




Better understanding of the labour market using Big Data

ALENA VANKEVICH

corresponding author

Vitebsk State Technological University, Moscow Av., 72, 210038 Vitebsk, Belarus


✉ vankevich_ev@tut.by

 orcid.org/0000-0001-8064-8250

IRYNA KALINOUSKAYA

Vitebsk State Technological University, Faculty of Economics and Business Administration,
Department of Management, Belarus

✉ i-kalinovskaya@yandex.by

 orcid.org/0000-0002-6622-2875

Abstract

Motivation: As the result of digitalisation of the economy, the number of Internet users is increasing, which leads to an increase in the number of vacancies posted on online platforms and services. The description of vacancies includes information about skills and competencies, which is the source of additional data for the labour market analysis. This information cannot be received through the analysis of statistical and administrative data. Therefore, it is important: — to learn how to evaluate new information sources, and use the data they generate; — to develop tools that people and organizations will use for finding an employee or a vacant post. The study focuses on the analysis and forecast of labour demand in the context of skills and competencies, which significantly enriches and adds to the information about the labour market and facilitates effective decision-making.

Aim: The main goals of this article are the following: (1) identification of the methodological approaches in the labour market analyses using Big Data; (2) assessment of the labour demand and labour supply in the context of skills and competencies listed in the vacancy description posted on job portals; and (3) determination of the matches (mismatches) between skills and competencies in order to help the companies and individuals get better employment and education. Empirical data used in the research were collected from



the description of job vacancies (16 401 vacancies) and CVs (227 215 CVs) from the most popular open job portals in Belarus through the scraping approach and classified according to the ESCO and ISCO codes. Quantitative analysis by the means of artificial intelligence was used in the research.

Results: The study results revealed that the information about the volume and structure of skills and competencies obtained by scraping data from vacancy descriptions and Cvs, which are posted on online portals, allows for more precise diagnostics of labour demand and supply and overcoming of bilateral information asymmetry in the labour market. Based on the analysis, the parameters of scarcity and excess in competencies for individual occupations in the labour market are determined (the level of the correlation ratio between applicants' competencies and those requested by employers in the context of occupations (four digits according to the ISCO classification) is less 0.8; the deviation of the ranks of competencies listed in CVs and vacancy descriptions according to the ESCO groups of skills/competencies and a sign of revealed deviations). The methodology is developed to set areas for necessary knowledge acquisition (by the analysis of competencies listed in CVs and vacancy descriptions at the 3rd and 4th digit level of ISCED classification) and skills (by the analysis of competencies at the 2nd digit level in ESCO groups). The paper illustrates limitations in using Big Data as an empirical database and explains the measures to eliminate those limitations.

Keywords: labour market; skills; competencies; Big Data; open-job portals

JEL: J01; J24

1. Introduction

New opportunities for labour market analytics and decision making have emerged as a result of the increasing volume of information and the ability to aggregate Big Data. The digitalisation of the economy has led to an increase in the number of various information sources where vacancies are posted. These are not only online portals but also websites of companies and recruitment agencies, professional networks (such as LinkedIn), online job search platforms (Upwork, Uber), educational platforms, social networks (Facebook, Twitter), websites of newspapers and announcement boards, university centres for promoting graduate employment, etc. Collectively, these sources expand the possibilities for labour market research and provide a means to measure what was previously unmeasured or unmeasurable (Horton & Tambe, 2015). In other words, they contain a lot of additional information about the labour market that cannot be obtained from statistical or administrative data sources (e.g. household employment surveys or information from the national employment service). Thus, it will be very important to: — use new information sources, and analyse the data they generate; — combine these data with traditional resources (statistical, administrative data) in order to enrich information about the labour market and facilitate effective decision-making; — develop tools that people and organisations will use to find an employee or a job.

