PROCEEDINGS OF SCIENCE

Methods and algorithms of the analytical platform for analyzing the labor market and the compliance of the higher education system with market needs

S.D. Belov[a,b], **I.A. Filozova**[a,b], **Y.E. Gavrilenko**[b], **A.V. Ilina**[a,b,*], **J.N. Javadzade**[a], **I.S. Kadochnikov**[a], **V.V. Korenkov**[a,b], **I.S. Pelevanyuk**[a,b], **D.I. Priakhina**[a,b], **R.N. Semenov**[a,b], **V.A. Tarabrin**[b], **and P.V. Zrelov**[a,b]

[a] *Joint Institute for Nuclear Research, Meshcheryakov Laboratory of Information Technologies, Joliot-Curie St., Dubna, Moscow Region, 141980, Russia*

[b] *Plekhanov Russian University of Economics, Stremyanny lane 36, Moscow, 117997, Russia*

E-mail: belov@jinr.ru, fia@jinr.ru, yulya952@rambler.ru, annailina@jinr.ru, jjavadzade@yandex.ru, kadivas@jinr.ru, korenkov@jinr.ru, pelevanyuk@jinr.ru, pry-darya@yandex.ru, roman@jinr.ru, vitaliy.tarabrin@gmail.com, zrelov@jinr.ru

The article discusses the methods and algorithms that underlie the analytical platform for automated monitoring and analysis of the labor market in the Russian Federation, as well as the analysis of the higher education system's compliance with the labor market's current needs. The study involved natural language processing methods and Big Data technologies. The general scheme corresponds to end-to-end processing – from data collection and storage, their transformation, analysis, and modeling, to visualization of results and decision-making. The analytical core of the system is a module for intellectual analysis of the texts of job advertisements in the labor market. The vacancies are collected from the most complete databases in Russia (namely HeadHunter, Work in Russia and SuperJob). Job descriptions of vacancies are matched with the official list of professions of the Ministry of Labor and Social Protection of Russia using semantic analysis based on neural models trained on large arrays of texts. Also, using semantic analysis, automated monitoring and intellectual analysis of the staffing needs of the all-Russian and regional labor markets are carried out according to the range of specialties of the university. Data gathering has been ongoing from 2015 up to now.

Keywords: Big Data, digital platform, labor market analysis, natural language processing, semantic analysis

The 6th International Workshop on Deep Learning in Computational Physics (DLCP2022)
6-8 July 2022,
JINR, Dubna, Russia

[*]Speaker

1. Introduction

It is impossible to base effective market needs forecasting on average statistical estimates, various selective indicators, and various sociological surveys. Instead, such forecasting is possible only based on monitoring various resources and services that provide information about employers' offers. Subsequent analysis of these data allows one to see the labor market's structure and its change over time and provides a forecast for the development of both the market and its sectors.

These functions are performed by specialized information and analytical systems [1]. Several factors must be considered when designing such systems. Many sources of information to be analyzed, the lack of unified systematic approaches to the analysis of the labor market in terms of changing requirements for the qualification of the labor force and reflecting future market needs in the content of educational programs. Some factors are related to the need for uniformity in the wording of employers' vacancies and candidates' resumes.

In the scientific literature, many works are devoted to studying various aspects of the labor market, a detailed review of which is beyond the scope of this article. We can only point out some of them that are combined by common "instrumental" means – machine learning, natural language processing algorithms, methods based on semantics, ontologies, as well as aimed at creating software complexes for automatic data processing – decision support systems, end-to-end data processing and analysis, etc. For example, the method of assessing the nesting of taxonomic data in the labor market is MEET [2], the eDoer recommendation system focused on the labor market [3], OntoJob – building ontologies of the labor market [4]. In [5], kNN classifiers are used to extract information from a given set of vacancies, and [6] presents an approach to automatically classifying vacancies on the Internet using machine learning according to the standard professions' taxonomy.

Related to this task is identifying the education system's degree of compliance with the labor market's needs. The interaction of the labor market and the vocational education system is a complex, ideally mutually agreed process. The labor market covers its needs for a qualified workforce in the person of graduates trained in the field of education. At the same time, the education system must quickly respond to changes in the economic situation. In the world literature, the issue of the discrepancy between the economy's demand and the supply of professions produced by vocational education systems has been discussed for a long time [7]. At the same time, as noted in [8], the relationship between the needs for certain specialists presented by the labor market and the range of specialists who graduated, in particular, from Russian universities, from different points of view, there are still only a few studies. Thus, work [9] is devoted to the development of the LO-MATCH web platform using ontological descriptions and tag clouds to identify the correspondence between the needs of students, on the one hand, and the capabilities of the education system, on the other, as well as between the profiles of applicants and the requirements of companies. The work [10] discusses the creation of an ontology of job knowledge (Job-Know), which combines elements of work and knowledge collected from various sources, which in the future, according to the authors, can ensure compliance with the requirements of the labor market and the education system. The work [11] considers an ontological model of a professional field for a selected educational profile, which considers the relationship between professional standards, labor market requirements, and a graduate's competency model.

