Human Factors, Business Management and Society, Vol. 135, 2024, 139–147
https://doi.org/10.54941/ahfe1004939

Designing a Proof of Concept for a Virtual Competence Assistant

Markko Liutkevičius, Madis Kõosaar, and Sadok Ben Yahia

Department of Software Science, TALTECH, Estonia

ABSTRACT

This paper addresses the critical gaps within public employment services (PES), with a particular focus on the deficiency in automation and AI-supported intelligent career recommendations. Notably, the advancements in this public sector's domain remain in their early stages, necessitating further exploration and development. The study sheds light on the underutilization and challenges faced in the PES sector, where scarcity of training data poses significant hurdles. Highlighting the potential of ESCO (European Skills, Competences, Qualifications, and Occupations) classification, the paper underscores its role in facilitating the alignment of occupations with specific skills through AI-driven approaches. By narrowing the training data to ESCO's research occupations and job posts collected from Estonian labor market, the paper lays down the preliminary foundation for constructing a digital career coach as a Virtual Competence Assistant (VCA). Ultimately, the envisioned proof of concept for an AI-based VCA holds the potential to revolutionize the delivery of PES services in the new era of efficiency and effectiveness.

Keywords: Public employment services, Virtual competence assistant, ESCO, Labor market, AI

INTRODUCTION

In the rapidly changing landscape of the 21st century, competency development has emerged as a fundamental element driving individual and societal progress. Technological advancements, economic shifts, and dynamic job markets necessitate a highly adaptable and skilled workforce. Unlike the past where qualifications ensured prolonged job security, the present reality emphasizes the crucial nature of competency development. This development is integral to personal growth, career advancement, economic prosperity, and social well-being. With the increasing integration of automated services, attention needs to be directed toward addressing citizens' needs during challenging times, particularly in the realm of public employment services (PES) supporting the unemployed. The EURES Regulation's Article 19 and its Implementing Decisions, endorsed by the European Commission in July 2018, introduce the use of the European Skills, Competences, and Occupations (ESCO) classification system (European Commission, 2018). It is crucial to highlight that ESCO serves as not only a taxonomy, formalizing hierarchical relationships among concepts and specifying terminology, but also as an ontology, identifying and distinguishing concepts based on their relationships to each other (SkillLab, 2022). A crucial deadline of 31st of

July 2021 has been set for Member States to map their national occupational classifications or national skills classifications to ESCO, or they could choose to directly adopt ESCO for effective implementation. Many developments in the PES sector are already promising in various aspects such as profiling jobseeker's probability of finding work (Desiere & Struyven, 2021) (Leinuste, 2021) and a more general impact evaluation of active labor market measures (Printsmann, 2023). An example of a system that aims to develop an automated skills-based matching tool is the Europass portal (European Union, 2024). Member States are mandated to supply job vacancies and CVs using ESCO codes, which define occupations and skills. These developments have however little influence in making the national PES services more efficient. Moreover, certain scholars are currently directing their attention towards the potential of ESCO in the categorization of occupational postings and corresponding skill sets (Niittylä, 2020), (Zhang et al., 2022), (Zhang et al., 2022), (Zhang et al., 2022). However, these developments are independent of the specific constraints and requirements set by PES. The incorporation and integration of artificial intelligence (AI) into PES activities bring forth a multitude of advantages and benefits, contributing significantly to the enhancement and optimization of various operational aspects and outcomes (OECD, 2022). The European Union (EU) Commission plays a crucial role by providing guidance, coordination, and financial support to ensure that national PES activities align with broader EU employment and social policy objectives. Nevertheless, the duty of designing and executing employment services lies within the jurisdiction of national governments. Estonia has recently announced a new vision called 'Personaalne Riik' (Personal Country), where the idea is to take advantage of the latest technologies already used in the private sector and develop the public services using smart solutions to be modern, efficient and based on the needs of the user (Ministry of Economic Affairs and Communications of Estonia, 2024). Within the Estonian Unemployment Insurance Fund (EUIF), The Estonian PES, individuals registering as unemployed are required to provide information about their job preferences through an e-service called 'e-töötukassa' (Estonian Unemployment Insurance Fund, 2024). Moreover, the service is accessible to all residents and citizens of Estonia. Since the service is linked to the national data exchange layer service (X-Road), the users of that service can effortlessly import data on their prior work experience and education (e-Estonia, 2024). Subsequently, individuals are assigned to a case worker consultant who personally discusses next career options. The limitation of this manual approach lies in the frequent occurrence of individuals aspiring to acquire skills diverging from their educational and experiential background. To address this issue, the development of an AI-enabled Virtual Competence Assistant (VCA) aims to facilitate citizens in effortlessly discovering suitable employment opportunities or receiving training suggestions tailored to their profile. The aim of this research is to lay down the preliminary foundation for constructing a digital career coach as a VCA for job recommendations considering the specific constraints and requirements of PES in Estonia which has been in our focus in previously published research (Liutkevicius & Erlenheim, 2021)

(Liutkevicius & Ben Yahia, 2022). A separate, yet unpublished study investigates the internal requirements of PES, along with user expectations of modern PES self-services and the experiences of AI-experts involved in internal AI-adoption in the EUIF. This investigation offers an insightful perspective on the adoption of AI-enabled services and user readiness, thereby complementing the insights presented in this research. Consequently, our research question is 'what approach can be used to design a proof of concept (POC) for a VCA recommending jobs for citizens in Estonia using the ESCO classification?' Eventually, the VCA will be designed to harness the power of artificial intelligence, leveraging ESCO to match citizens with job opportunities and training programs that are tailored to their skills and ongoing careers.

