AUTOMATED SYSTEM TO MONITOR AND PREDICT MATCHING OF HIGHER VOCATIONAL EDUCATION PROGRAMS WITH LABOUR MARKET

S.D. Belov[1,2], I.A. Filozova[1,2], I.S. Kadochnikov[1,2], V.V. Korenkov[1,2], R.N. Semenov[1,2], P.V. Zrelov[1,2]

[1] *Laboratory of Information Technologies, Joint Institute for Nuclear Research, 6 Joliot-Curie, Dubna, Moscow region, 141980, Russia*

[2] *Plekhanov Russian University of Economics, 36 Stremyanny per., Moscow, 117997, Russia*

E-mail: [a] sergey.belov@jinr.ru

Interaction of labour market and educational system is a complex process, with many parties involved (government, universities, employers, individuals, etc.). Both horizontal and vertical mismatch between skills and qualifications from one side and market's requirements from another are still widely observed in both developing and developed countries. To discover both qualitative and quantitative correlations between education system and labour market in a reasonable time, we proposed an intellectual system to monitor the demands of employers and match them with the educational standards and programs. The analysis is based on stringing together job requirements and single competencies from the educational standards, the lowest levels of the models of the labour market and the education system correspondingly. To automate the processing, we used machine learning technologies for semantic parsing and the vector representation of words and short sentences. Big Data approaches and technologies are in use for collecting and processing the data. The system allows to estimate a need for specific professions for regions, to consider matching of the professional standards with real market jobs, to plan the number of funded places in colleges and universities. Having historical data, it is also possible to make some further predictions.

Keywords: big data, machine learning, labour market, education, competencies

1. Introduction

Interaction of labour market and educational system is a complex process, with many parties involved (government, universities, employers, individuals, etc.). In the ideal world, this interaction would be coherent and perfectly balanced. Mostly it affects youth employment, so called school-to-work transition. Since the unfolding of the Great Recession in 2008, youth unemployment has been the forefront of political and academic debates. In most countries, the young employees have suffered more in recession than have more experienced ones [1]. High unemployment rate especially among young in a country or region [2] could cause the growth of social tensions or even be a breeding ground for extremism.

Many researchers are giving attention to volatile labour market and youths' complications with entering it. There are plenty of entities to influence the area, e.g. contract policies for new employees, state programs, etc. Governments still invest a lot in education, so do individuals. However, both horizontal and vertical mismatch between skills and qualifications from one side and market's requirements from another are still widely observed in both developing and developed countries [3, 4]. This may hinder the way of youths entering the labour market, causing the fall of education-related expectations, or make people inactive (out of employment, education, and not looking for a job).

From the employer's perspective, successful worker's qualifications and skills should be on the level required for the job. For the potential employees, education quality means competitive advantages. Most of the approaches to discover the real needs of the market use per-area employer and worker surveys. Conducting such polls takes certain time and resources, and couldn't ensure complete coverage of the labour market.

To discover both qualitative and quantitative correlations between education and labour market in a reasonable time, we proposed an intellectual system to monitor the demands of employers and match them with the standards and programs of higher education [5]. As a source of the real-life market needs, it was decided to use job advertisements from job search resources on the Internet (job hunting sites, state and city employment offices, etc.). For the education, texts of the state educational standards along with university educational programs are involved.

2. Links between the needs of the labour market and professional education

At the moment, Russian economy is characterized by the discrepancy between the quantitative and qualitative structure of graduates of universities and colleges to the needs of the labour market. Low level of graduates' employment is related with an imbalance of supply and demand in the labour market, quality of education and trainings, the mismatch of competencies of graduates with the requirements of the employer, as well as various social factors. As for the "relevance" of university graduates in the labour market, according to the portal *"Career.ru"*, in 2014 the list of "Top-20" Russian universities whose graduates were most in-demand, were two universities from St. Petersburg, the others – from Moscow [6]. This fact emphasizes the edge of the regional aspect of the problem. The analysis was conducted based on the search queries of employers. In 2016 *"Career.ru"* has published the rating of departments of the Moscow universities in eight vocational areas. In the ranking was used the data of graduates of 2015-2016 of faculties/departments of moscow universities who posted their applications on the website *"Career.ru"* (the youth branch of *HeadHunter* Internet-portal [7]). It was estimated the real demand for graduates basing on the actions of the applicants/graduates (profile of placement) and employers (invitations to interviews, salaries, which invited alumni and its comparison with General market salary) on the website *"Career.ru"*. The results of the research available at [8], and methodology is also published [9].