Unlike traditional approaches to comparing and analyzing the content of educational programs based on ontological models and expert systems, in [12, 13], the analysis of the compliance of educational standards with the needs of the labor market was first performed based on an algorithm for translating short sentences into a vector space using the neural network model word2vec [14]. The comparison is based on the hierarchical labor market and education models proposed in these works. The first results were obtained in 2016, and this direction continues to develop consistently [15-20].

The papers [21, 22] consider the method of decision support in the formation of educational programs, taking into account the labor market's needs and the semantic search for educational content for the given requirements of the labor market, determined by professional standards. Neural network models word2vec, fastText [23] and graph models of educational programs, professional standards and vacancies are used. In [24], devoted to the problem of correlating professional standards and the real demand presented by employers in the labor market in the field of top management, algorithms were used to quantify high-quality information regarding the knowledge, skills and abilities demanded by employers, as well as the method of grouping professional standards and vacancy announcements between yourself. It is also shown that many professions and their names from professional standards only partially correspond with the HeadHunter database and vice versa.

Both tasks - the analysis of the labor market as such and the compliance of educational standards with its needs - are successfully solved using the automated information and analytical system developed by the authors for the intellectual analysis of labor market needs. Information on used educational standards of higher education comes from open data sources. The system is aimed at a wide range of users: authorities and management of regions and municipalities; management of universities, companies, recruitment agencies; graduates and future students. Since 2015, a unique database of vacancies, applicants' CVs and information about employers with more than 20 TB accumulated.

Big Data approaches and technologies are used to collect and process data. The analytical block of the system is built on used machine learning technologies for semantic analysis and vector representation of words and short sentences. The system allows one to assess the need for specific professions for the regions, address the compliance of professional standards with real market jobs, and plan the number of funded places in colleges and universities. Based on the accumulated historical data, it is possible to make forecasts for the development of the labor market.

Currently, one of the projects implemented on the platform is the information and analytical resource "Professions, salaries, universities: the entrant's navigator 2022" created with MIA Rossiya Segodnya [25]. Its goal is to help high school graduates navigate when choosing a university.

2. Big Data platform of the Plekhanov Russian University of Economics

In Plekhanov Russian University of Economics [26] was developed a prototype of an analytical platform that implements the concept of end-to-end data processing based on Big Data technologies and machine learning. In addition to our developments, open-source software products are used, as well as a business intelligence system MS Power BI[1] (an option for presenting results).

The platform, called the Automated System for Monitoring, Analyzing and Forecasting the Development of the Labor Market in the Russian Federation, solves the tasks of monitoring the personnel needs of enterprises and organizations in real-time and the task of analyzing the that needs for the range of specialties of higher educational institutions. We use data from open sources – the Internet portals of the recruitment companies *HeadHunter*[2] and *SuperJob*[3], as well as the portal *Work in Russia*[4] (the federal state information system of the Federal Service for Labor and Employment). In addition, the register of approved professional standards[5] and the Federal State Educational Standards for Higher Education (FGOS HE)[6] are used as guiding documents.

The system allows for estimating the needs for specialists in a particular region, the proposed range of salaries and the level of requirements for job seekers, as well as understanding under what educational programs they can be trained.

The nature of vacancies is such that they first open, then they are in a published state for some time, and then they are closed, while it is vital to monitor the status of specific vacancies day by day to fix the moments of appearance and closing.

Data on vacancies undergo multi-level processing, including collection, filtering, transformation, analysis, modeling, and presentation in a form convenient for subsequent visualization.

The system is a set of specialized functional blocks:

- Text data collection block (operates automatically using open sources - Internet portals and recruiting agencies). The block includes filtering, pre-processing, transformation, and storage. The register of approved professional standards and the Federal State Educational Standards of Higher Education are used as guiding documents.
- Block of automatic processing – performs processing, analysis and automatic linking of labor market requirements and competencies of higher education (algorithms for machine learning, semantic and linguistic analysis).
- Modeling and visualization block – interpretation of results, identification of trends, and user interfaces for creating and displaying reports based on business intelligence technologies.

