learning tasks from the previous period can also be used as a basis.

2. Formation of the main learning tasks list based on the analysis of expert opinions.

The resulting list of learning tasks for the workshop kernel is passed to experts to assess their significance and complexity (qualification level). The information system offers learning tasks to the workshop kernel with an assessment of significance above the threshold set by the workshop developer. This stages of interaction with experts is repeated every year to update the list of tasks, assess their significance and technologies for solving them.

3. Formation of the workshop project with calculation of labor-intensity characteristics of learning tasks and division of these tasks into auxiliary ones. The labor-intensity of learning tasks in the workshop project is determined in accordance with the weight coefficient of each learning task. The weight coefficient is calculated using the formula 1

$$W_i = \frac{\sum_{j=1}^m C_i * K_j}{\sum_{i=1}^n \sum_{j=1}^m C_i * K_j} \quad (1)$$

where W_i – the weight coefficient of the i -th learning task, C_i – the expert evaluation of the i -th factor of the learning task in points (from 1 to 10), K_j – confidence level of the j -th expert, n – number of practical tasks of the workshop, m – number of evaluation factors.

According to the value of the weight coefficient, each main learning task is assigned a share of the total labor intensity of the workshop.

4. Formation of disciplines workshops projects of the education profile on the basis of educational process modeling.

At the previous stage, the main learning tasks are also divided into blocks of sub-tasks and analyzed for duplication – if the main learning task i contains one or more identical sub-tasks with the previous (within the educational process) learning task $i - 1$, then these sub-tasks are excluded from the preparatory set i . This allows you to reduce the "preparatory" trajectory of the main educational task without losing the quality of students learning. After that, the learning tasks are arranged in the optimal order for studying and distributed among the disciplines, forming the educational trajectory of the profile.

5. Processing the learning outcomes of the disciplines workshops and making recommendations for modifying the workshop content.

Based on the learning outcomes, the information system generates recommendations for modifying the electronic workshop content based on the results of its passage. According to the described approach, the implementation of the main learning tasks and sub-tasks is carried out using data

obtained at the stages of formation and expert evaluation of the workshop kernel. Technologies and tools selected according to the labor market needs and recommended by experts are included in the educational process. Since we are talking about training specialists in knowledge-intensive areas where popular technologies and tools are constantly changing, constant monitoring of the labor market and an annual survey of specialists allow us to timely update the workshop content implementation. In this work, we have implemented an approach to ontological modeling of labor market needs based on the results of systematic monitoring of these online-recruitment systems.

3.1 Modeling the current labor market needs within a specific educational profile

In order to clearly define the format, purpose of the extracted data and how to analyze it, it was decided to describe the ontological model of the professional area for the specific educational profile. Ontology in the context of this paper will be understood as a specific knowledge base [16–18] used for solving learning tasks.

$$O_{PA} = (E_{PA}, R_{PA}), \quad (2)$$

where E_{PA} is a set of entities of the professional area, R_{PA} is a set of relationships between entities in a professional area.

The set of $E_{PA} = V, L, S, Req$ entities is divided into four main subsets:

- sub-set V (vacancies) - a subset of professions described by professional standards (PS) and vacancies. Each profession is characterized by appropriate professional standards and a specific set of vacancies;
- sub-set L (labor tasks) entities that describe labor tasks from professional standards or from vacancies descriptions;
- sub-set S (skills) entities includes knowledge and skill requirements. These skills are described in the PS in relation to each labor task. The student must master these skills in order to be a sought-after specialist;
- sub-set Req (requirements) - this set includes the employers requirements - technologies and tools knowledge, or so-called key skills from the vacancies description that the applicant must have in order to solve labor tasks. For modeling knowledge-intensive areas this subset is key in the formation of actual content for practical learning of students in the educational profile.

Figure 2 shows a generalized model for matching professional standards entities and vacancies description entities in online-recruitment systems.

