European Network on Regional Labour Market Monitoring

NOTE

Seminars of the Big Data Knowledge Hub

Analysis of online job vacancies and occupations and skills in demand in North Macedonia

Notes of the online seminar promoted by the Big Data Knowledge Hub of the European Network on Regional Labour Market Monitoring (ENRLMM). June 29, 2023

The fifth of the Seminars of the Big Data Knowledge Hub took place on June 29, 2023. The aim of these series of seminars is to offer an opportunity to deepen the Network's knowledge on how to use Big Data for labour market research and consulting by presenting practical cases and demonstrations.

The Seminar included a presentation by **Aleksandar Kostadinov** (Researcher at IDEI-Skopje) who presented the project "Analysis of online job vacancies and occupations and skills in demand in North Macedonia".

The open discussion counted with the participation of: **Christa Larsen** (IWAK Goethe University Frankfurt am Main, Germany), **Aleksandra Webb** (University of the West of Scotland, UK) and **Christian Müller** (SECO, Switzerland).

Introduction

Eugenia Atin (Speaker of the Big Data Working Group of the ENRLMM) after the initial greetings and thanks to the participants, contextualised the session in the work being done by the Big Data Working Group of the ENRLMM (European Network on Regional Labour Market Monitoring).

Christa Larsen (Founder and Coordinator of the ENRLMM), then welcomed the participants in the name of the European Network on Regional Labour Market Monitoring (ENRLMM) http://regionallabourmarketmonitoring.net/, a nonprofit network with 400 members across Europe and beyond. The Network shares a common interest and performs observation on regional labour markets. We are always searching to bring expert knowledge for evidence-based policy making. Many of us do not have the resources or the knowledge to explore the demand side, so these Seminars of the Big Data bring the opportunity to Network members to work with data from the internet.

Presentation by Aleksandar Kostadinov

Aleksandar Kostadinov is a Skopje based Expert researcher who has worked as an economic analyst and researcher on various national and regional projects funded by the European Commission, ILO, IOM, OECD, USAID, EBRD, the British Council-Macedonia as well as local entities.

The project's inspiration was the Cedefop Skills Ovate back in 2019, which only included the EU countries but no western Balkan countries.

The idea was to use alternative and open-source information, in this case the job vacancies that are published on a web portal. For this purpose he created a code in Python to web scrape one job portal: Najdirabota.com.mk, one of the oldest and most popular portals. The term "big data" is relative because North Macedonia is a small country so we talk about thousands of vacancies, not millions like in other countries.

As a first step, he analysed the structure of the website and how the objects were inserted, the position of the vacancy, the qualifications, the short description, the announce date, the apply Deadline, the geographical areas, and the company name. He excluded fields such as contact information and personal data.

Aleksandar Kostadinov made a demonstration and opened the tool he used (Jupiter notebook) to show the participants the python code he created. The dates for scrapping are also in the code, so you can change them and then you download the new file. The code only works for this particular website because every website has its own structure, elements, objects… Also note that some websites have some barriers to stop web scrapping, for example, "I'm a human" filters.

When you run the code, the scrapping starts, for the inserted period, and creates an excel file in the desired folder. Aleksandar Kostadinov made the demonstration and immediately downloaded an excel file of 186 registers for the 3 weeks he had selected. The excel file contained all the fields mentioned.

Once we collect the data, he matched them with the ISCO. The positions are translated into the English code. Once we have the ISCO code then we can translate to any language.

Then he compared his results with the results of the statistical office. The results, as expected were different for some occupations because some occupations are more advertised on the internet than others. For example, "professionals", according to Aleksandar Kostadinov's results were nearly 28% of the vacancies, however the statistical office only counted 7.3%.

Aleksandar Kostadinov then showed a graph displaying the occupations' demand over time. The seasons can be appreciated. There are more vacancies in September-October, then decrease, then grow again in springtime, there are less during holidays. The timeline follows the business cycles, so it makes sense.

The dashboard was prepared with Tableau (first he used Power BI and then he switched to Tableau which proofed to be better for this task). It is an interactive dashboard, intuitive, easy to use, explore, it finds patterns… Power BI is a Microsoft product and Tableau has been bought by SalesForce.

You can visualise the data any way you want. Aleksandar Kostadinov showed for example ISCO 3-digit occupations by selecting the service and sales workers group in the interactive Tableau tool.