To obtain correct statistical results, the system is supplemented with tools that solve the following tasks:

- Search for duplicate vacancies (required because data comes from multiple sources).
- Classification of vacancies by industry.

---

[1] https://powerbi.microsoft.com/

[2] *HeadHunter* – one of the largest portals for finding jobs and employees worldwide (according to the *Similarweb* rating). The Internet resource contains about 51 million CVs, 1,032 thousand vacancies in the database (https://hh.ru).

[3] Official site *SuperJob* — An IT company that develops hiring and job search technologies (https://www.superjob.ru).

[4] Web portal *Work in Russia*— Federal state information system of the Federal Service for Labor and Employment (https://trudvsem.ru).

[5] Official Internet resource of the Ministry of Labor and Social Protection of the Russian Federation, providing information on the list of professional standards (https://profstandart.rosmintrud.ru/).

[6] Russian Federal state educational standards (http://fgos.ru).

• Analysis of the content of the job offer, analysis of individual requirements for skills and competencies.

The Big Data analytical platform integrates various resources, software components and services, including collection, storage, distributed computing and analytics, modeling, visualization, resource management and authentication.

3. Estimation of semantic similarity of text documents

Modeling a word's semantics (meaning) is one of the most important problems associated with natural language processing. Semantic analysis results are used in search engines [27], automatic translation systems, and other areas related to natural language text processing [28].

Currently, approaches to vector representations of words use, in particular, the so-called predictive model based on neural networks [29]. One of the main tools for the vector representation of words is the *word2vec* model [14].

The basic principle of *word2vec* is to find relationships between word contexts assuming that words that appear in similar contexts tend to indicate similar concepts (being semantically close). The problem solved by word2vec can be formalized as follows: minimize the distance between vectors of the words that appear next to each other and maximize the distance between vectors of words that appear pretty far away. "Nearby" in this case means "in similar contexts". For example, the words "analysis" and "research" are often found in similar contexts, *word2vec* analyzes these contexts and concludes that these words are close in meaning. Context analysis is carried out on large arrays of text. In our task, we used the corpus of the Russian Wikipedia and the national corpus of the Russian language, as well as the models of distributive semantics *rusvectōrēs* [30].

There are approaches to creating a predictive model for immersing the entire document into a vector space [31]. However, comparing short sentences by the similarity of meaning has specific features. Using existing models for translating words or documents into a vector space without changes gives an unsatisfactory result.

Considering that the wording of duties in vacancy announcements, as well as the wording of texts of professional educational competencies, contain about 10 words on average, the task of assessing the semantic similarity of two short sentences underlies the analytical part of the system. In 2016, the first results [12, 13] of the analysis of the labor market and the compliance of educational standards with market needs were obtained based on the proposed algorithm for translating short sentences into vector space using *word2vec*. Vector representation allows one to calculate the semantic "similarity" of words based on the calculation of the cosine distance. Projections of vectors of words with similar meanings are located close to each other and form some semantic clusters.

For two words, w_1 and w_2, represented as vectors, the formula for calculating the semantic proximity:

$$cos\left(\vec{V}(w_1),\vec{V}(w_2)\right) = \frac{\sum_{i=1}^{n} w_{1_i} \cdot w_{2_i}}{\sqrt{\sum_{i=1}^{n}\left(w_{1_i}\right)^2} \cdot \sqrt{\sum_{i=1}^{n}\left(w_{2_i}\right)^2}} \quad (1)$$

To calculate the semantic similarity of short sentences representing, on the one hand, the texts of duties in vacancy announcements and, on the other hand, the texts of educational competencies or the texts of the register of professional standards (professional guide), then used weighted values of their constituent words in vector representation.

The calculation of short sentence vectors $s = \{w_1, w_2, \dots, w_k\}$ defined as the weighted average of the component word vectors:

$$\vec{v}(s) = \frac{\sum_{i=1}^{k} p_i \cdot \vec{v}(w_i)}{\sum_{i=1}^{k} p_i}, \qquad (2)$$

where $w_i (i = 1,2, \dots, k)$ – words of the sentence, p_i – words' weights, k – number of words in the sentence. Word weights can be calculated in different ways, depending on the task, for example, as the ratio of the frequency of a word to the volume of the corresponding lexicon or using the TF-IDF (term frequency-inverse document frequency) metric.

In this study, various weights were used; the results presented below refer to the choice of TF-IDF as a metric.

The TF (term frequency) value reflects the frequency of occurrence of a given word in the current text fragment[7]:

$$TF(t, s_i) = \frac{n_t}{\sum_k n_k},$$

where n_t – number of words t in the document s_i, $i \in \{E, V\}$, E – index for the texts of educational system (competences), V – index for the texts of labor market (vacancies), $\sum_k n_k$ – total number of words in the text fragment s_i.