Analysis of vacancies posted on online portals is the first step in the methodology of using Big Data for the labour market. This generates general under-

standing of the dynamics of the digital labour market in the country, the types of economic activities and qualification segments it covers, and characteristics of vacancies posted and curriculum vitae (CVs) (Mezzanzanica & Mercurio, 2019). Artificial intelligence technologies make it possible to process large amounts of data, and a number of studies illustrate benefits from using them for the analysis of the labour market (Bejger & Elster, 2020; ILO, 2020; Vankevich & Kalinouskaya, 2020). This will create an opportunity to assess the demand for labour in the context of necessary skills and competencies and to determine the emergence of new professions (or, alternatively, the disposal of obsolete ones) through the dynamics of competencies and varieties of their combinations. Also, this allows to process large amounts of currently scattered information about a potential candidate for a vacancy, to consider more precisely all the available information about the candidate (from social networks, instant messaging platforms, search engines, special databases). Overall, this will reduce bilateral information asymmetry in the labour market and increase the transparency of decisions made for all participants (ILO, 2020; Vankevich & Castel-Branco, 2017).

Special attention should be paid to the forms for submitting the skills and qualifications required by employers, which contain templates for job descriptions, as well as the forms for submitting CVs (which also indicate skills and qualifications, personal and business skills, etc.). These forms allow the data to be processed using Big Data analytics. Analysis of these data helps to determine the labour demand and supply in the context of skills and competencies.

Alongside the expanding opportunities for analytics, Big Data creates new challenges, solutions to which are offered in this paper, including access to data (by a means of scraping), the quality of data (on the basis of their cleaning, deduplication and a unified way of their structuring), culture and ethics in publication and the use of personal data (by a means of arranging a single data array which excludes personal information).

2. Literature review

The classical school of thought sees the demand in the labour market as open vacancies and the supply as the labour force (the total of both the employed and the unemployed) (Borjas, 2015, p. 21, 84, 144), where the measures are to be taken at the market and institutions levels in order to maintain the equilibrium between the supply and demand. While education (investment in human capital) plays an important role in the labour market equilibrium. “... We stress the idea that our educational and training decisions play an important role in the determination of earnings” (Borjas, 2015, . 230). Education performs a signal function in the labour market. Under current conditions both employers and employees focus more on a set of specific skills (from various training areas) which an employee has and which will be demanded for new occupations, rather than on the level of education and a speciality received. Thus, if education

plays such an important role in the labour market equilibrium, there is a need for the indicators to determine the skill relevance and assess which skills give most benefits to people who have them. One of the key features of the labour market is bilateral information asymmetry. From a theoretical point of view, effective search and matching of the labour market requires reliable information about vacancies and employees, in particular their quantity, structure and quality characteristics. This simplifies the process of recruitment or job search.

In the economics literature the process of recruitment and hiring is largely viewed from an employee's perspective (Oyer & Schaefer, 2011). Information from the national employment agency or job ads placed by employers was used as an empirical database of the number and quality of vacancies. (CEDEFOP, 2019b, pp. 47–48; UK Commission for Employment and Skills, 2014, p. 3). The number of vacancies is used to characterize the unrealised demand (Kureková et al., 2015). However, there is no exact total number of vacancies. It is different on different portals, and there are duplications. To characterise the volume of supply in the labour market the number of the employed and the unemployed is used (Borjas, 2015). Whereas LFS (results of labour force surveys) characterises, to a greater extent, labour force supply rather than its demand and in terms of the number of employees instead of skills and competencies.

According to some studies, about 55–81% employees, who were newly hired using the information from the Mass Media, the Internet and social networks, receive additional training in the company (Roshchin & Solntsev, 2017, p. 183), which indicates the low efficiency of matching.

Thus, the problem of demand and supply match lies in matching of vacancy requirements to the applicant's skills rather than in matching the number of open vacancies to the amount of labour force. These studies are limited by their starting point which is a definition of the vacancy (it is viewed as a job, i.e. workplace for one employee). Moreover, there is no analysis of the requirements to applicants, their skills and competencies match. However, their combinations create many options of new vacancies and therefore demand for various employees (or create employment opportunities for various employees). To reduce the information asymmetry when looking for a vacancy of an employee the attention should be paid to the employee's set of competencies rather than to their level of education or the profession received. (Makovskaya, 2020; Vasiliev et al., 2016). Statistical and administrative data are not enough for the analysis of skills and competencies; it requires Big Data analytics.

