DISCOVERING JOB KNOWLEDGE FROM WEB DATA

Intellectual Output 5 (Report)

Job Knowledge Base Assessment and Reporting

Co-funded by the European Union

Erasmus+
Enriching lives, opening minds.

Content

Project DISKOW:

Discovering job Knowledge through Web Analytics towards facilitated mobility of European Professionals and Refugees Career Integration

https://www.diskow.eu/

An ERASMUS+ Project, Grant agreement n. 2018-1-DE02-KA202-005215
KA2 - Cooperation for Innovation and the Exchange of Good Practices
KA202 - Strategic Partnerships for vocational education and training Strategic Partnerships for vocational education and training

Priorities

HORIZONTAL: Transparency and recognition of skills and qualifications
HORIZONTAL: Development of relevant and high-quality skills and competences
HORIZONTAL: Open education and innovative practices in a digital era

It is currently a big challenge to discover required competences to specific jobs, update and adapt education as well as training programs in this regard and also provide further tools and support for strengthening the basic and key competences. Job seekers face difficulties in better understanding of the required job knowledge and competences through reviewing job profiles. Moreover, public and private sectors need IT-based tools to simplify transparent recognition of domain specific job knowledge while setting up their job profiles.

In comparison to traditional data sources, web data offer a range of advantages. They are available in real time and allow researchers to compile large, detailed and diverse datasets in an easy, fast, flexible and relatively inexpensive way (covering topics for which traditional sources are absent or weak; examples are self-employment and on-the-job search). In this regard, DISKOW discovers job specific knowledge form the web data and provides further recommendations and job knowledge base (JKB) that could strengthen the key competences of job seekers. As a result, job seekers will be able to use the JKB transparently to provision and develop their domain specific skills and competences. The JKB presents an architecture pipelining information from online job vacancies to the end-users in an easily accessible way, allowing comparisons between regions, occupations and in time.

Introduction

This Intellectual Output is the final IO of the DISKOW project and covers the activities carried out to assess the final prototype of the Job Knowledge Base platform. The first chapter is devoted to the analysis of the online job vacancies in the labour market, followed by the second chapter, which describes the JKB prototype as a final outcome of the project. The third and fourth chapter are focused on the assessment activities. To assess the accuracy of the data extraction, a sample of random job postings was selected and the extracted data were manually compared with the original text published online. The results were analysed and potential improvements, which could be achieved by the future users identified. Further, external stakeholders were approached and interviewed with the aim to evaluate the user experience with the JKB prototype. The answers obtained are analysed and summarized in the chapter 4 of the IO5.

1. Online job vacancies in labour market analysis

According to the European Commission (EC), proper development and use of job knowledge, competencies and skills are key factors for success in finding the job and keeping it [Bruges Communique (BC), 2010[1]]. In the last few years, substantial efforts have been made by governments, international organisations, and other institutes to improve our understanding of the labour market dynamics. These efforts have resulted in a series of applications, tools, and sources that cover labour demands, supply or matching, and shed more light on skill gaps and mismatches. ESCO (European skills, competencies, qualifications and occupations tool developed by the EC), Cedefop's skills panorama and EURES (EU's job portal), as well as European Qualifications Framework (EQF) are few examples of such efforts.

Researchers and policymakers rely on interviews, trade publications, surveys, and vacancies to get more insight into labour demands or supply. While these traditional data sources have some clear advantages, they are also characterised by limitations that can be addressed by using web-based data instead. The web is a gold mine for job knowledge discovery. Linked open data, job announcements, social media, job search engines, forums, wikis, data streams and interlinked information are few examples of such valuable job-related sources on the net. Together with (Hooley et al., 2012), we distinguish online data sources covering the Internet and those using the Internet to conduct research. Out of these, we focus on data covering the Internet.

Labour market research exploring online data already has a history of counted in decades. Its first wave explored online search trends in predicting (or now-casting) labour market development (Askitas & Zimmermann, 2009; Caperna et al., 2020; Fondeur & Karamé, 2013; Schmidt & Vosen, 2013; Tuhkuri, 2016). In the later wave, the interest was re-directed towards social networks data, such as Twitter (2014; Barberá & Rivero, 2015; Blank, 2017; Rafail, 2018). Lenaerts et al. (2016) provide an overview of the studies exploring non-vacancy online data for labour market analysis.

Social research specifically focusing on online job vacancy (OJV) data divides into studies processing time-series and trends on aggregate figures (Antenucci et al., 2014; De Pedraza et al., 2019; Lovaglio et al., 2020) and those relying on a more detailed analysis of the OJVs content (Deming & Kahn, 2018; Fabo et al., 2017; Kureková et al., 2015; Turrell et al., 2018). Cedefop (2019) maps the OJV data sources across the EU and explores the potential of this data in generating information on the demanded skill needs and complementing the official labour market statistics. Reusens et al. (2018) evaluate a public employment service job seekers to jobs recommender system supported by OJV data for the Netherlands.

OJV data attract a growing amount of interest in the labour market analysis. In this context, the DISKOW project aims to explore its potential in providing labour-market information closer to the end-user, the job seeker or the case-worker aiding special groups of disadvantaged job seekers (e.g. the immigrants).