IDF (inverse document frequency) is the value reflecting the frequency of occurrence of a given word in the original set of text fragments D_i, $i \in \{E, V\}$ for both the education system (E), and the labor market (V), and used to reduce the weights of the most common words (for the set of source documents D_i, $i \in \{E, V\}$ – the higher the value of the *IDF*, the less the weight of the word):

$$IDF(t, D_i) = log\left(\frac{|D_i|}{\left|\{s_{ij} \in D_i \middle| t \in s_{ij}\}\right|}\right),$$

where $|D_i|$– number of documents s_{ij} in D_i, $\left|\{s_{ij} \in D_i \vee t \in s_{ij}\}\right|$– number of documents s_{ij} from the collection D_i, where occurs the word t.

Thus, the value of *TF-IDF* for each word t of the text fragment s_{ij} is given by:

$$TF\text{–}IDF(t, s_{ij}, D_i) = TF(t, s_{ij}) \times IDF(t, D_i),$$

where the index $i \in \{E, V\}$ corresponds to either competencies or vacancies.

4. Analysis of the labor market based on the reference book of professions

First, it is necessary to classify vacancies for a quantitative study of the structure of the labor market's demands. Since in the ads, even the names of vacancies with similar or identical positions and labor functions can differ quite a lot, so a separate reference list is needed to compare vacancies with. The most detailed and complete classifier related to the labor market is the Directory of Professions of the Ministry of Labor of the Russian Federation [32]. In addition to the information needed directly for comparison with the vacancy, such as the name of the profession, the area of professional activity, position, description of the profession and keywords,

[7] It is the text listing the duties in the vacancy information for the labor market, and the name of the professional competence for the "education system".

PoS(DLCP2022)028

there is information about the relationship of the profession with data in other classifiers (professional standards, occupations, professions and positions, state educational standards, etc.).

Example of one of the direct methods of comparing vacancies and professions is, for example, the selection and direct comparison of sets of keywords for them. However, because the terminology in both cases is often markedly different, a direct comparison is ineffective. At the same time, the semantic similarity analysis makes it possible to obtain reasonably good results [15].

When performing semantic analysis, the methods of vector representation of words and phrases described in the previous section are used. For vacancies and professions from the directory, vectors are constructed in the semantic space. Then, by calculating the proximity coefficient for each pair of "vacancy - profession", the most suitable profession from the directory is determined.

Linear combinations of vectors for various attributes of vacancies and professions are used to construct the corresponding vectors in the semantic space. Vectors for text attributes are constructed based on the vectors of their constituent words in the *word2vec* representation:

$$\vec{v}(s) = \sum_{i=1}^{k} \vec{v}(w_i),$$

where $w_i (i = 1,2, \dots, k)$ – constituent words, k – words number in the attribute's text. The difference from the construction of similar vectors for short sentences is in the absence of weighting coefficients.

The vector for a vacancy is constructed as follows:

$$\vec{V}(t, d, r, a) = T_V \vec{v}(t) + D_V \vec{v}(d) + R_V \vec{v}(r) + A_V \vec{v}(a),$$

where string parameters *t* – the name of the vacancy, *d* – its description, *r* – job duties, *a* – area of professional activity; T_V, D_V, R_V, A_V – numerical coefficients.

Corresponding vector for the profession:

$$\vec{P}(t, d, p, k, a) = T_P \vec{v}(t) + D_P \vec{v}(d) + O_P \vec{v}(o) + K_P \vec{v}(k) + A_P \vec{v}(a),$$

where t – profession title, d – its description, o – possible jobs, k – keywords, a – area of professional activity; T_P, D_P, O_P, K_P, A_P – numerical coefficients. The values of the coefficients for vacancies and professions give the minimum average distance between a vacancy and a profession from the directory for a sample of 1000 vacancies when using an expert assessment of the correspondence between a vacancy and a profession.

In the next step, the distance between the vectors $\vec{V}$ and $\vec{P}$ is calculated, and the profession with the greatest semantic similarity is selected for the vacancy. Each vacancy from the announcement is compared with each job description from the directory. After matching, vacancies can be grouped, and further statistical analysis can be carried out. Additional information related to the profession from the reference books and classifiers mentioned earlier can also be used.

The accuracy and speed of the presented method can be increased both by introducing a preliminary clustering of vacancies (for example, by field of activity), and by a multi-level selection of the closest professions by gradually narrowing the comparison area (name of the vacancy and profession, possible positions, duties, etc.).