3. Big Data for labour market analytics

Online job portals can serve as a source for Big Data on the labour market (CEDEFOP, 2019a; 2019b; ILO, 2020; Mezzanzanica & Mercorio, 2019, p. 14; World Economic Forum, 2018, p. 4). They have a number of advantages over administrative databases in terms of cost and the ability to form different target samples (CEDEFOP, 2019b; Horton & Tumbé, 2015). The empirical database

for the analysis of data from the online portals is developed by a means of scraping, that is, collecting data on vacancies from the online sources with the help of a special application (scraper). The potential of this method to develop the empirical database for the research is described in works within the Burning Glass Research (CEDEFOP, 2019a; ILO, 2020; Mezzanzanica & Mercorio, 2019, p. 14; World Economic Forum, 2018, p. 4).

Scraping data has proven effective for conducting research with a specific purpose, that is, to form a database and test any hypothesis on it. The literature presents the results of individual research projects using Big Data collected from online job portals and other internet sources:

- The Burning Glass Research — the most feasible transitions and required skills were revealed by a means of analysis and clustering of jobs and skills. The data were collected by scraping from over 40 thousand unique online resources (over 50 million vacancies for 2016–2017). The skills were grouped into 550 clusters, and 958 types of jobs according to the O*NET classification and by their growth rate (O*NET — Occupational Information Network — standard taxonomy of professions in the USA);
- Where the Works.org Project (UK) is a model of demand and supply analysis in the labour market for jobs that require mid-level qualifications (Mezzanzanica & Mercorio, 2019, pp. 42–43). The project resulted in creation of the online portal that revealed discrepancies between the applicants' qualifications and the employers' requirements to the jobs that require mid-level qualifications. However, the analysis was done in the context of jobs but not competencies;
- Kureková et al. (2015) were a pioneers in the usage of the European-wide publicly administered job-vacancy portal EURES. They perform content analysis and carry out a comparative study of employers' demand for skills demand in small European economies. They have revealed that the mix of required skills is very diverse across the countries, implying that there is no universal set of requirements and also that domestic institutions and structures strongly affect how demand is formulated.

Thus, based on empirical results and methodological limitations, the research objectives can be defined as follows:

- the analysis of labour demand will help to determine the most required skills and competencies; the analysis is based on the classification of skills and competencies which are mentioned in the vacancy description and collected from the online portals;
- the analysis of labour supply will help to determine the most dominant skills and competencies among applicants; the analysis is based on the classification of skills and competencies listed in the applicants' CVs and collected from the online portals;
- comparison of labour demand and supply in the context of skills and competencies provides a more accurate image of matching (mismatching) between them. It serves as a basis for educational institutions to make decisions

about necessary areas of training, and for employees to make decisions about self-training and education that will be duly adopted to employers' requirements.

It should be noted that, in this way, the state of the labour market, the scope of requirements to the applicants, and applicants' set of skills and competencies will be significantly different from those presented by the national employment service or labour force surveys.

4. Labour demand and labour supply analyses with Big Data: empirical data, method, results

The empirical database of the research is represented by the data collected by the means of scraping from vacancies descriptions and CVs posted on the leading online job portals of Belarus. In the course of research the collection of data was analysed that included: 16 401 vacancies (89 864 competencies) and 227 215 CVs (1 569 290 competencies), in total 1 659 154 competencies from vacancy descriptions and CVs have been analysed (Table 1).

The applications specially developed on the basis of artificial intelligence were used as a tool for extracting the necessary digital data. The applications meet the requirement to the processing speed and volume of information and provide highly accurate results.

The process of extracting, processing and analysis of competencies includes the following stages (Scheme 1):

- collection of CVs and vacancy descriptions from various web sources. To collect CVs and vacancy descriptions the programme modules were developed on the basis of Scrapy framework that automatically perform collection and initial processing of the unstructured data according to the schedule with the help of Airflow task scheduler;
- processing of the extracted data that includes tokenization, removing of stop-words and punctuation marks, standardization, stemming and lemmatization;
- deduplication of data using artificial intelligence that helps to compute the similarity of text documents by their vector representation (Mezzanzanica & Mercurio, 2019; Vankevich & Kalinouskaya, 2020);
- extracting and classification of positions from CVs and vacancy descriptions according to ISCO (International Standard Classification of Occupations). USE (Universal Sentence Encoder) is chosen as a model classification as it supports multiple languages and encodes entire sentences which provides for the required accuracy of processing and comparison of text information;
- extracting and classification of competencies from CVs and vacancy descriptions according to ESCO (European Skills / Competences, Qualifications and Occupations);
- transferring of processed and classified data into ClickHouse Database and their visualization with the help of Superset.