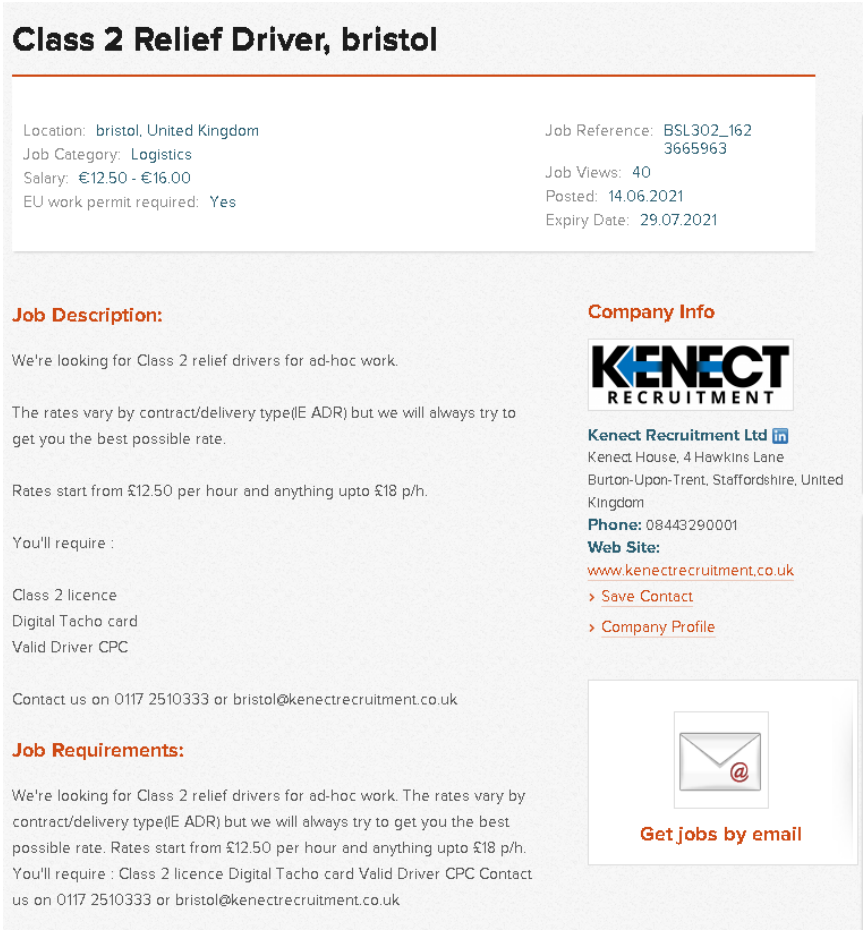
[1] https://ec.europa.eu/commission/presscorner/detail/en/IP_10_1673

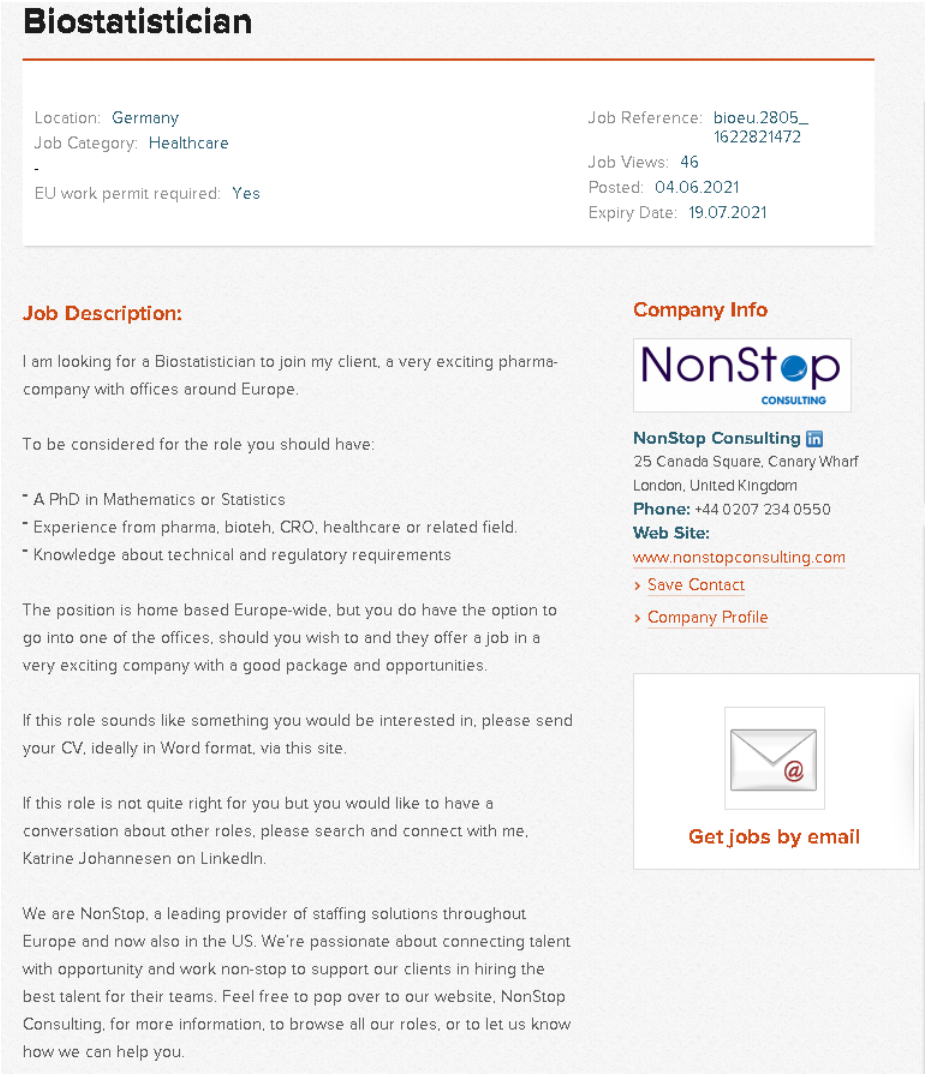
2. The Job Knowledge Base

Since the OJV data were not collected for the purpose of labour market analysis, they need further structuring in order to become employable in labour market analysis. The main problem in this regard is not in the availability of data and how to retrieve them, but in how to cleanse, explore, visualise and interpret such a huge volume of various web data: in order to streamline this process and make such data suitable for further exploitation (e.g. consumed by specialised mobile apps) an open Job Knowledge Base (JKB) is proposed that can be used by employers, employees, job seekers, labor market experts and policymakers.

The JKB contains information from different OJVs, allowing for a different structure of the OJV text. It extracts particular types of information, such as salary information, geographical and time information, together with the requested skills and other requirements. For the reason of correct identification of relevant pieces of information out of natural language text, an appropriate ontology was selected (see the Intelectual Output 1 of the DISKOW project). Thanks to a universally applicable ontology, multiple OJV data sources can be linked to the JKB. The overview of the potentially employable data sources across the EU countries can be found in the Intelectual Output 2 of the DISKOW project. In the JKB prototype version, one OJV data source was scraped to feed the JKB interface (eurojobs.com). Scraped data source contained OJVs in various structure and languages.