According to the research company MAR Consult, studied whether people are pursuing the profession obtained in university, the majority (52%) of poll participants are not. The survey was conducted in Moscow, St. Petersburg, Ekaterinburg, Nizhny Novgorod and Samara [10].

The problem of forecasting of economic development and educating of relevant professionals is challenging for many countries including European ones, where also becomes more popular the researches of the market's needs for skills on a regional and local levels, as well as for individual enterprises. The analysis of the experience in forecasting of the demand for qualifications in the EU shows that there are no elaborated unified system approaches to the analysis of the labour market from the perspective of changing requirements of the qualifications of the workforce and revealing of future needs of the labor concerning educational programs' content [11].

Making effective prognosis of skills requirements on the labour market is only possible on the basis of an objective assessment of the market. Scientific and practical interest to this problem is confirmed by the development of information-analytical systems intended for automation of data collection from popular recruitment services and its analysis aimed to identify the most demanded specialties and professions [12], estimation of the key status parameters of the labour market areas at the levels of districts and whole region [13].

From this perspective, it seems viable to develop and elaborate an automated information system to monitor the compliance between staffing requirements of the market and educational programs.

Do not include headers, footers or page numbers in your final submission. These will be added when the publication is assembled.

3. Competency-based approach to the description of graduate and specialist

The implementation of the competency-based approach to the training of university graduates in Russia is regulated by the Federal State Educational Standards of Higher Vocational Education , which are mandatory for all state-accredited universities, and involves the formation of students' set of general cultural, general and special professional competencies. Competence are interpreted as:

• ability to apply knowledge, skills and personal qualities for successful work in different professional situations;
• integral measure of interdisciplinary education quality.

Professional competencies are organized by activity types. As a professional integrity is meant the level of mastery of competencies, degree of readiness to apply the competencies in professional activities. For the implementation of the Federal Educational Standards in the relevant field of study, educational institution develops the principal professional educational program which includes educational plan, training schedule, working programs of disciplines (modules) and practices, instructional materials and other components. Planned results of capturing of the educational programs (competencies) are listed in the general description of the educational program. As a result, from the side of the educational system there are available the wordings of the competencies' contents.
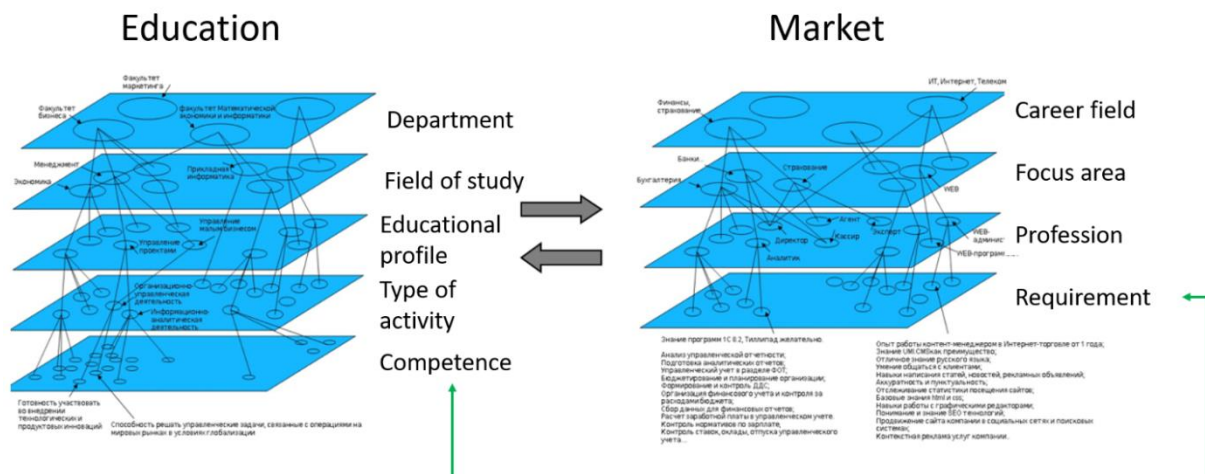

Figure. 1 Mutual mapping between models of the education system and the labor market at different levels of hierarchies

From the point of view of professional activity, it is possible to talk about a competency-based model of a specialist as a subject of demand in the labour market. This model is more complicated to describe because employers are not restricted to the formal framework while formulating job advertisements. As mentioned above, it is expected that the approved professional standards can be a link between the requirements to the qualifications from the market's perspective and requirements to the learning outcomes of education.