To determine the matching degree of vacancy requirements and applicants' CVs three attributes — competency, profession, and type of economic activity — were analysed by a means of correlation analysis. Determination of the matching degree of proposed and demanded skills considering the type of economic activity involved estimating the presence, power and direction of the correlation among the ranked competencies in the context of occupation-and-qualification groups by ISCO (1 digit). The nonparametric analysis was applied to the ordinal scales and the Spearman's rank correlation coefficient was calculated which is used to calculate the coefficient of rank correlation and does not require a normal law of the correlating series distribution (Table 2).

The analysis of the coefficient of correlation by ISCO occupation groups revealed the following:

- six of nine occupation-and-qualification groups demonstrate a high extent of direct correlation between competencies listed in CVs and those in vacancy descriptions (rank correlation coefficient is over 0.9);
- two occupation-and-qualification groups (plant and machine operators and assemblers, elementary occupations) show rather high, directly proportional correlation between competencies listed in CVs and those in vacancy descriptions (rank correlation coefficient is 0.8–0.9);
- the occupation-and-qualification group “skilled agricultural, forestry and fishery workers” has moderate direct correlation (rank correlation coefficient is 0.415).

The hypothesis about a zero Spearman's rank correlation coefficient has been rejected in all ISCO occupation groups which indicates a statically significant rank correlation coefficient and significant correlation. Thus, the analysis of matching the competencies listed in CVs and those in vacancy descriptions in the context of major occupation-and-qualification groups shows an entire picture of demand and supply correlation in the labour market and reveals directions that have the greatest imbalance (e.g. jobs for workers or in the agriculture). However, it cannot detect new skills and determine the extent of matching (mismatching) between skills which are the most/least required and those which are most/least proposed in CVs in the context of specific specialities and areas of training. Due to this fact, the analysis was deepened in the context of most demanded occupations (to the 4th digit by ISCO classification) (Table 3 shows calculations for the first 3 most demanded occupations).

The analysis of rank correlation coefficients for competencies listed in vacancy descriptions and CVs within the group of most demanded occupations revealed that:

- there is a direct correlation between competencies listed in CVs and those in vacancy descriptions for the “accountant” occupation (rank correlation coefficient is over 0.8), which signals about adequacy of the training system in terms of employer's requirement to these specialists;
- there is a moderate direct correlation between competencies listed in CVs and those in vacancy descriptions for the “sales engineer” and “software

developer” occupations (rank correlation coefficient is 0.6 and 0.5 correspondingly), which signals about the necessity to upgrade academic curricula and syllabi for these specialties as they do not train to all the necessary competencies.

Thus, the analysis of matching the competencies listed in CVs and vacancy descriptions across various occupations (4 digits by ISCO classification) revealed profiles with some mismatching (“sales engineer” and “software developer”). However, this analysis failed to reveal the extent of mismatching between the required skills and those listed in CVs in the context of specific skills. That is why the analysis was done in the context of ESCO groups of skills/competencies for these occupations.

5. The discussion of the results and recommendations

The analysis of mismatching between the required skills and existing ones according to the ESCO groups of skills/competencies was done by ranking the competencies listed in vacancy descriptions and CVs, determining the value and the sign of deviation in the ranks and calculating the percentage of the deviation. To visualize the results, Chart 2 display the rank of each competency listed in either vacancy descriptions or CVs. The competency is classified by ESCO groups of skills/competencies (corresponding color on the chart). The difference in the size of the corresponding colored segments signals about mismatching between the required skills and existing ones. The rank of a competency listed in the vacancy description was used as a basic component for the analysis of mismatching. If the size of the segment for the competency listed in a CV was smaller than that in a vacancy description, the resulting mismatching was related to the scarcity of competencies in the ESCO group of skills/competencies. If the size of the segment for the competency listed in a CV was larger, the resulting mismatching was related to the excess of the competencies.

