Advances in Intelligent Systems Research, volume 174
Proceedings of the 8th Scientific Conference on Information Technologies for Intelligent Decision Making Support (ITIDS 2020)

Knowledge Identification by Structured Data for Decision Making in Project Teams

Nafisa Yusupova*
Ufa State Aviation Technical University
Ufa, Russia
yussupova@ugatu.ac.ru

Olga Smetanina
Ufa State Aviation Technical University
Ufa, Russia
smoljushka@mail.ru

Ekaterina Sazonova
Ufa State Aviation Technical University
Ufa, Russia
rassadnikova_ekaterina@mail.ru

Aygul Agadullina
Ufa State Aviation Technical University
Ufa, Russia
aygul.agadullina@gmail.com

***Abstract*— The methodology of information support for decision-making in the formation of the project team was presented in this article. The object of the research is the process of the project team's formation before the start of the project. The subject of research is a set of methods and means of organizing information support for decision-making in the formation of a project team. Aim of this research work is organization of information support for decision-making in the process of forming a project team to improve the quality of decisions. The proposed methodology includes collection and preparation of data for analysis, identification of new knowledge based on similarity of objects using clustering, their integration with expert knowledge, formalization of knowledge and formation of the knowledge base, obtaining solutions while making use of knowledge and inference engine. Application of tools for data mining, namely the analytical platform Deductor Studio, are shown. The results of experimental studies based on the proposed method are provided.**

Keywords—Data mining, production rules, Kohonen maps/Self-organizing map, decision support

I. INTRODUCTION

Any corporate information system (CIS) is used by enterprises for several years and a large amount of data is accumulated for this time. The data correspond to the CIS functionality. There is to talk about information in the context of the tasks being solved in data processing. Increasing amounts of data and the use of modern analysis technologies can provide the solution new problems and reveal new useful knowledge for decision-making.

Such knowledge can be called tacit (non-obvious). The subsequent formalization of knowledge and their reduction to explicit knowledge is feasible. Thus, the DIKW pyramid (data, information, knowledge, wisdom) can be built. The initial data and semantics of the problem area are fundamental when building a model.

The issues of identification of knowledge based on structured data were considered in both theoretical aspects and practical implementation in the works of M. Polany [1], L.A. Amaeva[2] , E.V. Zubkov [3] , V.M. Belov [3], M. E. Pavlova [4], T. G. Guryeva [4], V. S. Abrukov [4], V. A. Duke [5], A.V. Flegontov [5] , I. K. Fomina [5] .

There are many definitions of tacit knowledge. So, the author [1] noted "tacit knowledge as knowledge that cannot easily be transferred to others". In studies, the following issues are considered: tacit knowledge in the cognitive characteristics of the information model is defined [6]; tacit knowledge is also considered as either "not visible", or "not perceived", or "not interpreted" by other people, or any combination of these [7]. The authors of this article are closer to the concept of V. Ya. Tsvetkov [7]. The issues of tacit knowledge are discussed in the works of I. M. Neisky [8], A. Yu. Fillipovich [8], V. Ya. Tsvetkov [9,10], E. A. Emelyanova, R. B. [11], Bolbakov [7], V. V. Volchik, E. V. Maslyukova [12] and etc [13-22].

In addition, issues of obtaining implicit knowledge based on Data Mining are discussed in the works of Evsyukov V.V. [23], Asci F.H. [24], Wei C., Xiaodong L., Lihong M., Liang C. [25], Yu L. [26], Bi Y., Jia L., Wang, QB [27], Li M. [28] et al.

Formalized data is used for decision making. Many specialists, both in Russia and abroad, in particular, S.N. Vasilyev [29], A.N. Raikov [30], L.J., Gressgard, T. Nesheim [31], J.L. Medina Moya, B. Jarauta Borrasc, J. Menegaz [32], et al., research issues of knowledge formalization. However, the interpretation of knowledge requires an analysis of problem areas, and, consequently, additional research.

In this article authors propose to develop the DIKW model for decision-making in forming teams. The first two levels of the model are represented by data from corporate information systems. Also they are formed by these systems on the basis of data and are depended on the functionality of systems.

The production rules system was developed by authors of this article. The consequent of the production rules is represented by recommendations for decision-making in the formation of a project team; the antecedent is represented by a cluster number and boundary values of one or more characteristics. A cluster is determined by performing cluster analysis. A neural network apparatus is used for this analysis.