The idea of describing the subject matter in a hierarchical model (figure 1), which is a directed graph, whose vertices correspond to the domain objects and edges specify relations between them, was adopted from the work [14]. Models built on this principle, the model allow to correspond market's requirements and competencies at various levels, based on the link between the lower levels – competencies and requirements.

4. Linking market requirements with educational competencies

As is known, modeling of the semantics (meaning) of the word is one of the key problems related to natural language processing. The results of the semantic analysis are used in search engines [15], automatic translation systems and other fields related to natural language text processing [16].

At the moment in the approaches of vector representations of words (word embedding), the leading place is taken by the so-called predictive model based on the use of neural networks [17]. One of the principal tools for vector representation of words is *word2vec* [18].

The basic principle of *word2vec* is to find relations between contexts of words according to the assumption that words that appear in similar contexts tend to indicate similar things (that is, being semantically close). The problem solved by *word2vec* could be formalized as following: to minimize the distance between the vectors of words that appear next to each other, and maximize the distance between the vectors of words that appear quite far. "Near" 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 their meaning. Analysis of contexts is performed on large corpora of text, in our task we used the corpus of the Russian Wikipedia and a national corpus of the Russian language, as well as models of distributional semantics RusVectōrēs [19].

There are attempts to create a predictive model for the translation of a document to a vector space [20]. However, the task of comparing short sentences on the similarity of meaning has certain characteristics and the use of existing models for translating words or documents to a vector space, without modifications gives an unsatisfactory result.

Considering that the text of the language of educational competences, as well as the wording of the requirements in the vacancy announcements, contain an average of about 10 words, the task of evaluation of the semantic closeness of two short sentences is in the basis of the analytical part of the system. Authors have developed the algorithm [21] of sentences translation to vector space based on *word2vec*.

Thus, each word is mapped to a vector of dimension *n* (this parameter affects to the accuracy of the model). Metric space of mappings of words is used to be called *semantic*. Projections of the vectors of the words close by meaning are close together as well and form some semantic clusters.

Vector representation allows to calculate the "similarity" of words based on the calculation of cosine distance.

So, for two words w_1 and w_2, represented in the form of vectors $\vec{V}(w_1)$ and $\vec{V}(w_2)$, the semantic closeness is calculated by the formula:

$$cos\big(\vec{V}(w_1), \vec{V}(w_2)\big) = \frac{\vec{V}(w_1) \times \vec{V}(w_2)}{|\vec{V}(w_1)| \cdot |\vec{V}(w_2)|}. \qquad (1)$$

By analogy with the calculation of the similarity of words, it is calculated the semantic proximity of the competencies and requirements, which are short statements having 10 words in

average. The calculation of the vector of described sentences $\vec{V}$ *(s)*, where $s = \{w_1, w_2, ..., w_k\}$, is defined as a weighted average of vectors of the constituent words:

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

Where p_i is the weight of a word, which is calculated as the ratio of the frequency of use of the word to the dimension of the lexicon of the selected level of the hierarchy on the side of the education system or labour market,

k is the number of words in a sentence.

Then it is calculated the semantic proximity of the sentences using the formula given above. It is worth noting that words that have no particular meaning (conjunctions, particles, prepositions, pronouns and so on), do not participate in formation of the vector for the sentence.

5. Prospects of the approach's development

Due to the fact that the compared sentences have are narrow focused, and the Russian Wikipedia and a national corpus of the Russian language cover a vast number of areas and activities, the model becomes quite blurred with respect to the problem. This is mainly manifest itself in the lack of vectors for some words or their variations. To partially eliminate this effect, it was decided to make the two-level model: the second level represents the same comparison algorithm as described above, however, it does not work with words, but with their stems, that is, with their unchanging parts. Authors suppose that the accumulation of the vacancies database could allow forming of a unique corpus, taking into account the special terminology of the labour market by industries, which can then be used to train models. Also in the course of accumulation of statistics it is planned forecasting of demand for various specialities and individual qualifications in relation to professions.

It is also worth to point that the confirmation of the adequacy of the results of the comparisons is possible with the use of expert knowledge, however, the volume of the received results testifies an actual inability to fully verify them by a human within a reasonable time. Therefore, the authors are developing the methods that will allow automate the verification of this model.