Going further, the size, which is presented as percentage of deviation in ranks of skills listed in vacancy descriptions and CVs, and the sign of this deviation are determined for the mismatching in ESCO groups of skills/competencies. The analysis of the matching between ESCO groups of skills/competencies within specified occupations revealed the groups of skills with mismatching and classified them on the basis of scarcity or excess of competencies.

1. the scarcity of competencies listed in applicants’ CVs was detected in the following groups:
 - “knowledge” (3%) and Skills 1 “communication, collaboration and creativity” (2%) in the occupation “sales engineer”;
 - “knowledge” (14%) in the occupation “software developer”;
 - “knowledge” (4%) and Skills 2 “information skills” (5%) in the occupation “accountant”;

2. the excess of competencies listed in applicants' CVs was detected in the following groups:

- “attitudes and values” (3%) and Skills 2 “information skills” (3%) in the occupation “sales engineer”;
- Skills 1 “communication, collaboration and creativity” (5%) and Skills 5 “working with computers” (4%) in the occupation “software developer”;
- Skills 5 “working with computers” (3%) and Skills 6 “handling and moving” (3%) in the occupation “accountant”.

Also, the analysis revealed the necessity for the applicants to acquire knowledge in the occupations above according to the competencies included into the group “knowledge”. It was found that the areas of knowledge acquisition are determined with the help of analysis of competencies listed in vacancy descriptions and CVs within the 3rd and 4th digit group by ISCED (International Standard Classification of Education). The list shown in Table 4 presents areas and specialisations of knowledge acquisition (by ISCED classification) according to the detected mismatching of competencies listed in the group “knowledge” for the above mentioned occupations.

Missing skills within the group “communication, collaboration and creativity” for the occupation “sales engineer” and group “information skills” for the occupation “accountant” were detected using the data from ESCO S1 and S2 groups of skills/competencies respectively (Table 5). The detection of missing competencies in the context of ESCO groups was done by two digits (S1.x and S2.x).

Thus, the detailed analysis of the correlation between proposed and required skills and competencies grouped in accordance with ESCO classification within each occupation helps to reveal either the excess or the scarcity of competencies the applicants have in comparison to those demanded in the labour market (and listed in vacancy descriptions). And the analysis of the mismatching of competencies within an individual ESCO group of skills/competencies helps to determine the areas for training programmes offered to the applicants who need additional knowledge and skills.

Results of the research make it possible to determine the training areas for the competencies which:

- are listed as the most required in vacancy descriptions;
- demonstrate the greatest mismatching with the competencies listed in applicants' CVs. But in this case, the problems of representativeness of the sample and data quality are paramount (Horton & Tumbe, 2015). The economics literature suggests the following actions to address the problem of the representativeness of data on vacancies obtained from online job portals: — combine different data sources, for example, administrative data, company surveys (supplemented by interviews with representatives of HR departments).



Promising areas for further research can be the following:

- search for the correlation between the demand for skills and employability based on the combination of employee’s skills (listed in the CV);
- enrichment of the ESCO taxonomy by adding new skills into occupations and creating new combinations of them;
- development of a graphical database in countries that do not use the EU taxonomy to get the opportunity for correct benchmarking.

6. Conclusions

Big Data analytics is a promising tool, however, its results cannot be considered as pertinent and useful for decision making unless experts provide their support (ILO, 2020, p. 24, 33, 49). The analysis has helped to determine the following mismatches in the Belarus labour market:

- by the type of economic activity (in the context of occupation-and-qualification ISCO groups by the 1st digit) the greatest mismatch between the required and proposed competencies was determined within the group of occupations for workers and in agriculture;
- in the context of competencies within the group of the most demanded occupations the maximum mismatch between the required and proposed competencies was determined in the occupations “sales engineer” and “software developer” (the analysis at the 4 digits level of ISCO classification);
- in the context of ESCO groups of skills/competencies within the group of the most demanded occupations the scarcity of skills and knowledge was detected in the groups “knowledge”, “communication, collaboration and creativity” and “information skills”.