5. Labor market and educational standards

The analytical system provides additional opportunities for identifying qualitative and quantitative links between education and the labor market. The implementation of a competency-based approach to the training of university graduates is regulated by the Federal State Educational Standard of Higher Education and involves the formation of a set of general cultural, general professional and specialized professional competencies among students. Professional competencies are grouped by type of activity. From the perspective of professional activity, we can talk about the competence model of a specialist as a subject in demand in the labor market. Occupational standards can link the requirements for qualifications in the world of work and the requirements for learning outcomes in the field of education.

The labor market and the education system are presented in simplified hierarchical models (for details, see [12]), which makes it possible to link market requirements and educational competencies at different levels. At the same time, the educational model of the university contains 5 hierarchical levels ("faculties", "training areas", "profile", "type of activity", "competence content"), and the labor market model contains 4 levels ("fields of activity", "directions", "professions", "requirements"). The mapping of one model to another occurs through the establishment of links at their lower levels: for the university, these are "competences"; for the labor market, they are "requirements (duties)". Implementing connections at the lower levels allows, going from bottom to top, to obtain connections at any of the selected levels (depending on the task). For example, to display the links "field of study" – "professions".

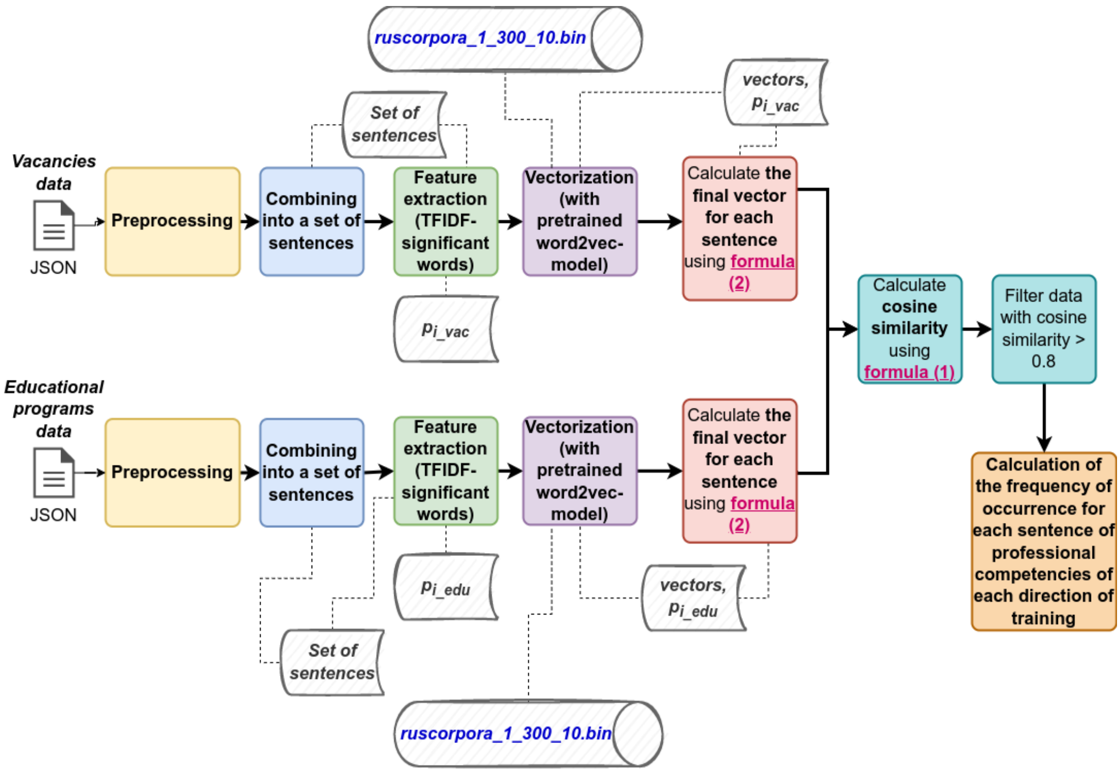

Figure 1: Block diagram of the algorithm for calculating semantic similarity between the texts of educational programs and job advertisements

As well as in the case of linking market requirements with a directory of professions, linking with educational standards occurs by determining semantic similarity by calculating a cosine measure of proximity between vector representations of texts in the higher education system and the labor market.