II. STATEMENT OF THE PROBLEM

The value of data and knowledge is constantly growing for organizations, this allows them to gain competitive advantages. New knowledges are generated from data and semantics. Each step up the DIKW pyramid [33] raises the status of the initial data, enriching the data with sense and context.

Consider as an example the company, which is a form of activity can be defined as a project. One of the primary tasks of functional leaders with project activities is the effective management of human resources.

It is expertly determined that when organizing information support for human resource management, it is necessary to fulfill the following points: the uniqueness of the project (that is select the necessary characteristics of personnel, the level of personnel experience and the uniqueness of knowledge); time constraints (that is clear planning of staff workload, direction and distribution of staff workload in terms of the effectiveness of the existing levels of knowledge, skills and abilities); resource constraint (that is effective use of the existing composition and distribution of resources so that there are no "overlaps", namely, one employee worked on several projects); the required quality of results (this means that it is necessary to identify risks in the field of personnel, to regulate in time and take measures to limit the impact on the project). The fishbone (or Ishikawa) diagram developed to identify factors which affect the effectiveness of the project (that is successful completion GM> 60%). The successful completion of the project is influenced by each individual person involved in the project, the nature of the team, services, characteristics of the manager, management tools. Table 1 contains the description of these "root" causes and factors describing the nature of the collected data.

The selection is made of the data available in the databases of functioning corporate information systems (in particular, SAP R/3) and ready-made reports (information) based on identified factors for subsequent analysis (Table 2).

The analysis will allow to evaluate:

- compliance of the quantitative composition of the teams with the qualitative composition, and how command structure correlates with the goals and requirements of the project;
- the effectiveness of group work of a particular team;
- compatibility of team members;
- determination of motivation and incentive to work.

III. FORMAL STATEMENT OF THE PROBLEM

Given: D is the data set, P is the set of tasks to be solved, F is the set of data conversion and processing functions, R is the set of obtained solutions (data processing results), K^E is expert knowledge of the problem area and methods of solving problems, I is the set of information.

Thus, in order to build a DIKW model for decision making in the formation of teams, it is necessary to find and determine: K is the set of knowledge for decision making in the formation of teams, *Met* is the set of processing, analysis and inference methods, P' is the set of tasks to be solved for knowledge and their interpretations, *Mod* is knowledge representation model, K'^E is expert knowledge of the problem area.

The purpose of the generalized decision-making task is to identify knowledge about objects that are similar in their characteristics and that are of greater importance to efficiency. To determine the similarity of objects, a particular problem of clustering is solved after assessing the quality of the data, conducting correlation and factor analysis

IV. DATA PROCESSING AND EXPERIMENT

To solve the problem of constructing a DIKW model for decision-making in the formation of teams, a hypothesis was put forward that the project's effectiveness, which is defined by the concept of Gross Margin, is influenced by the following factors: client, partner and project manager, group, type of services; contract value, revenue, time spent on the project (total and by job position), quantitative and qualitative composition of the team, amounts allocated to other functions, hourly rates for grades; the effect of currency rate fluctuations, GSR (the sum of hours total of grades on hourly rates), implementation rate, average project cost per hour.

The collection of the above data was carried out by uploading various reports from SAP R/3 and consolidating them into a pivot table (Table 3). To identify knowledge, it is proposed to use the Deductor platform.

At the stage of revealing implicit knowledge, many data processing and analysis methods are defined. A preliminary assessment of the quality of the data is carried out.

Depending on the recommendations, the "editing of outliers and extreme values" was carried out (for each of the factors that fell into the preprocessing, we selected methods for processing extreme values and outliers,"filling in the missing data". Since the number of omissions was insignificant, it was decided to delete these recordings.

After the manipulations with the data set, all factors became suitable for analysis.

The next step for screening factors is a correlation and factor analysis.

According to the results of the correlation analysis, we can conclude that the factors "Implementation rate,%", "Average cost, $ / hour", "TotalFees" (amounts given to other functions) have the greatest influence on the output variable. As a result of the factor analysis, the number of variables was reduced from 16 to 9. The system suggests leaving the factors "Partner", "Manager", "Group", "Type of service", "Forex" and "Indirect costs" as they are combine the others into groups. So "Client" and "Revenue" are combined into one factor. The "Number of hours", "Direct costs", "GSR" - in another, "Amount of the contract" and "Payment to others" - in the third. Such a combination is quite logical and easily explained by the mathematics embedded in the economics of the company's projects.