The research resulted in the development of the method of scraping and competencies analysis that helped to reveal scarcity of knowledge in the group “knowledge” — languages, business and administration, information and communication technologies (ICTs), social and behavioural sciences, law (the analysis was done at the 3rd and 4th digit level of ISCED classification); and scarcity of skills in the groups “communication, collaboration and creativity” and “information skills” — negotiating, liaising and networking, presenting information, promoting, selling and purchasing, information skills, conducting studies, investigations and examinations, managing information, analysing and evaluating information and data (the analysis was done at the two digits level of ESCO groups of skills/competencies).

Thus, the newly developed method of the labour market analysis with the help of Big Data enables to focus on the characteristics of vacancies (structure of skills and competencies listed in vacancy descriptions) rather than on their number which contributes to better understanding of matching between employers’ requirements and applicants’ competencies. It adds to the understanding of the situation in the labour market that is outlined by the national employment service on the basis of statistical data on jobs and information about

graduates employment provided by educational institutions. However, it should be noted that Big Data does not provide a complete description of an occupation. That's why the role of experts in the labour market is becoming greater.

References

- Bejger, S., & Elster, S. (2020). Artificial Intelligence in economic decision making: how to assure a trust? *Ekonomia i prawo. Economics and Law*, 19(3), 411–434. <https://doi.org/10.12775/EiP.2020.028>.
- Borjas, G.J. (2015). *Labor economics* (7th ed.). McGraw–Hill Education.
- CEDEFOP. (2019a). *Online job vacancies and skills analysis: a CEDEFOP pan-European approach*. <https://doi.org/10.2801/097022>.
- CEDEFOP. (2019b). The online job vacancy market in the EU: driving forces and emerging trends. *CEDEFOP Research Paper*, 72, 1–96. <http://data.europa.eu/doi/10.2801/16675>.
- Horton, J.J., & Tambe, P. (2015). Labor economists get their microscope: big data and labor market analysis. *Big Data*, 3(3), 130–137. <https://doi.org/10.1089/big.2015.0017>.
- ILO. (2020). *The feasibility of using big data in anticipating and matching skills needs*. Retrieved 04.11.2021 from https://www.ilo.org/wcmsp5/groups/public/--ed_emp/---emp_ent/documents/publication/wcms_759330.pdf.
- Kureková, L.M., Beblavý, M., & Thum-Thysen, A. (2015). Using online vacancies and web surveys to analyse the labour market: a methodological inquiry. *IZA Journal of Labor Economics*, 4, 18. <https://doi.org/10.1186/s40172-015-0034-4>.
- Makovskaya, N. (2020). *Modern features of the labor market functioning in belarus. Monograph*. A.A. Kuleshov Mogilev State University. Retrieved 04.11.2021 from <https://libr.msu.by/bitstream/123456789/12909/1/3345n.pdf>.
- Mezzanzanica, M., & Mercurio, F. (2019). *Big data for labour market information: an introductory guide*. Retrieved 04.11.2021 from <https://www.etf.europa.eu/sites/default/files/2019-06/Big%20data%20for%20LMI.pdf>.
- Oyer, P., & Schaefer, S. (2011). Personnel economics: hiring and incentives. In D. Card D., & Ashenfelter, O. (Eds.), *Handbook of labor economics*, 4B, 1769–1823. [https://doi.org/10.1016/S0169-7218\(11\)02418-X](https://doi.org/10.1016/S0169-7218(11)02418-X).
- Roshchin, S.Y., & Solntsev, S.A. (2017). How companies search employees: empirical evidence from Russian companies. *Russian Management Journal*, 15(2), 173–192. <https://doi.org/10.21638/11701/spbul8.2017.203>.
- UK Commission for Employment and Skills. (2014). *Information to local intelligence: guidance on local labour market assessments*. Retrieved 04.11.2021 from https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/346923/Local_LMI_Guidance_Aug14_FINAL_with_covers_v3.pdf.