The distributive semantics model *rusvectōrēs ruscorpora_1_300_10.bin* [30] containing about 185000 words (vector size 300, window size 10 words) was used to obtain word vectors. Before calculating the *TF-IDF* metric and applying the *word2vec* model, the name of each professional competence and each sentence from the “duties” field was subjected to preliminary processing: removal of non-printable characters (formatting marks), removal of punctuation marks, removal of words that do not carry a semantic load (prepositions, conjunctions, pronouns, interjections, etc.), tokenization and lemmatization of words, markup by parts of speech (using *Universal PoS* tags [33]).

Based on the received representation, an analysis is made of the demand for both professional competencies and the direction or profile of training for any parameters of interest (regions, professional areas, salaries, etc.). The demand for professional competence in the labor market is determined by counting its occurrence in the resulting data frame. The block diagram of the algorithm is shown in Figure 1.

6. Data analysis and presentation of results

The automated information and analytical system have two options for working: with a web interface and a business intelligence system (currently Microsoft Power BI). The web interface is most convenient for demonstrating and analyzing (primarily visual) prepared charts and graphs. The business intelligence system is intended for specialists who know the data structure and its topology.

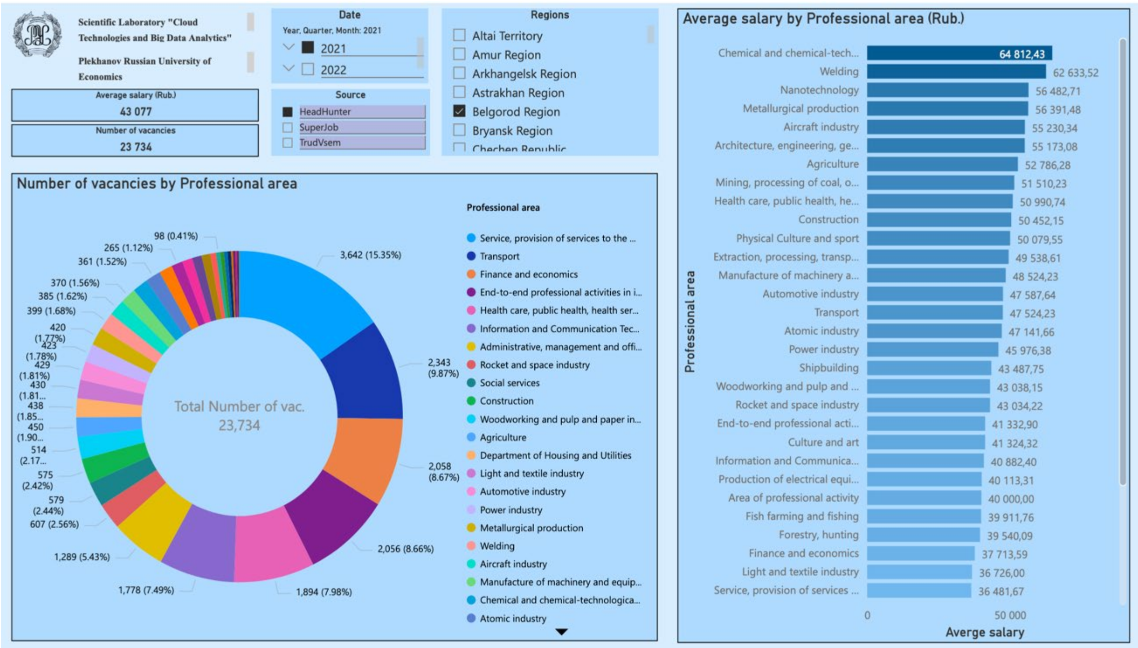

Figure 2: Distribution of the number of vacancies and the average salary by professional fields for 2021-2022

Figure 2 shows one of the analytical dashboards of the system with the distribution of vacancies and average salaries by professional area for the period 2021-2022 (using the Microsoft Power BI business intelligence system).

7. Conclusion

An analytical platform has been developed that implements automated monitoring and analysis of the labor market in the Russian Federation, as well as monitoring the compliance of employers' staffing needs with the level of training of specialists.

The platform is based on Big Data solutions and technologies. It implements a complete data processing cycle – from collecting, storing, filtering, transforming, semantic analysis and modeling to services for visualizing results and making decisions. Open sources provide the information base – databases of vacancies *HeadHunter*, *Work in Russia* and *SuperJob*, the official list of professions of the Ministry of Labor and Social Protection of the Russian Federation, higher education standards.

Data collection has continued since 2015 to the present; a unique database of vacancies, resumes of applicants and information about employers with a volume of more than 20 TB is gathered. The analytical core of the platform is built on the methods and algorithms developed by the authors using natural language processing methods, in particular, semantic analysis based on neural network models trained on large arrays of texts.