Figure 3 shows an ontological model of the professional area of a certain educational profile, formed on the basis of an analysis of the labor market needs and the competencies of the graduate.

These models show the main data that should be processed and compared in the process of forming the actual workshop kernel of a certain educational profile - labor tasks and skills of professional standards, requirements, labor tasks and key skills from employers.

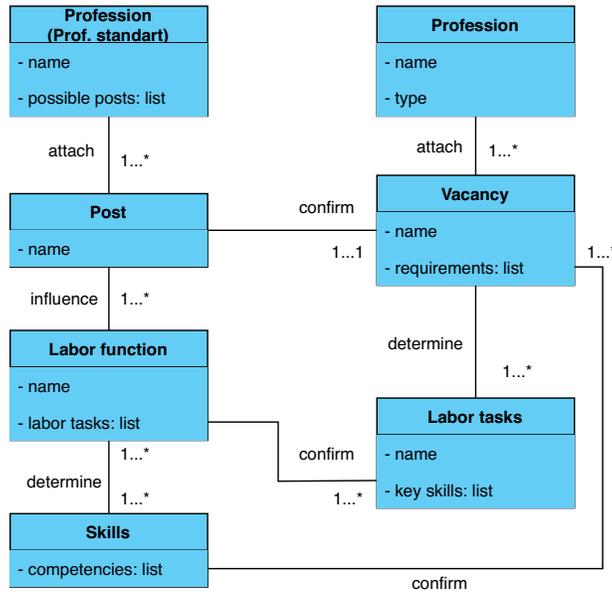


Fig. 2. Generalized model for matching professional standards entities and vacancies description entities in online-recruitment systems

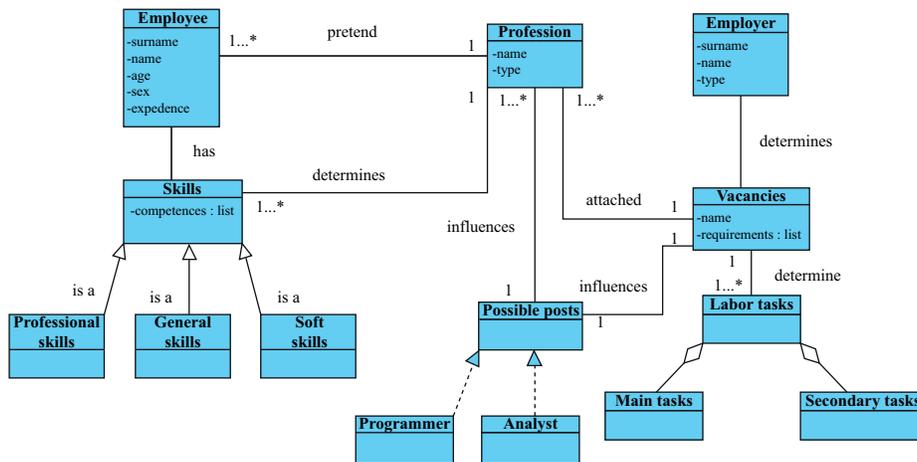


Fig. 3. Ontological model of the professional area of a specific educational profile, formed on the labor market needs and the competencies of the graduate

3.2 Method for forming the electronic workshop kernel based on the analysis of current labor market needs

The proposed method is based on semantic analysis of current requirements of employers for specific projects (tasks) in knowledge-intensive areas. Based on the

results of this analysis, the main labor tasks that a graduate of the profile must be able to solve in order to be competitive in the labor market are highlighted. The collection of labor tasks is carried out using the analysis of vacancies corresponding to the profile from online-recruitment systems. The required skills and responsibilities of applicants are highlighted from the vacancies description. The data obtained are processed using semantic analysis technologies, clustered by topic. The most popular labor tasks and technologies (tools) that need to be solved are highlighted. If the workshop design is not the first iteration, then the list of learning tasks from the previous learning period can also be used as a basis.