Examples are provided below:





Biostatistician

Location: Germany
Job Category: Healthcare
-
EU work permit required: Yes

Job Reference: bioeu.2805_1622821472
Job Views: 46
Posted: 04.06.2021
Expiry Date: 19.07.2021

Job Description:

I am looking for a Biostatistician to join my client, a very exciting pharma-company with offices around Europe.

To be considered for the role you should have:

- A PhD in Mathematics or Statistics
- Experience from pharma, bioteh, CRO, healthcare or related field.
- Knowledge about technical and regulatory requirements

The position is home based Europe-wide, but you do have the option to go into one of the offices, should you wish to and they offer a job in a very exciting company with a good package and opportunities.

If this role sounds like something you would be interested in, please send your CV, ideally in Word format, via this site.

If this role is not quite right for you but you would like to have a conversation about other roles, please search and connect with me, Katrine Johannesen on LinkedIn.

We are NonStop, a leading provider of staffing solutions throughout Europe and now also in the US. We're passionate about connecting talent with opportunity and work non-stop to support our clients in hiring the best talent for their teams. Feel free to pop over to our website, NonStop Consulting, for more information, to browse all our roles, or to let us know how we can help you.

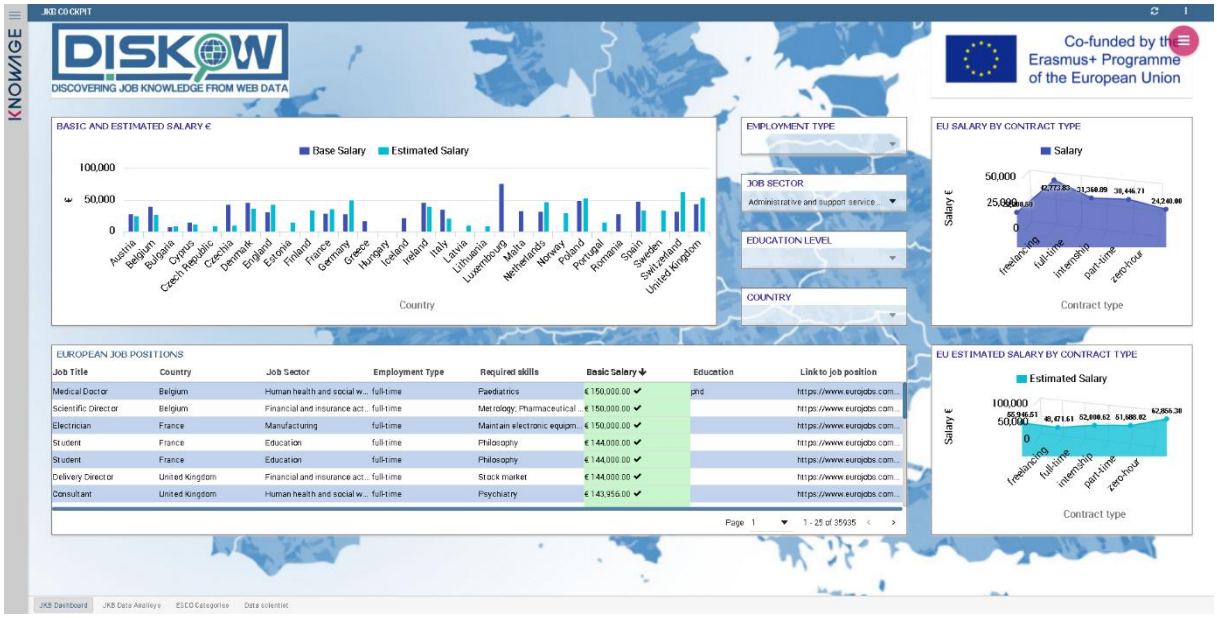
Company Info

NonStop CONSULTING

NonStop Consulting
25 Canada Square, Canary Wharf
London, United Kingdom
Phone: +44 0207 234 0550
Web Site:
www.nonstopconsulting.com
› Save Contact
› Company Profile

Get jobs by email

The graphical user interface was designed using an open-source platform KnowAge. The final prototype of the user interface, used for the stakeholder feedbacks is shown below.



The project DISKOW aims at creating a JKB prototype based on an existing open-source Business Intelligence platform in order to cover the most important factors in this regard, such as required job knowledge for a specific job. In line with the project proposal, the JKB prototype was build using open-source tools and is provided to the community as an open-source tool for further development.

JKB platform consists of four user sheets, similar to the Excel format. Once logged in, users are brought to the JKB sheet called JKB Dashboard. Dashboard is the core of the project and its aim is to provide the user with the structured database of the jobs available within the EU. Navigation through job postings is done by four filters, where user can filter out desired vacancies based on the Employment type (full-time, internship, part-time), Job sector based on NACE, Education level (bachelor, masters, diploma, phd) and the country of the vacancy. On the sides of the filter are graphs depicting the average salaries based on filtered choices. Base salary is calculated from the salaries published directly in the scrapped data, while estimated salary is predicted based on the values reported in the European Social Survey. Bottom left of the sheet obtaines the table filtering available vacancies and structured information on them. Once mouse is centered on the table, one can find additional filter in the upper right corner, that can be used to further narrow the user choice. Once desired jobs are filtered and scammed through, user can proceed to the source posting for the further information by copying the link provided in the final column.