On the created platform, several interesting results were obtained in the study of the all-Russian and regional labor markets. In addition, new results were received on the analysis of the relationship between the education system and the labor market; in particular, the demand for professional competencies was determined for twenty-three areas and training profiles in the main professional educational programs of the bachelor's degree at the Plekhanov Russian University of Economics.

The system is intended for heads of regions, universities, companies, and recruitment agencies.

8. Acknowledgements

The study was carried out at the expense of the Russian Science Foundation grant (project No.19-71-30008).

References

[1] T. Sparreboom, *Labour market information and analysis systems*, Perspectives on labour economics for development, Geneva: ILO, 2013, pp. 255-282.

[2] Malandri, Lorenzo et al., *MEET: A Method for Embeddings Evaluation for Taxonomic Data*, 2020 International Conference on Data Mining Workshops (ICDMW) (2020): 31-38.

[3] Tavakoli, Mohammadreza et al., *An AI-based open recommender system for personalized labor market driven education*, Adv. Eng. Informatics 52 (2022): 101508.

[4] Vrolijk, Jarno et al., *OntoJob: Automated Ontology Learning from Labor Market Data*, 2022 IEEE 16th International Conference on Semantic Computing (ICSC) (2022): 195-200.

[5] I.A. Wowczko, *Skills and Vacancy Analysis with Data Mining Techniques*, Informatics 2015, 2, 31-49; DOI:10.3390/informatics2040031.

[6] Boselli, Roberto et al., *Classifying online Job Advertisements through Machine Learning*, Future Gener. Comput. Syst. 86 (2018): 319-328.

[7] F. Cörvers & H. Heijke, *Forecasting the labour market by occupation and education: some key issues.* Researchcentrum voor Onderwijs en Arbeidsmarkt, Faculteit der Economische Wetenschappen, ROA Working Papers, 2004, no. 4.

[8] A.A. Tkachenko & A.B. Ginoyan (2021), *Assessment of the Occupational Structure of Employed in Regions. Ekonomika regiona* [Economy of regions], 17(4), 1224-1238 (In Russ.) DOI: https://doi.org/10.17059/ekon.reg.2021-4-13.

[9] V. Gatteschi, F. Lamberti, A. Sanna and C. Demartini, *Using Tag Clouds to Support the Comparison of Qualifications, Résumés and Job Profiles*, ICETA 2011 9th IEEE International Conference on Emerging eLearning Technologies and Applications, October 27-28, 2011, Stará Lesná, The High Tatras, Slovakia.

[10] M. Khobreh, F. Ansari, M. Fathi, R.Vas, S.T. Mol, H. Berkers & K. Varga*, An ontology-based approach for the semantic representation of job knowledge,* IEEE Transactions on Emerging Topics in Computing, 2016, 4(3), 462-473; DOI: https://doi.org/10.1109/TETC.2015.2449662.

[11] Elena Boldyreva and Vadim Kholoshnia, *Ontological Approach to Modeling the Current Labor Market Needs for Automated Workshop Control in Higher Education*, MICSECS (2019). URL: http://ceur-ws.org/Vol-2590/paper29.pdf.

[12] P.V. Zrelov, V.V. Korenkov, N.A. Kutovskiy, A.S. Petrosyan, B.S. Rumiantsev, R.N. Semenov, I.A. Filozova, *Monitoring of the Labour Market Needs for University Graduates Based on Data Intensive Analytics,* Proceedings of the XVIII International Conf. «Data Analytics and Management in Data Intensive Domains» (DAMDID/RCDL'2016), October 11-14, 2016, pp. 124-131 (In Russ.).

[13] Petr V. Zrelov, Vladimir V. Korenkov, Nikolay A. Kutovskiy, Artem S. Petrosyan, Boris D. Rumyantsev, Roman N. Semenov, Irina A. Filozova, *Automated system for monitoring and analysis of compliance of the labour resources needs according to the specialties' nomenclature of higher educational institution,* Federalism, 2016, № 4 (84). pp. 63-76 (In Russ.).

[14] Tomas Mikolov et al., *Efficient Estimation of Word Representations in Vector Space* [arXiv:1301.3781v3, 2013].

[15] S.D. Belov, I.A. Filozova, I.S. Kadochnikov, V.V. Korenkov, R.N. Semenov, P.V. Zrelov, *Automated system to monitor and predict matching of higher vocational education programs with labour market,* CEUR Workshop Proceedings. 26. Ser. «Selected Papers of the 26th International Symposium on Nuclear Electronics and Computing, NEC 2017», 2017, pp. 98-104.