This approach assumes the following steps:

1. determining the current labor market needs based on the analysis of online-recruitment systems;
2. formation (addition) of an ontological model of the professional area to identify new trends in the labor market;
3. formation of the actual workshop kernel and recommendations for the implementation of the workshop content in this educational profile in accordance with the professional area needs.

3.3 Determining the current labor market needs based on the analysis of online-recruitment systems

Professional standards are basic normative documents that define the legally fixed requirements of the professional area. The main drawback of such documents is their inertia – the term for changing such documents can be 5-10 years. In addition, professional standards reflect skills, actions, and functions in general, without delving into specific tools and technologies. However, the demand of the labor market for specialists of a certain profile and the requirements for the technological and tool base change faster than professional standards. In order to track these changes in a timely manner, you need to contact online-recruitment systems.

The texts of vacancies descriptions in such online-recruitment systems have a similar structure – each vacancy contains a block describing the duties and requirements for the applicant. Depending on the employer, the difference may be that the data may be indicated by different words, such as "Requirements" "Required" "It is necessary to...". The scheme of the algorithm for extracting and processing information from the vacancies description is shown in Figure 4.

As part of the information system, the analysis of vacancies received using the services API hh.ru is currently implemented. First, the workshop developer selects the required professional profile. The system sends a request to the API and receives a list of vacancies for this profile. Since this list offers vacancies descriptions in an abbreviated format, vacancies identification number is extracted from this list. Next, the algorithm sends a request to get a full description of each of the vacancies by its ID. Data is provided in json format, and descriptions

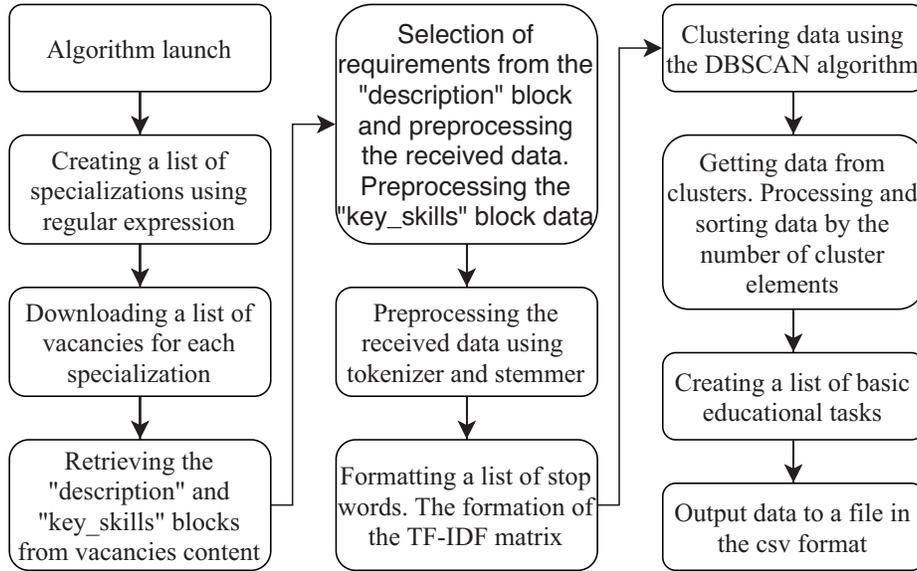


Fig. 4. Algorithm for extracting and processing information from vacancies descriptions

are contained in the *description* block. A parser was implemented in Python, extracts the "responsibilities" and "requirements" blocks and makes a list of them for further processing. In addition, the description contains the *key_skills* section, which describes the tools and technologies that the applicant must possess for the position.

At the processing stage, punctuation marks (commas, dots, and semicolons) were removed. Abbreviations like "SW" were replaced with "software". Normalization was carried out by bringing words to their initial form. In this task, we decided to use stemmer, because due to the specific terminology of vacancies descriptions and the specifics of the Russian language, the exclusion of endings does not distort the meaning of the phrase. This system implements the Porter stemmer [19]. At the output, we get an array of normalized tokens for employers' requirements.