6. Automated monitoring system for the labour market

The aim of implementation of the information systems for monitoring and forecasting the situation on the labour market and analysis of staffing requirements is to provide additional opportunities to identify qualitative and quantitative relationships between education and the labour market.

The system is developed for a wide range of users and is intended primarily for heads of regions, universities, companies, recruitment agencies. It is expected that the project will provide a tighter link between educational system in the country and the labour market, will give the opportunity to adjust curricula, to open new educational programs or to adjust the existing ones in accordance with the economic objectives of the regions, to implement efficient recruitment and training. After that, it is assumed that the system will become a useful tool for young professionals, starting seeking for a job in their chosen profession, and also the persons trying to choose their professionalization.

As a data source on vacancies, the following internet resources are used: portal "Work in Russia" (information website of the Russian labour agency), portals of staffing companies *HeadHunter* and *SuperJob*. As the governing documents the registry of approved professional standards and Federal state educational standards of higher education are used [22]. The subject of a separate study is evaluation how complete do the job ads represent the real demands of the market.

Implemented prototype of the automated information system is a web-oriented application with an intuitive user interface, ensures reliable data storage.

The system is built on a modular principle and include, first, the module collecting textual data (operating in automatic mode with the use of open sources, which are Internet portals and recruitment agencies).

Second, the load module and data storage, consisting of a distributed data store (provides replication and archiving).

Third, the automatic processing module performing the preparation of information for analysis, automatic linking requirements and competences, and machine learning.

Fourth, a user interface to generate and display reports based on business intelligence technologies.

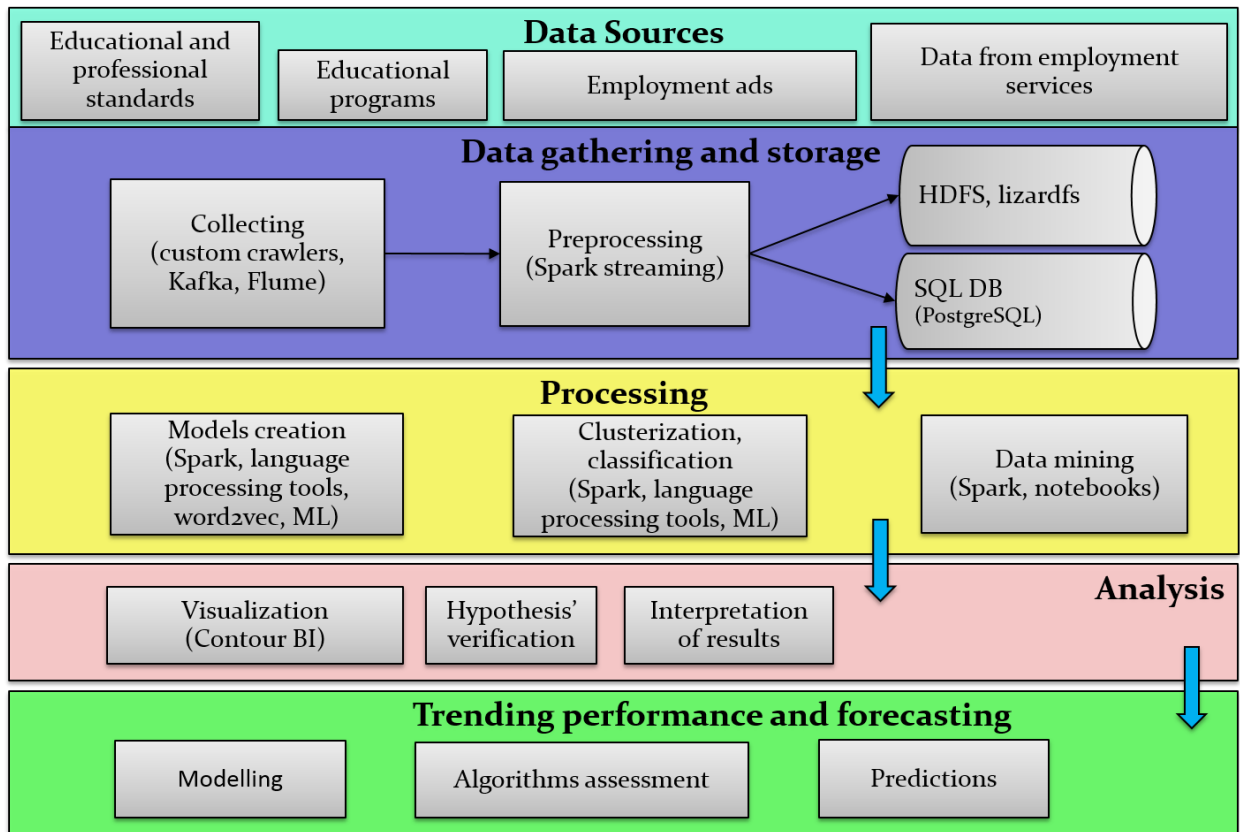General scheme of data processing is shown in figure 2.