[16] S. Belov, I. Filozova, I. Kadochnikov, V. Korenkov, R. Semenov, P. Smelov, P. Zrelov, *Labour market monitoring system*, CEUR Workshop Proceedings, GRID 2018, Selected Papers of the 8th International Conference «Distributed Computing and Grid-Technologies in Science and Education», 2018, pp. 528-532.

[17] Sergey Valentey, Petr Zrelov, Vladimir Korenkov, Sergey Belov, Ivan Kadochnikov, *Monitoring of matching of vocational education with the needs of the labour market*, Social Sciences and Contemporary World (ONS), 2018, № 3, pp. 5-16 (In Russ.).

[18] S.D. Belov, J.N. Javadzade, I.S. Kadochnikov, V.V. Korenkov, P.V. Zrelov, *Big Data technologies for labour market analysis*, CEUR Workshop Proceedings, NEC 2019, Proceedings of the 27th Symposium on Nuclear Electronics and Computing, 2019, pp. 469-472.

[19] S.D. Belov, I.S. Kadochnikov, V.V. Korenkov, P.V. Zrelov*, Intellectual texts processing in socio-economic applications,* CEUR Workshop Proceedings. Ser. «SCG 2020, Proceedings of the Big Data Analysis Tasks on the Supercomputer GOVORUN Workshop», 2020, pp. 65-71.

[20] S. Belov, A. Ilina, J. Javadzade, I. Kadochnikov, V. Korenkov, I. Pelevanyuk, R. Semenov, P. Zrelov, V. Tarabrin*, Analytical platform for socio-economic studies,* CEUR Workshop Proceedings. 9. Ser. GRID 2021, Proceedings of the 9th International Conference «Distributed Computing and Grid Technologies in Science and Education», 2021, pp. 619-623.

[21] D.S. Botov, *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. (in Russ.) DOI: 10.14529/ctcr190101.

[22] D.S. Botov, Yu.V. Dmitrin, Yu.D. Klenin, *Neural Network-Based Semantic Search of Educational Programmes Fitting Labor Market Requirements*, Bulletin of the South Ural State University. Ser. Computer Technologies, Automatic Control, Radio Electronics, 2019, vol. 19, no. 2, pp. 5-15. (in Russ.) DOI: 10.14529/ctcr190201.

[23] P. Bojanowski, E. Grave, A. Joulin, T. Mikolov, *Enriching Word Vectors with Subword Information,* arXiv preprint [arXiv:1607.04606], 2016; DOI: 10.1162/tacl_a_00051.

[24] A.A. Ternikov, *Professional standards and demand in the labor market in the sphere of higher management in Russia,* Vestnik of Saint Petersburg University, 2022, Management 21 (1): 131-148. DOI: http://doi.org/10.21638/11701/spbu08.2022.106 (In Russ.).

[25] *«Professii, zarplaty, vuzy: navigator abiturienta 2022»* [Elektronnyi resurs] [Professions, salaries, universities: entrant Navigator 2022]. URL: https://na.ria.ru/20220323/vuzy-1779514893.html? (Accessed 17.07.2022).

[26] *Rossijskij ekonomicheskij universitet imeni G. V. Plekhanova* [Elektronnyi resurs] [Plekhanov Russian University of Economics]. URL: https://new2.rea.ru/ (Accessed 01.08.2022).

[27] Amir Efrati, *Google Gives Search a Refresh*, The Wall Street Journal, Retrieved July 13, 2012.

[28] Eva Martínez Garcia, Cristina España-Bonet, Lluís Màrquez, *Document-Level Machine Translation with Word Vector Models*, Proceedings of the 18th Annual Conference of the European Association for Machine Translation (EAMT), 2015, pp. 59-66.

[29] Barkan Oren, *Bayesian Neural Word Embedding* [arXiv:1603.06571], 2015.

[30] A. Kutuzov, E. Kuzmenko, *WebVectors: A Toolkit for Building Web Interfaces for Vector Semantic Models*. In: Ignatov D. et al. (eds)*,* Analysis of Images, Social Networks and Texts, AIST 2016, Communications in Computer and Information Science, vol. 661, Springer, Cham.

[31] Le Quoc et al., *«Distributed Representations of Sentences and Documents»* [arXiv:1405.4053].

[32] *Spravochnik professij Ministerstva truda i social'noj zashchity RF* [Elektronnyi resurs] [Directory of Professions of the Ministry of Labor and Social Protection of the Russian Federation]. URL: https://spravochnik.rosmintrud.ru/professions (Accessed 01.08.2022).

[33] *Universal POS tags* [Electronic resource]. URL: https://universaldependencies.org/u/pos/ (Accessed 01.08.2022 г.).

PoS(DLCP2022)028