To pick the other sheets of the JKB, user has to use the buttons on the very bottom of the platform. JKB Data Analysis provides the user with the information on the amount of the jobs based on the country, education and sector specification. To navigate within the data filters are available on the left side of the sheet. Overwiev is provided below:

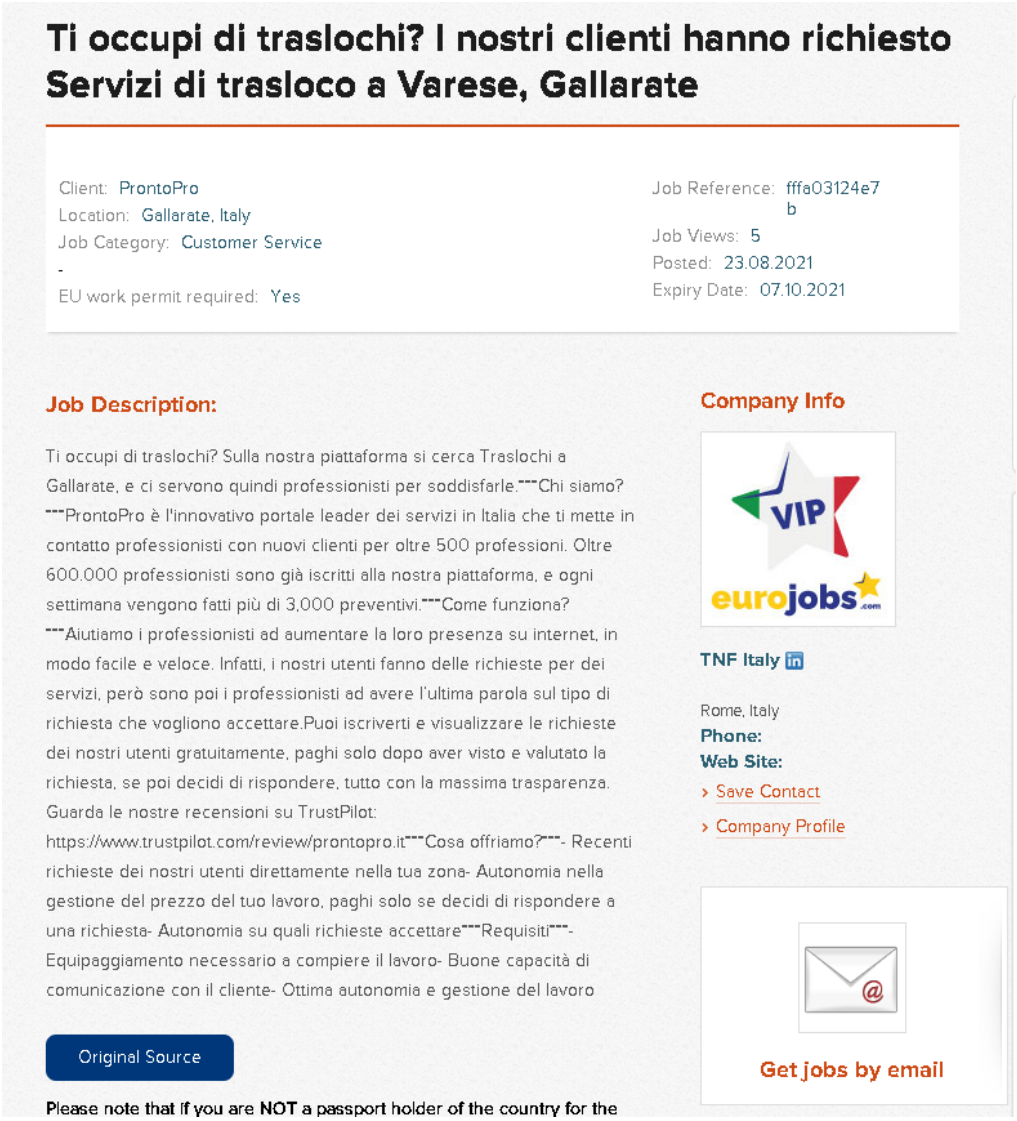
Third sheet, called ESCO categories provides the dictionairy of the skills assigned to specific occupations by European Skills, Competences, Qualifications and Occupations database. Last sheet of JKB, Data scientist, should be focused on the specific vacancies relevant for the Data scientists, however is currently not working.

Two moments are crucial in evaluating the JKB prototype. First, it is the accuracy of data extraction using the developed algorithm. Second, it is the evaluation of JKB interface from the perspective of the potential end-user. These two moments are going to be addressed in the following text.

3. Evaluation of the accuracy of the extracted data

In the first step, information needs to be extracted from OJV postings written in natural languages. An automated algorithm was developed for this purpose. This is published https://git.l3s.uni-hannover.de/diskow/diskow

To assess the accuracy of the data extraction, we have picked 100 random job postings and compared the information displayed by JKB platform after the data processing with the information from the original job posting published on eurojobs.com. Among these 100 postings are texts written in English, French, German, Italian and Dutch. Regarding the postings in Italian, the structure of the posting did not allow us to successfully evaluate the information gathered. These postings contain one sentence regarding the job vacancy and further several paragraphs on the providing external company, not mentioning any description of the job. Example is provided below:



Ti occupi di traslochi? I nostri clienti hanno richiesto Servizi di trasloco a Varese, Gallarate

Client: ProntoPro
Location: Gallarate, Italy
Job Category: Customer Service
-
EU work permit required: Yes

Job Reference: fffa03124e7b
Job Views: 5
Posted: 23.08.2021
Expiry Date: 07.10.2021

Job Description:

Ti occupi di traslochi? Sulla nostra piattaforma si cerca Traslochi a Gallarate, e ci servono quindi professionisti per soddisfarle.""Chi siamo? ""ProntoPro è l'innovativo portale leader dei servizi in Italia che ti mette in contatto professionisti con nuovi clienti per oltre 500 professioni. Oltre 600.000 professionisti sono già iscritti alla nostra piattaforma, e ogni settimana vengono fatti più di 3.000 preventivi.""Come funziona? ""Aiutiamo i professionisti ad aumentare la loro presenza su internet, in modo facile e veloce. Infatti, i nostri utenti fanno delle richieste per dei servizi, però sono poi i professionisti ad avere l'ultima parola sul tipo di richiesta che vogliono accettare.Puoi iscriverti e visualizzare le richieste dei nostri utenti gratuitamente, paghi solo dopo aver visto e valutato la richiesta, se poi decidi di rispondere, tutto con la massima trasparenza. Guarda le nostre recensioni su TrustPilot: https://www.trustpilot.com/review/prontopro.it""Cosa offriamo?""- Recenti richieste dei nostri utenti direttamente nella tua zona- Autonomia nella gestione del prezzo del tuo lavoro, paghi solo se decidi di rispondere a una richiesta- Autonomia su quali richieste accettare""Requisiti""- Equipaggiamento necessario a compiere il lavoro- Buone capacità di comunicazione con il cliente- Ottima autonomia e gestione del lavoro

Original Source

Please note that If you are NOT a passport holder of the country for the

Company Info

VIP eurojobs.com

TNF Italy

Rome, Italy
Phone:
Web Site:
› Save Contact
› Company Profile

Get jobs by email

The focus of the evaluation was set on the variables Salary, Currency, Education requirements, Employment type, Job title, Job location, Industry and Skills. We have also tried to assess the

extraction of the working hours; however, the information is mentioned in the job posting so rarely (less than 10%) that we dropped from this intention and disregarded the beforementioned variable.