- Vankevich, A., & Kalinouskaya, I. (2020). Artificial Intelligence technologies in human resource management. *Belarusian Economic Journal*, 2(91), 38–51. <https://doi.org/10.46782/1818-4510-2020-2-38-51>.
- Vankevich, A., & Castel-Branco, E. (2017). Information-analytical system of the labour market and forecasting of demand for personnel: content and directions of formation in the Republic of Belarus. *Belarusian Economic Journal*, 2(79), 73–92.
- Vasiliev, D.S., Roshin, S.Y., Solntsev, S.A. (2016). *Evolution of job search technologies in the labor market in the internet age*. National Research University Higher School of Economics. Retrieved 04.11.2021 from https://wp.hse.ru/data/2016/05/11/1129129931/WPI5_2016_02____.pdf.
- World Economic Forum. (2018). *Towards a reskilling revolution: a future of jobs for all*. Retrieved 04.11.2021 from https://www3.weforum.org/docs/WEF_FOW_Reskilling_Revolution.pdf.

Acknowledgements

Author contributions: authors have given an approval to the final version of the article. Authors contributed to this work as follows: A.V. developed the concept and designed the study, I.K. collected the data, A.V. and I.R. analysed and interpreted the data, A.V. and I.R. prepared the draft of the article, A.V. revised the article critically for important intellectual content.

Funding: this research was undertaken as part of the state program of scientific research of the Republic of Belarus *Society and humanitarian security of the Belarusian state* under the title *Digitalization of the labor market in the innovative economy of the Republic of Belarus: directions, methods of analysis, regulation, forecast*. Some of ideas were initiated under ETF project *Big data for labour market information* and TOR project *Landscaping of the Web Labour Market in Belarus, and ranking of online job vacancy sources (websites)*. This research was fully funded by a grant (FTF.94906).

Supplementary information: authors acknowledge following FTF for help in understanding the possibility of using Big Data in the labor market.

Note: the results of this study were presented at an webinar *Labour market information in transformation focus on big data for LMIS* (December, 10, 2020).



Appendix

Table 1.
Characteristics of the research empirical database

ISCO Group	Vacancies, items	Competencies listed in vacancy descriptions, items	CVs, items	Competencies listed in CVs, items
1. managers	1 450	9 290	23 500	191 000
2. professionals	8 560	52 700	99 900	852 000
3. technicians and associate professionals	868	4 530	9 310	63 000
4. clerical support workers	714	3 420	9 290	61 700
5. service and sales workers	2 500	12 500	49 800	277 000
6. skilled agricultural, forestry and fisheries workers	8	24	75	190
7. craft and related trades workers	1 030	3 340	11 900	44 100
8. plant and machine operators and assemblers	605	1 950	17 400	61 800
9. elementary occupations	667	2 110	6 040	18 500

Source: Own preparation.

Table 2.
Spearman's correlation coefficient between competencies listed in vacancy descriptions and CVs in the context of occupation-and-qualification ISCO groups (1 digit)*

Occupation-and-qualification ISCO group (1 digit)	Rank correlation coefficient	Critical point
1. managers	0.951	0.07
2. professionals	0.954	0.06
3. technicians and associate professionals	0.914	0.09
4. clerical support workers	0.945	0.07
5. service and sales workers	0.944	0.07
6. skilled agricultural, forestry and fisheries workers	0.415	0.29
7. craft and related trades workers	0.909	0.09
8. plant and machine operators and assemblers	0.835	0.12
9. elementary occupations	0.815	0.13

Notes:

* level of the coefficient significance — 0.05; t-Student coefficient — 2.27; connection strength was determined using the Chaddock scale.

Source: Own preparation.



Table 3.

Spearman's correlation coefficient between competencies listed in vacancy descriptions and CVs in the context of three most demanded occupations by ISCO (4 digits)*

Occupation	ISCO group and code	Number of vacancies	Number of competencies listed in vacancy descriptions	Number of CVs	Number of competencies listed in CVs	Rank correlation coefficient	Critical point
sales engineer	"professionals" code 2433	2 070	15 000	20 800	181 000	0.589	0.36
software developer	"professionals" code 2512	1 240	7 150	7 410	66 200	0.541	0.38
accountant	"professionals" code 2411	451	2 270	9 730	76 600	0.836	0.25

Notes:

* level of the coefficient significance — 0.05; t-Student coefficient — 2.368; connection strength was determined using the Chaddock scale.