TABLE I. CHARACTERIZATION OF "ROOT" CAUSES AND FACTORS

Root Cause	Factor	Factor Description
Person	Training	Share of time spent on self-development, corporate trainings, and other training courses
	Health	The proportion of time during which a person was on sick leave
	Experience	The period of work in the company
	«Recycling»	Time share - hours per year allocated to the time of work in projects (in general)
Command	Composition	Grade (job) composition
	Problems	It can be represented by expert evaluation
	Cohesion	Cohesion level team, determined on the basis of criteria and indicators
	Team Relations	Relations level in the team, determined on the basis of criteria and indicators
Services	Kind	Code
	Type	Code
	Volume	Predicted value / report data
	Cost prediction.	Predicted value / report data
Means	Hardware, software, organization methods, technologies	Possible "downtime" due to failures
Manager	Leadership qualities, management methods, analysis and forecasting skills, delegation, and distribution of tasks.	Personal qualities assessment based on various methods

TABLE II. COMPLETE LIST OF VARIOUS CHARACTERISTICS

Objects	Complete list of various characteristics
Client	- customer name; - location country; - location region; - location city; - sector of activity; - type of client; - local leader (the person who brought or works with this client in the company); - customer risk level; - presence of subsidiaries / quantity
Employees	- surname, name and code of team member; - country / region in which the employee works; - office / department in which the employee works; - industrial group; - gender; - date of entry to work; - employee position / status display of promotion / demotion; - employee's grade; - the number of hours spent on the project, - the number of hours outside the project (vacation, sick leave, time off); - the number of hours spent on training; - work period (dates) in the project; - the number of work days or hours on weekends and holidays; - work types performed on projects; - employee utilization (this is the ratio of the number of hours spent on projects to standard hours); - travel expenses for the project.
Team	- the number of each grade people; - the number of hours spent by each grade; - the sum of actual costs of each grade; - the number of those attracted from other offices / groups.
Project	- project code; - name of project; - country of the project; - region of the project; - project office; - industrial group; - type of work; - start / end date of work; - the amount of the contract; - actually costs (multiplication of time and cost); - additional expenses (business trips, courier services, etc.); - proceeds from the project; - information about invoices for the project (date of issue, amount, date of payment, debt); - the amounts for other offices or groups; - gross margin of the project at the time of closing.

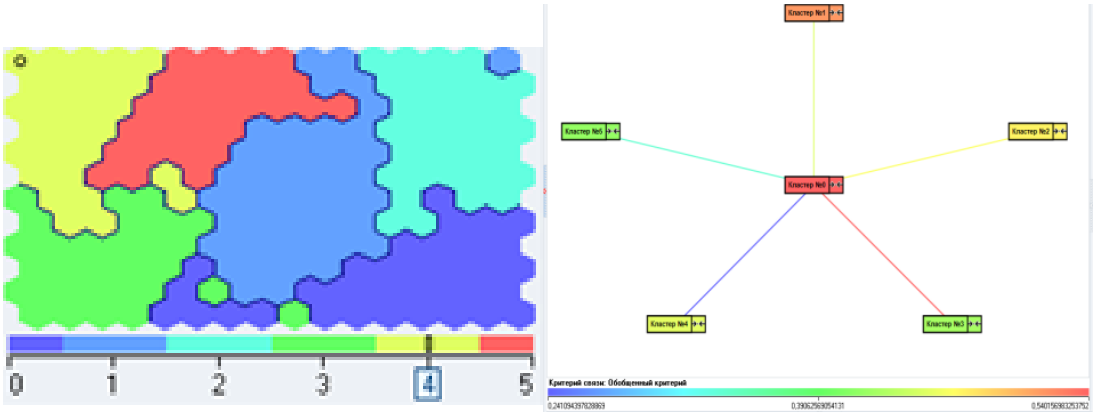

Fig. 1. Map of Kohonen

TABLE III. A FRAGMENT OF THE SOURCE DATA, INFORMATION

Client	Project	Partner	Manager	Type of services	Hours	Contract amount, $	Sum of GSR	Effect of currency rate fluctuations, $	Revenue, R	…
10000009	C0031086	1469	8831	5001	542	80000	102549	1022	79729	…
10000009	C0031087	1469	8831	5001	1724	110000	286246	-1243	90379	…
10000185	C0031809	1519	2281	5001	115	79352	19944	5077	72316	…
10000230	C0032729	1450	6146	5001	178	31182	27670	492	24211	…
10000341	C0032071	1346	2281	5001	584	80000	83581	247	79981	…
10000535	C0032930	1476	7434	5001	351	21000	69645	345	19473	…
10000661	C0031099	8682	5407	5001	558	25000	93190	-585	18126	…