From the randomly picked observations, approximately 25% possessed the information on the salary. If the amount published was set to be salary per year or month, the JKB code extracted the data and displayed salary, calculated per year, correctly in 73% of the cases. Part of the inaccuracy is also caused by the row data structure because a small portion of the job postings is not giving information on a single job but rather a set of positions within the company, mentioning several salaries at once. In cases of the vacancies situated in England, the processing code converted it with the exchange rate GBP/euro from the date of the sample construction accurately. If the salary was published as a range, the code took the upper bound. On the contrary, in cases of the salaries set on an hourly basis, the code could not extract the data and fill in the estimated yearly salary. For these salaries, the accuracy is 0%, and improvement would be desirable.

Trying to overcome the issue of unavailability of the salary data, we created a table of the estimated salaries given the three-digit ISCO code for a given country based on the data from the European Social Survey (https://www.europeansocialsurvey.org/). However, when checked for the postings with the salary given by the employer, we conclude that it is highly inaccurate, sometimes giving more than two-fold amounts. This is possibly due to the low number of observations in the abovementioned survey. Because of the abovementioned, the estimated salary should be interpreted with caution.

Due to the source used, the possible currencies were the euro and British pound. The accuracy of this category is 70%. The 30% inaccuracy accounts for the postings where the code was not able to extract salary because it was given in hourly values. In these cases, JKB shows incorrectly that the data on salary, and thus currency, is not provided. However, if the currency is being identified by the JKB, it is always the correct one.

The requirements on the education are, in some way, given in 35% of the job postings. The accuracy of our code for its extraction is 30% in cases of the postings in which information is given. If we assume that the postings without the information of the educational requirement are the ones where no specific level is required, the accuracy goes up to 70%. Relatively low accuracy may be explained by the differences in educational systems between countries, where a considerable number of possible qualifications exists, and usually the requirement is mentioned as a specific certificate or diploma, and one needs to go through the educational system of a given country to find what level of education is needed obtain that certificate. There is also a particular issue with the string categorising. In many posts, it is explicitly mentioned that some *skill should be mastered*, which is mistakenly sorted to the group University degree by requiring jobs. To overcome this issue, we are implementing the code which, in addition to string, would also take the three-digit ISCO categorisation into account. This should improve the accuracy significantly.

Regarding the employment type, the accuracy of the code is 89% (where both, part/full time are suitable, 50% accuracy is assigned). Even the remaining 11% inaccuracy is not due to the mistake, but rather because in some postings, it is explicitly mentioned that the position is suitable for full-time as well as part-time, but our categorising process considers just the first option.

The accuracy of the extracted job title is 51%. Here we would like to point out that approximately in one out of ten postings, it is not clear what is the position, or it is not a position at all. Firstly, as mentioned above, some postings consist of several job positions plugged into one text and secondly, several posts are looking for someone to move the kitchen or for disposal of some furniture. In these cases, it is difficult to assess the job title even manually. Once these posts are excluded from the

evaluation, the accuracy rises to 65%. Some of the remaining inaccuracy can be attributed to the simplification of the occupations. For example, positions such as Real estate agent or Breeding agent are both classified simply as Agent. Similarly, disregarding the vehicle, Telehandler driver and Courier driver are both classified simply as drivers. Although the simplification leads to the loss of accuracy, it helps to keep displayed data sensible and user friendly, and the trade-off is desirable.

Extracting the information on job location at the country level, our code is 96% accurate. We had 9 countries in our sample – Germany, UK, Italy, France, the Netherlands, Belgium, Malta, Cyprus, and Switzerland. There seems to be an issue with the Netherlands as out of 5 jobs situated there just 1 was correctly associated with that country. For the rest of the countries, the JKB code was 100% accurate.

Moving to the extraction of information about the industry (economic branch), where the NACE classification was used, JKB shows 31% accuracy. The industry classification is based on the skills identified in the job postings, which are, based on the ESCO dictionary translated to the NACE category. There are three reasons for the low accuracy. The first issue is that several skills might be identified for a given posting that can lead to various NACE categories. This seems to be the aspect with the greatest room to be improved. The second problem arises from the structure of the data. There are postings that are two lines long, and no skill extraction is possible at all. As an example, in posts where someone is looking for an unspecified helping hand with the renovation or kitchen installation, the string analysis is not able to identify any skill and thus, the industry is not classified at all. These account for approximately 10% of the posts. The third issue comes from the translation. We have postings in several languages, and we are currently using a free translator, which is less accurate than professional but expensive translators such as the one from Google. Therefore, some inaccuracy arises from the wrongly translated texts. For potential use in the future, it could be viable to pay for the professional translation of the posts.

The last category assessed was the accuracy of extracted skills. The skill extraction is based on the keyword analysis of the strings. JKB reports the precision of the extracted skills at the level of almost 69.6%. We have disregarded the observation where the structure of the text did not allow the extraction of the skills needed. As already mentioned, these are the posts where someone looks for the unspecified helping hand. Part of the inaccuracy is again arising from the translation. However, during the evaluation, we have identified specific skills which are usually inaccurately associated with the jobs. These are Energy, Scaffolding components and Keep company skills. As these seemed not to be relevant for most of the jobs that the code associated them with, we have decided to turn these skills off. Once done, accuracy rises to more than 71%. We would like to point out, that the inaccuracy is caused by the false positive cases only, taking into account the falsely assigned skills not related to the job. As the number of potential skills is rather extensive we did not evaluated false negatives, because that would be highly subjective, generated potential additional skills for most of the postings.

To conclude, the JKB shows relatively high accuracy in variables of salary (where given), currency, employment type, locations, skills and partly also job title. On the other hand, room for improvement was identified in categories of Education and Industry. Overall, the level of accuracy of the information extracted by the JKB platform prototype is 70 %.