Aleksandar Kostadinov performed another task for one particular occupation "software and applications developers and analysts". He took the field related to the qualifications and with Orange3 by University of Ljubljana (a machine learning software that allows to create text analysis) identified the words that are more frequently mentioned in the text. He then obtained a word cloud of skills where javascript and English were the most demanded skills in the vacancies.

With the same word cloud (an occupation) he looked into the analytical cable. The next challenge will be to use natural language processing (NLP) and Artificial Intelligence for network analysis.

Aleksandar Kostadinov has used this technology for another purpose back in 2020. He used GPT-3 (previous to ChatGPT- Chat GPT is based on GPT models) in a project commissioned by the Swiss Agency for Development and Cooperation Agency, named "Beneficiary Assessment of Civica Mobilitas" where the aim was to see how the donor founding affected the community. There were interviews and textual data was collected from vast amount of sources and stakeholders. Aleksandar Kostadinov analysed all this text with GPT-3 to connect the topics that people mentioned in the interviews across the country and identify by community which are the things that the programme affected the most. The GPT tool was very helpful to identify the connections and to present it to the public.

Another use of the online job vacancy analysis was during lock-down. Aleksandar Kostadinov ran the analysis for the period January- April 2020 to see how curfew and lockdowns imposed by governmental authorities affected the labour demand. Policy makers did not know what to do during these periods and LFS was postponed by the State Statistical Office, so authorities relied on the administrative data from the PES, but many people could not make the registrations so it was difficult to estimate what

was going on. Through the OJV analysis it was easy and free of charge to see what was happening, what the trends were.

The Ukrainian Central Bank also started to issue similar graphs, which demonstrate that their OJVs were normal before the Russian invasion and then just dropped. UCB also measures the demand for certain occupations.

This indicator could be used by Eurostat or National Banks because it is reliable, although you need the collaboration of the sources. There are also tools that could be used like the memorandum of collaboration in order to secure non restricted flow of data.

Aleksandar Kostadinov performed the analysis without the website's consent although he obtained it later. When doing web scrapping, you burden their servers so portals may block the process, which poses a risk for continual research.

As a conclusion, this task was scholar, but very useful information was extracted so it has a huge potential for the future. The main conclusions of this particular analysis in North Macedonia are:

- Requested skills and occupations can be easily tracked, and the information utilised to quickly adjust policy responses to ensure a better matchmaking process in the labour market;
- The level of demand for technicians and associate professionals is similar and relatively high in both sets of data, while requests for plant and machine operators and assemblers is high only in the survey-based data;
- Issues with the sources related to: Agricultural workers, Public sector jobs;
- Possibilities for Geolocation of the posts, obtaining through API or data sharing;
- Better data registry system could significantly increase the effectiveness and comparability of data.

Aleksandar Kostadinov recommends using some basic rules – for example, introducing ISCO occupations, using ESCO libraries for skills and qualifications, properly entering job location, time and NACE sector – this would produce a more useful and relevant dataset for analysis. Aleksandar Kostadinov used ISCO, and ESCO will probably offer more alignment with Cedefop's Skill Ovate, data sets will be more comparative.

OPEN DISCUSSION

Christa Larsen (IWAK Goethe University Frankfurt am Main, Germany) was impressed by the use of network analysis for visualisation purposes. She wanted to know in what context this might be useful and who could be interested in network

relations? She also liked how trends are shown overtime and the relations with major events. In regions we have lots of transformation purposes, so it could be very interesting to follow trends over time. Could trends be analysed at this regional level for this purpose?

Aleksandar Kostadinov responded that the network analysis he performed was in the context of the "Beneficiary Assessment" project which collected a huge amount of textual data on different topics. It was not related to the labour market research and was textual data obtained from many respondents across the country. For the analysis, Aleksandar used GPT-3 in order to connect the data, topics and insights. GPT-3 as a system provided many insights and connections among the topics spoken by the respondents.

GPT -3 is a powerful system developed on AI language model, trained on huge amount of data available from the entire net. Data for training the model was obtained through various methods, such as web scrapping, libraries, registers, databases etc. Web scrapping process however was also done on social media like Twitter and Facebook, which later imposed limits to access of their content.
Given that the most of the social media platforms are US based companies (such as Twitter, Facebook, Instagram and similar) data ownership is theirs and the access of data to European companies is somehow limited, despite the fact that lot of personal data contained on such platforms comes from European citizens. Such status is advantageous to USA companies that can process and analyse the data, perform profiling and monitor the trendlines. Ideally, this data could be shared with researchers to see the labour market trends and correlated with green jobs for example.