Figure 2. Information workflow in the labour market monitoring system

7. Conclusion

Most approaches to identify the real needs of the market primarily used surveys among employers and employees. Conducting such surveys requires a certain amount of time and resources, and cannot provide full coverage of the labour market. To identify both qualitative and quantitative correlations between education and labor market within a reasonable time of proposed intelligent system of monitoring of the needs of employers and the analysis of their compliance with existing professional and educational standards. Results of this analysis may be recommendations for changes in educational programs.

In the framework of the project it was created a prototype of an automated information system for monitoring and analysis of employment needs of regions and identify the relationship to market demand with educational and professional standards. The system included in the software and technological solutions to the Situation centre for socio-economic development of Russia and subjects of Federation.

With this system the analysis of the constantly updating large amounts of data it is possible to determine how the training programs of higher education correspond to current market expectations, to anticipate changes in those expectations and automatically provide recommendations on adjustment of training programs to the most exact conformity to these expectations. Development and adaptation of the system can be carried out in accordance with the requirements of the customer depending on the specifics of the task – characteristics of the region, university, etc. We believe that the created system, and the algorithms and principles which it is based on, can be used to solve a wider class of topical challenges. For this, the system can be reconfigured depending on the peculiarities of the task statement and the nature of input data.

References

[1] J. Dolgado et al., No Country for Young People? Youth Labour Market Problems in Europe, 2015

[2] European Commission. Labour Market and Wage Developments in Europe. Annual Review, 2016

[3] European Commission. From University to Employment: Higher Education Provision and Labour Market Needs In the Western Balkans. Synthesis Report, 2016

[4] A. Wolf, Review of Vocational Education – The Wolf Report, 2011

[5] P. Zrelov, Automated system of monitoring and analysis of staffing needs for the nomenclature of specialties of the university (in Russian), "Federalism" journal,
№4 (84), 2016

[6] https://career.ru/article/15115 (in Russian)

[7] HeadHunter staffing agency: https://hh.ru/

[8] http://mel.fm/2016/10/21/rating_career (in Russian)

[9] http://mel.fm/2015/10/21/metod_career (in Russian)

[10] Pogorelov E., The problem of demand for university graduates on a contemporary labour market, Proceedings of V International student electronic scientific conference, «Student scientific forum» 15 February - 31 Marchs 2013 (in Russian)

[11] Oleynikova O.N., Muravjeva A.A., Forecasting of needs for skills and vocational education and training – the EU experience, The Center for the study of problems of vocational education - http://www.cvets.ru/Modules/SNA-EC.pdf - 22.07.16. (in Russian)

[12] Cheremisina E.N., Belaga V.V., Samoilenko Yu.I. Informational and educational environment for teaching information technologies on the basis of the Institute for System Analysis and Management of the University "Dubna" // "Open Education", 2/2014 - P. 59-65. (in Russian)

[13] Petrunina O.E. Designing of information-analytical system of management of the regional labor market, Modern science-intensive technologies. - 2005. - No. 5 - P. 75-78. (in English)

[14] Gushchin A.N., Providing an educational process built on the standards of the GEF-3, by means of information technologies // Educational Technology. 2013. № 4. P. 84-89. (in Russian)

[15] Efrati Amir. «Google Gives Search a Refresh». The Wall Street Journal. Retrieved July 13, 2012.

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

[17] Barkan Oren (2015). «Bayesian Neural Word Embedding». arXiv:1603.06571.

[18] Mikolov Tomas et al. «Efficient Estimation of Word Representations in Vector Space». arXiv:1301.3781v3 [cs.CL] 7 Sep 2013.

[19] Kutuzov A., Kuzmenko E. (2017) 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

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

[21] P. Zrelov et al., Monitoring of the labour market needs for university graduates based on data intensive analytics (in Russian), Proceedings of the XVIII International Conference DAMID/RCDL'2016, October 11-14, 2016, Ershovo, Moscow Region, Russia

[22] Professional standards in Russia: http://profstandart.rosmintrud.ru