4. Evaluation of the user experience

As a part of the JKB evaluation, we asked external stakeholders to provide us with user experience feedback during the online meetings. These stakeholders consisted of six non-governmental organisations working with the job seekers as a part of their mission. They were included in the project evaluation and asked for their feedback because of their interest in our project during the Multiplier events, where working versions of the JKB platform were presented. Due to the scarcity of time on their side, we have conducted four questionnaires (in the first wave) and three in-depth online interviews (in the second wave). Prior the in-depth interviews, stakeholders were asked to test prototype of the JKB platform and during online meetings they have provided feedback regarding their experience by answering the questions from the prepaired questionnaire (see Appendix). *Človek v ohrození, Fondazione Le Vele* and *Career & Life Planning,* were involved in the testing exercise.

4.1 Description of the stakeholders

Človek v ohrození is a Slovak non-profit, non-governmental organisation providing help to people deprived of dignity and freedom for various reasons. Regarding job seekers, they are trying to decrease long term unemployment within Slovakia by helping jobseekers from Roma communities to find suitable positions. Based on the filled-in questionnaire, depending on the month, the organisation has up to 7 counsellors, and approximately 20% of their clients find a suitable job within Slovakia successfully each month. Their usual client can be characterised as low-skilled and long-term unemployed with unfinished secondary education.

The second stakeholder, *Fondazione Le Velle*, is a non-profit organisation situated in Pavia that provides training courses, counselling, and assistance to the disadvantaged groups of job seekers in Italy. On an average month, they provide counselling to approximately 60 clients, and between 20 to 40% of these do find a suitable job in Italy successfully each month. Jobseekers could be characterised as low-skilled, youth or long-term unemployed, out of which some are disabled. They do not possess a University degree.

Lastly, the feedback interview was held with Mr. Joe Delaney, CEO of *Career & Life Planning*. is a professional training and coaching company with offices in the USA and Ireland. The company performs various activities to support employability, one of which is helping the job seekers to find suitable jobs, mainly in Ireland. Mr. Delaney provides employment counselling to approximately the 20 clients per month, however, their number, success rate and general characteristics differ significantly over time and group. They have several groups of clients that can be characterised by all educational levels depending on the group.

4.2 Summary of the main findings from the interviews

The stakeholders reported that they have spent up to 30 minutes assessing the prototype of the JKB platform. All of them used Google Chrome as a browser, and they did not encounter any technical bugs making the JKB unusable. The only issue reported was that respondents from *Človek v ohrození* experienced long loading times once filters were applied; however, none of the other users, nor we as a developer, encountered this issue. A possible reason might be using the less-performing laptop by the respondent or the internet connection. However, although with some time lag, the platform

worked as it should. To conclude, none of the testing users encountered login issues or disconnection; thus, the platform is deemed stable.

Regarding the JKB impression, stakeholders rated the platform prototype positively, with the exception of Mr. Delaney, who suggested that there is room to improve. He considered JKB in its current form to be a bit unclear, and he would appreciate some enclosed manual or even a video following an avatar explaining the use of the JKB prototype and its functionalities. The other two stakeholders did not have any comments, positive nor negative, on the intuitiveness of the platform use.

The most significant potential for improvement was identified within the filtering functionality. Firstly, *Fondazione Le Vele* stressed out that the filter for educational requirements was not working correctly. Thanks to the bug, it was possible, by the specific combination of the choices, to break the code and report a message "NO AVAILABLE DATA", even though the data were there. The developers solved the bug, and once the code was rewritten, none of the other two users, delivering feedback later, encountered this issue. Filters worked as they were designed to work. However, later respondents reported that they would appreciate enhanced filtering opportunities, to access the vacancies for the less educated/disabled or otherwise disadvantaged people. They could not benefit from the JKB fully in its current state, as it was not focused on their targeted population. Although this point arises mainly from the choice of the stakeholders and is not entirely relevant in connection with our goal to be focused on the Data scientists, in the future, it could improve the potential and user value if the filters are to be expanded. Lastly, a minor suggestion was made to link the filters between Dashboard and Data analysis tool. However, a relevant point was picked up only by one respondent and seemed not to influence the user experience of the other interviewees.

Assessing the graphs, users on average considered visualisation to be neither good nor bad. Graphs depicted information based on the filtered selection correctly, and users were able to interpret the information shown without problems.

Regarding the content, we need to mention that all the stakeholders were negatively surprised by the fact that JKB is currently scraping, processing, and depicting a single webpage - eurojobs.com. This results from the copyrights, as one needs special permissions to assess the data published on various websites with own scraping code. However, once the copyrights are bought, the code would be able to process any number of web pages desired. Thus, we consider this issue to be more of a legal nature than a technical complication.

The variables processed and displayed were considered by the test users as satisfactory. They did not identify any of them to be redundant. However, as mentioned above, stakeholders suggested adding filters for the jobs requiring lower levels of education. *Fondazione Le Vele* also recommended adding a filter for jobs suitable for disabled people. However, this is not quickly done, as the suitability for the disabled is usually not stressed in the job posting and it is not clear from neither job position nor industry filtered. There is not much that could be done related to this specific suggestion with the current structure of the published job postings.

Further, stakeholders suggested making links interactive, as in their current form, it needs to be manually copied and plugged into the browser. Although not essential, it could improve the user experience.

Lastly, respondents from *Človek v ohrození* and *Career & Life Planning* emphasised that the current form of the JKB, which reminds the excel sheet system, is not the most user friendly. They wrongly assumed that the switch between the modules of the JKB could be found under the menu either in

the top left or top right corner. As it was not there, they thought that they are not authorised to access anything. Although, in the end, they managed to find choices at the bottom of the platform, they strongly recommend either somehow highlight the selection or moving it to the top.

Resulting from the previous answers, all three stakeholders considered the JKB platform not beneficiary to them in its current form but identified its potential once there is a bigger focus on the disadvantaged categories of jobseekers. In general, they recognised and appreciated the potential of the platform.