Source: Own preparation.

Table 4.

Areas and specialisations of knowledge acquisition (by ISCED classification) according to the detected mismatching of competencies from the group "knowledge" listed in vacancy descriptions and CVs

Occupation	Area (by 3 digits in ISCED)	Specialisation (by 4 digits in ISCED)
sales engineer	023 languages	0231 language acquisition
	041 business and administration	0414 marketing and advertising
		0416 wholesale and retail sales
software developer	023 languages	0231 language acquisition
	061 information and communication technologies (ICTs)	0613 software and applications development and analysis
	041 business and administration	0413 management and administration
accountant	031 Social and behavioural sciences	0311 economics
	041 Business and administration	0411 accounting and taxation
	042 Law	0421 law

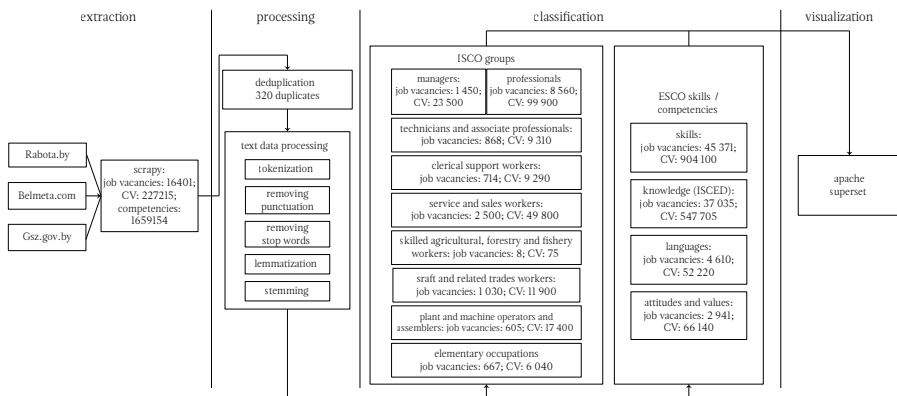
Source: Own preparation.

Table 5.
The list of missing skills in the context of the ESCO groups S1 “communication, collaboration and creativity” and S2 “information skills”

Occupation	Skills
sales engineer	S1.1 — negotiating
	S1.2 — liaising and networking
	S1.4 — presenting information
	S1.6 — promoting, selling and purchasing
accountant	S2.0 — information skills
	S2.1 — conducting studies, investigations and examinations
	S2.3 — managing information
	S2.7 — analysing and evaluating information and data

Source: Own preparation.

Scheme 1.
Stages of extracting and processing data from online sources

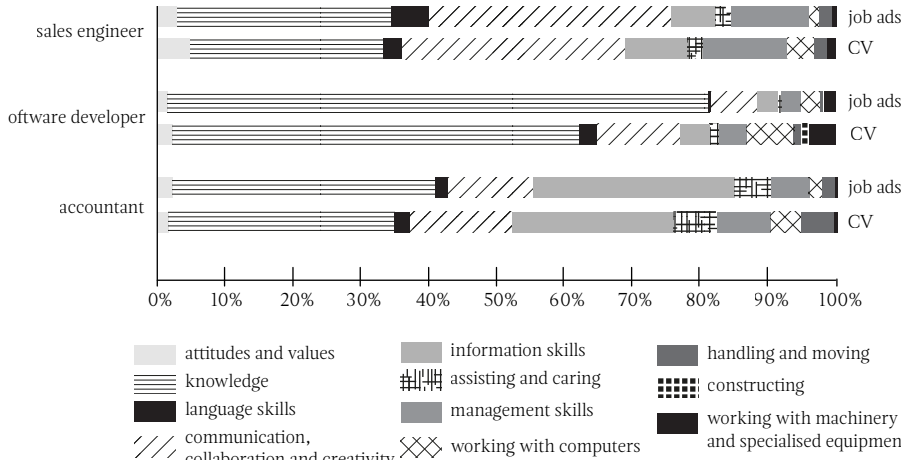


Source: Own preparation.



Chart 2.

The analysis of mismatching between the proposed and required skills by ESCO groups of skills/competencies



Source: Own preparation.