Next, the TF-IDF weight matrix is created. We consider each selected request record as a separate document. The matrix is compiled using `TfidfVectorizer` (sklearn package), the block of stop words is taken from the *ntlk* case and supplemented in accordance with the specifics of the professional area. DBSCAN, MiniBatchKMeans, and agglomerative clustering algorithms were applied to the resulting matrix.

The choice of the DBSCAN algorithm is due to the fact that with this spatial clustering algorithm, you don't need to specify the number of clusters in advance - it is difficult to predict the number of thematic clusters in the variety of employers' requirements. In addition, the list of requirements often contains requirements

of the type "well understand what it is about". These records may be considered noise and are not included in any of the clusters. Clusters of the following types are highlighted: corresponding to the subject area and noise requirements (the latter are removed from the list of requirements). Table 1 shows examples of such clusters.

Table 1. Cluster’s content examples (DBSCAN)

| Cluster type | Cluster content |
|--------------|--|
| Noise | [user support, cost management, PC repair, responsibility, competent speech', ...] |
| Cluster 1 | [ms sql development experience from 2 years, experience with industrial databases: oracle ms sql postgresql, experience with ms sql server from 1 year, experience with ms sql server (distributed databases), practical experience with ms sql server 2008/2016, experience with ms sql server profiler, experience with ms sql server / oracle database, ...] |

When selecting the clustering algorithms MiniBatchKMeans and agglomerative clustering, you must specify the number of clusters. It is based on the number of main labor tasks from the professional standards for this profile, which has been doubled (since about 50 % of the requirements in the vacancy are insignificant for the formation of this profiles workshop). This increase allows you to take into account " noise" and allocate additional clusters for storing such records.

3.4 Formation of an ontological model of the professional area

At this stage, cluster data is converted to ontological entities and described using OWL. An example of an intelligence map of an ontological model fragment for Cluster1 from Table 1 is shown in Figure 5.

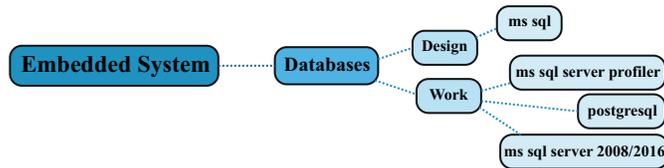


Fig. 5. Fragment of the ontological model mind map for Cluster1

The generated ontological model is offered to the developer for analysis and selection of the final formulation of the main learning tasks and tools and technologies for their solution. This set of learning tasks will be submitted for further evaluation and modification to a group of experts from employers.

4 Practical application

The proposed approach and implemented tools were used to form the workshop kernel on the profile “Embedded and cyber-physical systems”. As a basic data set for forming a kernel fragment dedicated to embedded system software, requirements and key skills for software - vacancies of 6 different specializations were taken and analyzed - 11183 vacancies and 121698 requirements. The total execution time of the algorithm with data loading and processing was 31.5 minutes.

Visualization of the DBSCAN and MiniBatchKMeans clustering algorithms is shown in Figure 6 and 7.

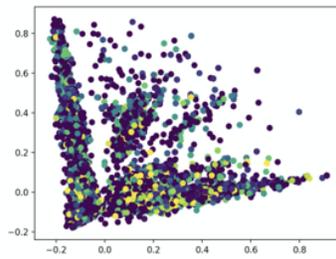


Fig. 6. Visualization of the DBSCAN clustering algorithm

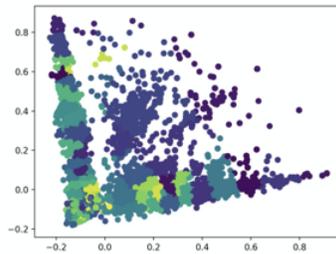


Fig. 7. Visualizing the operation of the MiniBatchKMeans clustering algorithm

The DBSCAN algorithm showed the best results when setting the distance between the eps clusters=0.001. With these parameters, about 270 clusters are allocated that contain data that is significant for this profile. Using agglomerative clustering 100 significant clusters and 96 clusters of the MiniBatchKMeans algorithm were identified.