5. Conclusions

During the assessment of the prototype of the JKB platform, the accuracy and user experience were evaluated. The overall accuracy of the extracted data was apx. 70%. The data processing is the most accurate for the categories of *job location*, *employment type*, *currency* and *skills required for the job*. On the other hand, potential space for improvement was identified in categories of *educational requirements* and *industry*. The first mentioned could be potentially improved by following the ISCO guidelines, assigning the job positions based on ISCO codes to the educational degrees. Pairing the vacancies to the ISCO codes is already done for the category estimation of the salary. This would decrease the number of missing values, where the degree is not mentioned in the posting. Also, some postings are looking for interns or apprentices, which would be incorrectly paired to the ISCO of the thoroughly educated professional in the given field. On the other hand, filtering vacancies based on the comparison of the text strings to our dictionaries are considered to render more precise results for the postings explicitly mentioning the required education.

Although a considerable part of the inaccuracy arises from the structure and content of the job vacancies posted on the https://eurojobs.com/ website, the accuracy might be further enhanced by the future users by the implementation of more robust dictionaries, allowing the more precise translation from other official European languages to English. As the final code is available online on the GitHub link: https://git.l3s.uni-hannover.de/diskow/diskow, it gives the user endless possibilities to insert any dictionary they feel might be desired.

Following the questionnaires, the user experience can be considered to be primarily positive, while several suggestions for further improvement were brought up by the stakeholders. Firstly, we can conclude that the JKB prototype was stable, and none of the respondents reported platform crashing bugs. During the first interview, the minor issue with the filter was identified, giving no results for certain filters, but this was promptly fixed, and in the later interviews, filters and graphs were assessed as working without further technical issues. However, with respect to the filters, respondents jointly reported that more categories representing less educated people should be added to the options. This is a result of the stakeholders' orientation and can be implemented in the future by the broadening of the dictionaries.

Regarding the visualisation and UI several suggestions, respondents reported that they managed to get familiar with and use the platform after the initial tryout. However, they stated that the menu for several sheets provided on the bottom of the platform is not easily recognisable and could be either highlighted, enlarged or moved to the top. One of the stakeholders did not manage to find the buttons at all before being instructed.

In overall, respondents identified the potential of the platform, however in its current state JKB is not well suited for their client structure. Once the filters are broadened and the job postings sources more extensive, they might reconsider.

To conclude, the project resulted in the initialisation of the JKB platform prototype, with the source code being published on GitHub, allowing those interested in modifying it based on their needs and desires. In its current state, it shows a satisfactory level of data processing accuracy, stability of the platform and provides the tools necessary for the data analysis and the search of the relevant job positions. We believe that the prototype of the JKB delivers a basepoint for further use in both

academic as well as practical use. It allows to scam through the structured summary of the jobs within EU disregarding the original language of the posting and thus helps in the work of the HR specialists and career consultants.

References:

Antenucci, D., Cafarella, M., Levenstein, M., Ré, C., & Shapiro, M. (2014). Using Social Media to Measure Labor Market Flows. *National Bureau of Economic Research*. https://doi.org/10.3386/w20010

Askitas, N., & Zimmermann, K. F. (2009). Google Econometrics and Unemployment Forecasting. *Applied Economics Quarterly*, *55*(2), 107–120. https://doi.org/10.3790/aeq.55.2.107

Barberá, P., & Rivero, G. (2015). Understanding the Political Representativeness of Twitter Users. *Social Science Computer Review*, *33*(6), 712–729. https://doi.org/10.1177/0894439314558836

Blank, G. (2017). The Digital Divide Among Twitter Users and Its Implications for Social Research. *Social Science Computer Review*, *35*(6), 679–697. https://doi.org/10.1177/0894439316671698

Caperna, G., Colagrossi, M., Geraci, A., & Mazzarella, G. (2020). Googling Unemployment During the Pandemic: Inference and Nowcast Using Search Data. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3627754

Cedefop. (2019). *The online job vacancy market in the EU Driving forces and emerging trends*. https://doi.org/10.2801/16675

De Pedraza, P., Visintin, S., Tijdens, K., & Kismihók, G. (2019). Survey vs Scraped Data: Comparing Time Series Properties of Web and Survey Vacancy Data. *IZA Journal of Labor Economics*, *8*(1), 1–23. https://doi.org/10.2478/izajole-2019-0004

Deming, D., & Kahn, L. B. (2018). Skill Requirements across Firms and Labor Markets: Evidence from Job Postings for Professionals. *Journal of Labor Economics*, *36*(S1), S337–S369. https://doi.org/10.1086/694106

Fabo, B., Beblavý, M., & Lenaerts, K. (2017). The importance of foreign language skills in the labour markets of Central and Eastern Europe: assessment based on data from online job portals. *Empirica*, *44*(3), 487–508. https://doi.org/10.1007/s10663-017-9374-6

Fondeur, Y., & Karamé, F. (2013). Can Google data help predict French youth unemployment? *Economic Modelling*, *30*(1), 117–125. https://doi.org/10.1016/j.econmod.2012.07.017

Hooley, T., Marriott, J., & Wellens, J. (2012). What is Online Research? : Using the Internet for Social Science Research. In *What is Online Research? : Using the Internet for Social Science Research*. Bloomsbury Academic. https://doi.org/10.5040/9781849665544

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*(1). https://doi.org/10.1186/s40172-015-0034-4

Lenaerts, K., Beblavý, M., & Fabo, B. (2016). Prospects for utilisation of non-vacancy Internet data in labour market analysis—an overview. *IZA Journal of Labor Economics*, *5*(1), 1–18. https://doi.org/10.1186/s40172-016-0042-z

Lovaglio, P. G., Mezzanzanica, M., & Colombo, E. (2020). Comparing time series characteristics of official and web job vacancy data. *Quality and Quantity*, *54*(1), 85–98. https://doi.org/10.1007/s11135-019-00940-3

Rafail, P. (2018). Nonprobability Sampling and Twitter. *Social Science Computer Review*, *36*(2), 195–211. https://doi.org/10.1177/0894439317709431

Reusens, M., Lemahieu, W., Baesens, B., & Sels, L. (2018). Evaluating recommendation and search in the labor market. *Knowledge-Based Systems*, *152*, 62–69. https://doi.org/10.1016/j.knosys.2018.04.007