TABLE IV. CHARACTERISTICS FOR CALCULATING GSR AND DIRECT COSTS

Грейды	INT	A1	A2	A3	SNR	MGR	SMG	DIR	PTR
X, hours	X1	X2	X3	X4	X5	X6	X7	X8	X9
У, gross rates	У1	У2	У3	У4	У5	У6	У7	У8	У9
A, cost rates	A1	a2	A3	A4	A5	A6	A7	A8	A9

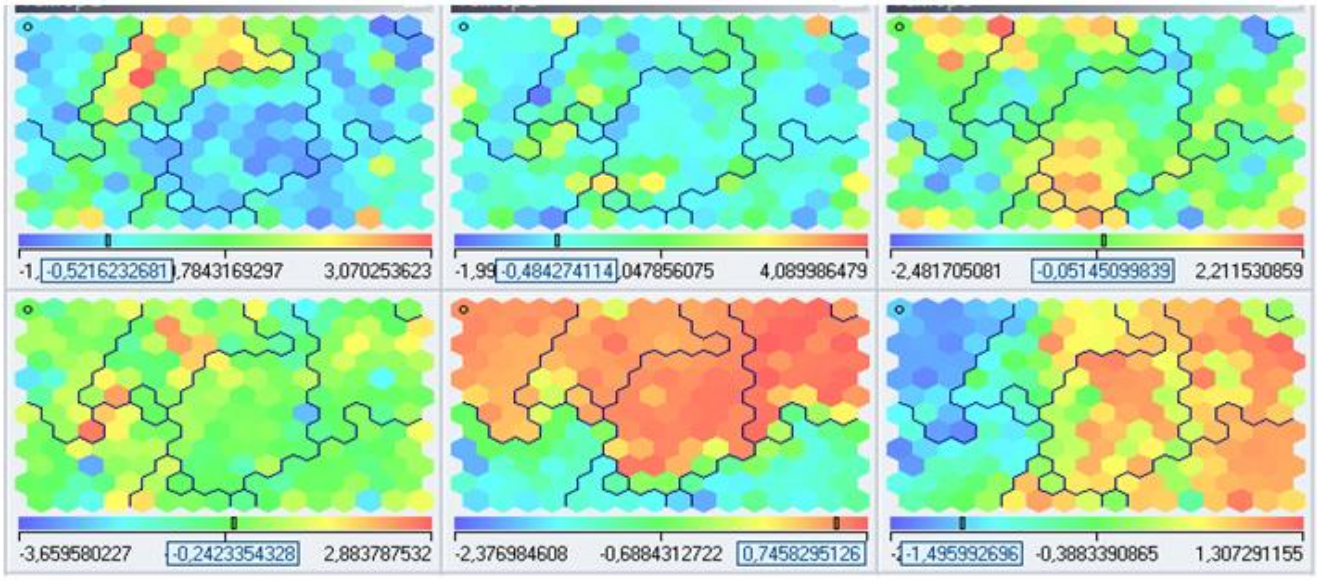

Fig. 2. Cluster analysis results

As a rule, the composition of the team is very diverse, since both experienced employees and beginners participate in the project. For each grade, the company has determined the values of the rates of gross hourly cost of work (gross rates) and the values of cost per hour (cost rates). At the same time, as a result of work on the project, the grade spends a certain number of hours. The presented characteristics (Table 4) were used to calculate GSR and direct costs.

We get that $GSR = \sum_{i=INT}^{PTR} x_i * y_i$, and direct costs = $\sum_{i=INT}^{PTR} x_i * a_i$, that is, these values are interrelated.

Further, for the clustering of objects, the Kohonen Self Organizing Map was used (Fig. 1). The number of clusters was determined by the system automatically (6 clusters).

The data obtained as a result of clustering were used at the stage of formalizing knowledge to form production rules, on the basis of which decision-making in the proposed DSS is based.

Examples of rules:

Rule 1. IF «type_of_service»= «StatutoryYearendAudit» OR «type_of_service»= «Statutory Referral Work» AND «project_partner»=9 AND «total_hours»/ «total_person»<=20, THEN «project_manager»=21;

Rule 2. IF«type_of_service»= «StatutoryYearendAudit» AND «project_partner»=8 AND «total_hours»/ «total_person»<=16 AND «contract_value»>4000000, THEN «project_manager»=7;

….