Further, the clusters were correlated with the labor tasks of professional standards, and they were chosen as close as possible to the formulations of labor tasks. Since it is difficult to assess the semantic proximity of texts in Russian, a “manual” assessment of the semantic proximity of the selected clusters and the main labor tasks was carried out. According to the results of this assessment, the DBSCAN clustering algorithm showed better results - the selected clusters were more “thematic” and contained fewer insignificant requirements.

5 Conclusion

The article describes an approach to creating the electronic workshop content of a certain education profile for knowledge-intensive areas. These areas have annually strong changes in the issues of technologies used and requirements for the skills of industry specialists. An ontological approach to modeling the actual labor market needs based on a systematic analysis of online-recruitment systems is proposed.

Ontological model of the professional area have been developed that take into account: the relationship between professional standards and labor market requirements, and the competence model of the graduate. This model show the main data that must be processed and compared in the process of forming the actual workshop kernel for a specific educational profile - labor tasks and key skills from employers. As part of the paper, the analysis of vacancies obtained using the services API hh.ru was implemented. For the profile “Embedded and cyber-physical systems” 11183 vacancies and 121698 requirements were analyzed for the workshop kernel segment of the embedded systems software. According to the results of the expert evaluation, the DBSCAN clustering algorithm showed better results - the selected clusters were more “thematic” and contained fewer insignificant requirements.

An ontological model of the professional area of the profile was developed based on the results of the analysis of vacancies descriptions, and a list of current main learning tasks and tools for solving them was defined in the kernel block of the discipline workshop “Embedded systems software” for transfer to experts-employers for evaluation.

Based on the described approach, a software package for extracting and processing current labor market needs from vacancies descriptions of online-recruitment systems has been developed, and a mechanism for forming and supplementing an ontological model of the professional area has been implemented.

As a direction for further research, it was decided to compare the cluster analysis of vacancies descriptions and classify them using neural network models.