Schmidt, T., & Vosen, S. (2013). Using Internet Data to Account for Special Events in Economic Forecasting. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2200402

Tuhkuri, J. (2016). *ETLAnow: A Model for Forecasting with Big Data*. *54*. https://doi.org/10.4995/carma2016.2016.4224

Turrell, A., Speigner, B., Djumalieva, J., Copple, D., & Thurgood, J. (2018). *Using job vacancies to understand the effects of labour market mismatch on UK output and productivity*. www.bankofengland.co.uk/working-paper/staff-working-papers

Appendix: JKB GUI evaluation questionnaire

Introduction:

The aim of this questionnaire is to assess and evaluate the user interface and visualisation of a prototype of a platform developed to assists job seekers in finding a suitable job in one of the EU countries. Online job advertisements from several online job posting sources are aggregated into a single Job Knowledge Base (JKB), which feeds the final version of the prototype of our platform. The platform is accessible online via PC or laptop. The main goal is to provide an intuitive, fast, and cheap source of information on job opportunities and skill requirements, as specified in the ISCO coding. The platform should reduce the time needed to find a job in the EU and has the potential to optimize the matching of clients' skills to the actual demand for labour.

We would like to kindly ask you to fill in the following questionnaire and thus to evaluate the final product of our efforts. Your answers will be used to assess the overall functionality, intuitiveness, and visualisation of the prototype of a platform, its content, and features. Being a person from the field of matching job seekers with job positions, your answers will provide valuable feedback.

Section 1: User case

1. **Do you, in your work, provide employment counselling (assistance in looking for a job)?**

☐ Yes
☐ No

(If No: do not answer any more questions - end of the questionnaire)

2. **You would describe yourself as** (Please, choose one or multiple options, if applicable)**:**

☐ Employment counsellor
☐ Career counsellor
☐ Social worker
☐ Case-worker providing complex assistance to vulnerable groups (e.g. immigrants)
☐ HR specialist
☐ Other, please specify: ……

3. **On an average month, to how many clients do you provide employment counselling?** (Please, provide approximate number)**:**
Number: ……

4. **On an average month, what share of your job-seeking clients finds job successfully?** (Please, choose one or multiple options, if applicable)**:**

☐ 0-20%
☐ 20-40%
☐ 40-60%
☐ 60-80%
☐ 80-100%

5. **Do you, in your work, use the job-posting websites as a source of the job offers?**

☐ Yes
☐ No
If No, please specify what other sources do you use (newspapers/ government websites/ direct offers from partner organisations, etc..): …….

6. **Are you, or your clients, interested only in the vacancies situated in Italy, disregarding other EU countries?** (Please, choose the most suitable option)**:**

☐ Yes
☐ No

7. **How would you describe your clients?** (Please, choose one or multiple options, if applicable)

☐ Low-skilled
☐ Asylum seekers
☐ Youth unemployed
☐ Elderly
☐ Disabled
☐ Long-term unemployed
☐ Other, please specify: ……

8. **Your typical client can be characterised by** (Please, choose the most suitable option):

☐ Unfinished secondary education
☐ Secondary education
☐ Bachelor's degree
☐ Master's degree or higher
☐ Do not know

9. **While assessing the job vacancies, how would you rank following variables based on their importance?** (Rank form 1 to 8, 1 being the most important)

Salary (monthly, weekly or hourly wage brutto or netto)
Required education (secondary, bachelor, master, phd…)
Employment type (full time, part time, sessional work, internship, self-employed)
Industry/Job sector (tourism, ICT, health service, agriculture, manufacturing…)
Country/Location (Germany, Slovakia, Italy… or more precise)
Occupation category (cashier, manager, technician, driver, nurse, accountant…)
Skills/prerequisites (PC skills, MS office, languages, certificates, experience with…)
Hiring organization (Amazon, Dell, Google, IGN,…)

Section 2: Use of the JKB

10. How much time, approximately, did you spend browsing the JKB platform?

☐ Less than 10 minutes
☐ 10 – 20 minutes
☐ 20 – 30 minutes
☐ More than 30 minutes

11. Which browser did you use to access the platform?

☐ Google Chrome
☐ Internet Explorer
☐ Opera
☐ Safari
☐ Other: ……

12. Using the JKB platform, did you encounter any technical issues? (Button not working/ / lags / disconnecting / etc.)**:**

☐ No
☐ Yes (Please specify): …….

13. How would you describe your general impression of the platform? (Please, choose the most suitable option)**:**

☐ Very good
☐ Good
☐ Normal
☐ Bad
☐ Very bad

14. How would you rate your experience of the JKB regarding its intuitiveness? (Please, choose the most suitable option)**:**

☐ Intuitive
☐ Somewhat intuitive
☐ Neither intuitive nor counterintuitive
☐ Somewhat counterintuitive
☐ Counterintuitive

15. How would you rate your experience of the JKB prototype regarding its visualisation? (Please, choose the most suitable option)**:**

☐ Very good
☐ Good
☐ Normal
☐ Bad
☐ Very bad

16. **Would you, on average, consider your clients to be able to independently use the JKB platform to find the job?** (Please, choose the most suitable option)**:**

☐ Yes
☐ Yes, to a limited extent
☐ Probably not
☐ Not
If not, please specify the reason (Internet access/ Lack of technical skills/ Insufficient language skills): …….

17. **Do you think JKB platform could help to boost the share of your job-seeking clients getting employed?** (Please, choose the most suitable option)**:**

☐ Yes
☐ No

18. **Supposing we do not have any legal restrictions, are there any specific websites which you would like to get scraped for the JKB platform?:**
Please specify: ……

19. **In overall, would you consider using the JKB platform in your further work?** (Please, choose the most suitable option)**:**

☐ Yes
☐ No
(If No: skip questions 21 and 22 and finish the questionnaire)

20. **Would you consider using our platform after adding certain features?** (Please, choose the most suitable option)**:**

☐ No
☐ Yes
(If No: skip question 22 and finish the questionnaire)

21. **Could you provide short feedback on what features (information) essential for your work are missing within the JKB platform?** (up to 100 words):

☐ Answer: ……

Thank you for your answers!