Aleksandra Webb (University of the West of Scotland, UK) thanked the speaker for acknowledging that through the vacancies of the job portals we do not have the complete picture of the labour market because some occupations are not represented. She asked the speaker about his position towards geolocation, in particular with regard to digital nomad working cultures, to what extent the skills profiles are made according to what the companies demand, rather than the region. Because this could have huge implications for policy responses in skills and development.

Aleksandar Kostadinov said that a Serbian thinktank developed a Gig Meter that scrapes freelancer platforms (around 10 countries). They perform analysis but do not include job analysis. They developed a paper for the ETF last year and talk about 2 types of platform workers, those working on global platforms and those working on local platforms.

Christian Müller (SECO) raised the returning question of the quality of data. There is a part of the labour market, e.g. the gastronomy sector, which recruits though word of mouth and is not present in the job portals. Another issue already raised in the last bullet of the conclusions is to have the big portals to agree on some standards for the data. There is competition between the job portals. They want to use unique systems to differentiate from each other instead of using ISCO or some other standard, so what can we do it about it? How to encourage job portals to standardise the way they publish the vacancies? We could try to give them something in exchange, give the job portals analysis or research products. Aleksandar Kostadinov responded that in his case he did not approach the job portal to ask for a different data entry but does not think that they would be willing to change their system. Us, researchers, could compare several job portals and adjust. Another way is to use the PES portal, in Macedonia for example, it is mandatory to register the job offer in the PES, although you can either make it public or private posting. He agrees that it is not sufficient data to have the complete picture, but overtime it improves. With AI is going to improve even more, we will soon be able to identify hard to find jobs, hard to find skills, things are moving very fast.

Aleksandra Webb (University of the West of Scotland, UK) pointed out that there is little collaboration but there is a lot of opportunity. We should try to suggest action points and create spaces for all to come together and collaborate, employers, public, job portal and PES.

Since there are no further comments about Aleksandar Kostadinov's presentation, he now explains some potential lines which could be further researched:

1) The touristic booking portals (e.g. Booking.com, Trivago) hold a lot of data about accommodation in certain areas. They could share their statistics with the region so that the regions can prepare accordingly. The hospitality sector would be aware of the occupancy rate in the different periods, and they could forecast the number of staff needed.

2) LinkedIn can also be a source of skills supplies and movements of professionals. There are tools to web scrape LinkedIn even though they do not want to be scrapped.

3) APIs can be used in many web pages. Many sites do not wish to be web scraped without permission, so they grant you access through API (e.g. New York Times, Guardian). In these newspapers for example you can perform sentiment analysis through machine learning tools using keywords.

4) WikiData is web scraping Wikipedia, you download the data though the WikiData Portal. With a few lines of codes you can search in Wikipedia. You can then extract lots of correlations. It is not a scientific source and it is not always reliable but it can be used for certain purposes.

5) One possible research is the Digital Intensity Occupations (DIO) index. It is not part of the Skills Ovate Portal. The digital skills demand is more intense in some occupations than others and in some regions than others. There is going to be a gap that could be measured with this DIO index. This gap is expected to grow in the future. For the same occupation/ industries, some countries/industries will demand more digital skills than others.

Aleksandra Webb (University of the West of Scotland, UK) is concerned about the privacy issues that are raised by the use of the data for example in the case of booking.com. She is also worried that we could further deteriorate the jobs and the working conditions of workers, meaning that when we know that the job has a high seasonality, then we have more tools and employer excuses to lay off the worker. We should promote innovative practices such as good contracts that find innovative responses when there is low demand.

More information will be circulated as always through the Network's newsletter. Also, if you are interested in sharing your experiences with the Network, please do so through the Knowledge Hub https://bigdatahub.uvt.ro/ and contact Eugenia Atin if you wish to present your project through a Seminar.

Bilbao, July 2023

References

Speaker: Aleksandar Kostadinov
Researcher at IDEI- Skopje
Email: aleksandarkostadinov@yahoo.it
EUROPEAN NETWORK ENRLMM
Email: jenny@jennykipper.de
Email: c.larsen@em.uni-frankfurt.de
Website: www.regionallabourmarketmonitoring.net
BIG DATA WORKING GROUP
Eugenia Atin
E-mail: e.atin@prospektiker.es
Website: https://bigdatahub.uvt.ro/