References

1. Mann, Anthony & Archer, Louise. (2014). Understanding employer engagement in education: theories and evidence.
2. Tomlinson, Michael. (2018). Employers and Universities: Conceptual Dimensions, Research Evidence and Implications. *Higher Education Policy*. 10.1057/s41307-018-0121-9.
3. Elias, K. L. (2014). Employer perceptions of co-curricular engagement and the co-curricular record in the hiring process. https://tspace.library.utoronto.ca/bitstream/1807/67968/1/Elias_Kimberly_L_201411_MA_thesis.pdf. Last accessed 02 Jan 2020.
4. Balganova E. V., Bogdan N. N. assessment by employers of competencies of future specialists in the field of personnel management as a basis for improving the educational process. *Professional education in the modern world*. 2016. Vol. 6. No. 2. Pp. 290-296. DOI: 15372/PEMW20160214.
5. Chung H., Kim J. An Ontological Approach for Semantic Modeling of Curriculum and Syllabus in Higher Education. *International Journal of Information and Education Technology*, 2016, vol. 6 (5), pp. 365–369. DOI: 10.7763/IJJET.2016.V6.715.
6. Lisitsyna L.S., Pirskeya A.S. [Automation of Management of Educational Trajectories for the Development of Modular Competence-Oriented Educational Programs of the University]. *Sbornik trudov Vserossiyskoj nauchno-prakticheskoy konferentsii s mezhdunarodnym uchastiem. Informatsionnye tekhnologii v obespechenii novogo kachestva vysshego obrazovaniya*. [In Proceedings of the All-Russian Scientific-Practical Conference with International Participation “Information Technology in Providing a New Quality of Higher Education”]. Moscow, 2010. pp. 75–86. (in Russ.)
7. Kharitonov I.M. [The Study Plan Forming Algorithm Based on the Study Discipline Formalized Presentation Procedure (by the Example of “System Simulation” Discipline)]. *Bulletin of Astrakhan State Technical University. Series: Management, Computer Science and Informatics*, 2011, no. 2, pp. 178–185. (in Russ.)
8. Sibikina I.V., Kvyatkovskaya I.Y. [Construction of Linguistic Scales with the Purpose of Revelation of Important Disciplines Developing the Competence]. *Bulletin of Astrakhan State Technical University. Series: Management, Computer Science and Informatics*, 2012, no. 2, pp. 182–186. (in Russ.)
9. Zrelov P.V., Korenkov V.V., Kutovskiy N.A., Petrosyan A.S., Rumyantsev B.D., Semenov R.N., Filozova I.A. [Automated System for Monitoring and Analysis of Compliance of the Labour Resources Needs According the Specialties’ Nomenclature of Higher Educational Institution]. *Federalism*, 2016, no.4 (84), pp. 63–76. (in Russ.)
10. Stain D.A., Verbitskaya N.O., Kalugina T.G. [Qualification-Oriented Expert System of Management of Educational Process of Higher Education in Modern Processes of Continuing Qualification Development of Personnel in Russia]. *Bulletin of the South Ural State University. Ser. Education. Educational Sciences*. 2018, vol. 10, no. 1, pp. 27–36. DOI: 10.14529/ped180104 (in Russ.)
11. Botov D.S. Intelligent Support Development of Educational Programs Based on the Neural Language Models Taking into Account of the Labor Market Requirements. *Bulletin of the South Ural State University. Ser. Computer Technologies, Automatic Control, Radio Electronics*, 2019, vol. 19, no. 1, pp. 5–19. DOI: 10.14529/ctcr190101 (in Russ.)
12. Boldyreva E. A. Approach to the automation of design processes for a workshop based on the views of employers // *Bulletin of the Astrakhan State Technical University. Series: Management, Computing and Informatics*. 2020. No. 1. P. 94–104. DOI: 10.24143 / 2072-9502-2020-1-94-104. (in Russ.)

13. Chookaew, Sasithorn Howimanporn, Suppachai Sootkaneung, Warin Pradubsri, Yoothai. (2014). Computer assisted learning based on ADDIE instructional development model for visual impaired students. Proceedings of the 22nd International Conference on Computers in Education, ICCE 2014.
14. Ngussa, B. M. (2014). Application of ADDIE Model of Instruction in Teaching-Learning Transaction among Teachers of Mara Conference Adventist Secondary Schools , Tanzania. *Journal of Education and Practice*, 5(25), 1–11.
15. Sivogolovko, E., Thalheim, B. Semantic approach to cluster validity notion. // *Advances in Databases and Information Systems / Ed. by Tadeusz Morzy, Theo Harder, Robert Wrembel. — Springer Berlin Heidelberg, 2012 — Vol. 186 — P. 229–239.*
16. Ward, J., and A. Aurum. 2004. Knowledge management in software engineering – Describing the process , 137–146., 15th Australian Software Engineering Conference (ASWEC 2004) Melbourne, Australia: IEEE Computer Society Press.
17. Van de Ven, A.H. 2005. Running in packs to develop knowledge-intensive technologies. *MIS Quarterly* 29(2): 365–378.
18. Guide to the Software Engineering Body of Knowledge. Version 3.0 (SWEBOK.v3). A Project of the IEEE Computer Society <http://www.computer.org/portal/web/swebok>. Last accessed 02 Jan 2020.
19. Willett P. The Porter stemming algorithm: then and now // *Program: Electronic Library and Information Systems. — 2006. — Vol. 40, no. 3. — Pp. 219-223.*