Rule n. IF«type_of_service»= «StatutoryYearendAudit» AND «project_partner»=2 AND «total_hours»/ «total_person»<=180 AND «contract_value»>30600, THEN «project_manager»≠2.

V. CONCLUSION

In this article the problem of forming a project team was analyzed, the work of other researchers and approaches to solving this problem were studied. Researches in the field of tacit knowledge were analyzed by authors of this article.

The methodology for solving the problem of decision support in the formation of a project team is proposed in the research work, the analysis of methods, criteria and means for assessing the effectiveness of the project, methods for the formation of project teams is carried out, the mathematical and algorithmic support of the decision support system in the formation of a project team has been developed.

Analytical platform Deductor Studio is used for Data Mining, in particular, for assess of data quality in the form of a comprehensive quality assessment of data sets based on gaps number, outliers and detected extreme values; Varimax method is used for factor analysis; Kohonen's neural network is applied for clustering objects.

According to the assessment of the methodology of the Cumulative Economic Effect, the develop and implementation of the proposed solution will have a clear positive effect both on the decision-making process in the formation of the project team, and on the economic results of the company.

ACKNOWLEDGMENT

The results of the research presented in this article were supported by Grants RFBR 19-07-00709 and 18-07-00193.

The study is conducted with financial support from the Ministry of Education and Science of the Russian Federation as part of the basic part of the state assignment to higher education educational institutions # FEUE-2020-0007.

REFERENCES

[1] Polany, M.: Personal knowledge. On the way to postcritical philosophy. Progress, Moscow, 1998.

[2] Amaeva, L.A.: Using data mining methods for user modeling. Vestnik of Kazan Technological University 18(1), 320-322 (2015).

[3] Zubkov, E.V., Belov, V.M.: Methods of data mining and intrusion detection // Vestnik SibGUTI 1, 118-133 (2016).

[4] Pavlova, M.E., Guryeva, T.G., Abrukov V.S.: Modeling the educational process in a university using methods of data mining // Conference: Mathematical models and their applications collection of scientific papers. Cheboksary, 2018 Publisher: Chuvash State University named after I.N. Ulyanova (Cheboksary). Не нашла статью, не знаю, как правильно оформить, не хватает выходных данных.

[5] Dyuk, V.A., Flegontov, A.V., Fomina, I.K.: The use of technologies for data mining in the natural sciences, technical and humanitarian fields. Izvestia: Herzen University Journal of Humanities & Sciences 138, 77-84 (2011).

[6] Tsvetkov V.Ya.: Spatial Information Models. European Researcher, vol. 60, 10-1, 2386-2392 (2013).

[7] Bolbakov, R.B.: The relationship between explicit and implicit knowledge. Perspectives of Science and Education 1(13), 10-16 (2015).

[8] Neiskiy, I.M., Filipovich, A.Yu.: Development of a tariff policy for brokerage services clients based on adaptive clustering methods. Applied Informatics 1(31), 3-11 (2011).

[9] Tsvetkov, V.Ya.: Analysis of implicit knowledge. Perspectives of Science and Education 1, 56-60 (2014).

[10] Tsvetkov, V.Ya.: Externalization of implicit knowledge. International Journal of Applied and Fundamental Research 12-1, 156-161 (2016).

[11] Emelyanova, E.A.: The transfer of implicit knowledge as a factor in the formation of clusters and increase the innovative activity of business. Modern problems of science and education 1, (2012). http://www.science-education.ru/en/article/view?id=5439.

[12] Volchik, V.V., Maslyukova, E.V.: The trap of metrics, or why implicit knowledge is underestimated in the process of regulating the sphere of education and science. Journal of institutional studies, vol. 10, 3, 158-179 (2018).

[13] Zvyagintsev, I.V., Ruzakova, O.V.: Formation of a knowledge management system in an organization using implicit knowledge. Fundamental Research 2-20, 4464-4468 (2015).

[14] Kimble, C.: Knowledge management, codification and tacit knowledge. Information Research 18(2), 577 (2013).

[15] Aamodt, A., et al.: Different roles and mutual dependencies of data, information, and knowledge—an AI perspective on their integration. Data & Knowledge Engineering 16(3), 191-222 (1995).

[16] Duan, Y., Shao, L., Hu, G., Zhou, Z., Zou, Q., Lin, Z.: Specifying architecture of knowledge graph with data graph, information graph, knowledge graph and wisdom graph. International Conference on Software Engineering Research, pp.327-332. IEEE (2017).

[17] Grishman, R.: Information extraction: Techniques and challenges. In: Pazienza M.T. (eds) Information Extraction A Multidisciplinary Approach to an Emerging Information Technology. SCIE 1997. Lecture Notes in Computer Science (Lecture Notes in Artificial Intelligence), vol. 1299. Springer, Berlin, Heidelberg (1997).

[18] Zins, C.: Conceptual approaches for defining data, information, and knowledge. Journal of the Association for Information Science & Technology 58(4), 479-493 (2014).

[19] Cowan, R., David, P.A., Foray, D.: The explicit economics of knowledge codification and tacitness. Industrial and Corporate Change 9(2), 211-253 (2000).

[20] Baskarada, S., Koronios, A.: Data, Information, Knowledge, Wisdom (DIKW): A Semiotic Theoretical and Empirical Exploration of the Hierarchy and its Quality Dimension. Australasian Journal of Information Systems 18(1), 2013. https://doi.org/10.3127/ajis.v18i1.74818. 10.3127/ajis.v18i1.748.

[21] Allen, J., Khader, M., Rosellini, A.: Investigating wisdom: Call for research. In: Proceedings of the Association for Information Science and Technology 56(1), 599-600 (2019). DOI: 10.1002/pra2.102

[22] Yao, X., Jin, H., Zhang, J.: Towards a wisdom manufacturing vision. International Journal of Computer Integrated Manufacturing 28(12), pp. 1291-1312 (2015). DOI: 10.1080/0951192X.2014.972462

[23] V. V. Evsujkov, "Data Mining as a tool for decision support in the banking system financial management", Izvestiya Tul'skogo gosudarstvennogo universiteta. Gumanitarnye nauki [Izvestiya of Tula State University. Human Sciences], vol. 4(1), 2014, pp. 374-384. (in Russian).

[24] F.H. Asci, "The effects of physical fitness training on trait anxiety and physical self-concept of female university students", Psychology of Sport and Exercise, vol 4, 2003, pp.255–264. URL: http://dx.doi.org/10.1016/S1469-0292(02)00009-2.

[25] C. Wei, L. Xiaodong, M. Lihong, and C. Liang, "Comparison of the promoting effect of different training programs on the physical and psychological quality of college students based on data mining algorithm", BoletinTecnico/TechnicalBulletin, vol. 55 (19), 2010, pp. 583-588.

[26] L. Yu, "Occupation Oriented and Data Mining Based Personalized Physical Quality Promotion" in the Proceedings of the 2nd International Conference on Information and Communication Technology for Education (ICTE 2015), 2015, pp. 234-237.

[27] Y. Bi, L. Jia and QB Wang, "Formulation of Scheme for Sports Training Management Based on System Dynamics" in the Proceedings of the 4th International Conference on Machinery, Materials and Information Technology Applications (ICMMITA), 2016, pp. 1500-1503.

[28] M. Li, "Investigation of Action Decomposition of Track and Field Sports Based on Data Mining" CHIMICA OGGI-CHEMISTRY TODAY, vol. 36(6), 2018, pp. 746-749.

[29] S.N. Vasiliev, "Knowledge formalization and management on basis of positive-educated languages"[Formalizaciya znanij i upravlenie na osnove pozitivno obrazovannyh yazykov] in Informacionnye tekhnologii i vichslitel'nye sistemy (Journal of Information Technologies and Computing Systems), vol. 1, 2008, pp. 3-17. (in Russian).

[30] A.N. Raikov, "Automated Synthesis of Cognitive Model on the Base of Big Data Analysis and Deep Learning" in International conference Internet and Modern Society (IMS-2018), St.Petersburg, Russia, vol. 2, 2018, pp. 103-111. (in Russian).

[31] L. J. Gressgård, T. Nesheim, T., "Knowledge Management Systems and Work Improvements: The Moderating Effects of Work Characteristics" in Journal of Information & Knowledge Management (JIKM), World Scientific Publishing Co. Pte. Ltd., 17(04), 2018, pp. 1-18.

[32] J. L. Medina Moya, B. Jarauta Borrasc, and J. Menegaz, "The formalization of professional knowledge in the curriculum" in Revistaibero-americanadeestudosemeducacao, vol. 13(2), 2018, pp. 588-603.

[33] Wan Yusof, W. S. E. Y., Zakaria, O., Zainol, Z. & Ananthan S.: DIKW Application on Knowledge Based Framework with Situational Awareness. International Journal of Academic Research in Business and Social Sciences 8(6), 1110-1120 (2018).