METHODOLOGY

The research method employed allowed for the design of the POC for the Virtual Competence Assistant (VCA) within a simulated environment tailored to the Estonian context. Given the challenge of working with Estonian language, a low-resource language with limited available data from the labor market, the research utilized tools and models best suited to overcome this constraint and optimize the effectiveness of the POC design process.

STUDY FINDINGS

Data Collection and Cleaning

In the course of this investigation, obtaining an adequate volume of job advertisement data presented a challenge. Numerous websites in Estonia feature job posts (JP); nonetheless, not all are accessible to the public. Starting from 2021 we retrieved job postings through the EUIF open data API (Estonian Unemployment Insurance Fund, 2020). This API granted access to a significant volume of JP with daily updates. However, in November 2023, the API was unexpectedly discontinued from public access. Additionally, despite the EUIF's JavaScript Object Notation (JSON) dataset comprising 53 metadata fields, an issue was identified with respect to the linkage to specific ISCO (The International Standard Classification of Occupations) and ESCO codes. 'ESCO has been built as an extension of the International Standard Classification of Occupations (ISCO)' (European Commission, 2017). Having access to specific dataset linked to classification identifiers would have made the training of the data much easier. It is crucial to underscore that EURES mandates the exchange of labor information, including active job post data from EUIF, with the EU utilizing ISCO and ESCO codes (European Commission, 2024). This implies that information pertaining to ESCO codes linked to job posts already exists in the backend of the e-töötukassa service. The identified deficiency was communicated through meetings and correspondences with EUIF representatives already in 2021. Regrettably, the data was still not added to the public job posts. In addition to JP, we obtained the classification dataset in CSV files from the ESCO portal, selecting the Estonian language for analysis and later training of the model.

In the data cleaning phase, the primary objective was to clean and remove evidently irrelevant columns and documents from the gathered datasets.

Firstly, among all the collected JP data, only the most relevant information was retained for the purpose of training models to identify skills with a clear correlation to the skills in question. The job titles and descriptions were concatenated together and in most cases, they were kept as they are. The texts were cleaned of unnecessary whitespaces and commas. Additionally, when the job description was not presented in the Estonian language it was dropped from the end document. Secondly, from the ESCO dataset, the 'preferredLabel' and 'altLabels' columns were put together as a single column. The column 'skillType' was used to filter with the value 'skill/competence' to get only skills that are relevant to this research study. Additionally, the data field 'conceptURI' was kept for later use connecting the unique skills to the ESCO data portal, which encompasses further relations to specific occupations. All other columns were dropped, as those were not deemed data fields that would have provided additional accuracy for skill classification. Thirdly, from the ESCO dataset, the fields 'prefferedLabel' and 'altlabels' were used for identifying occupations. As most of the occupations vary in the data quality, we decided to narrow the identification only to scientific occupations, which were well described, from both 'prefferedLabel' and 'altlabels' fields. ESCO's 'researchOccupationsCollection' table consisted of 122 professions such as 'biomedical engineer', 'criminologist' linked to a comprehensive list of alternative labels. However, since 'altLabels' were missing translation in the Estonian version. ChatGPT was used to translate the required data into Estonian from the English ESCO version. Additionally, ESCO codes, which are unique identifiers of occupations, were merged from main occupations table of ESCO as they were not present in the 'researchOccupationsCollection' table. During this research, specific information of CV-s obtained from the Estonian University websites, was used only to test the final model.

Data Preprocessing and Model Development

Upon analyzing the JP data, it became apparent that frequently, job descriptions explicitly outline expectations from candidates, specifying mandatory or desirable skills. This kind of information allows for the identification of relevant skills. In line with the study conducted by Zhang, Jensen, and Plank in 2022, focusing on both Danish and English languages, similar findings were observed, enabling the application of analogous logic to extract skills and knowledge from sentences as shown on Figure 1 (Zhang et al., 2022).

The job postings underwent tokenization, and EstNLTK morphological analysis was employed to examine the linguistic structure of the job postings (Laur et al., 2020). Morphological analysis was utilized on JPs to investigate neighbouring words within the participle, detecting particular word combinations such as 'noun + noun in a partitive case'. This analysis facilitated the extraction of additional results to contribute to the input data for skill prediction models. The outcomes played a role in a subtask related to information extraction within the field of Natural Language Processing (NLP), specifically, Named Entity Recognition (NER). The primary goal of NER tagging is to identify and classify named entities within the text into predefined categories, such as individual names, organizational names, locations, and other relevant types of information. Consequently, we followed the Estonian language rules to establish a potential combination of words that is corresponding as a skill.

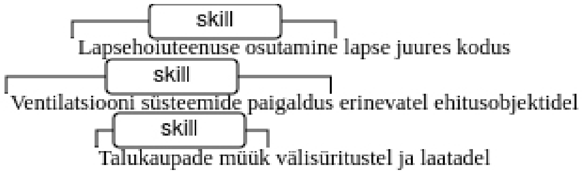

Figure 1: Skill extraction from JP by the example of Zhang, Jensen, and Plank in 2022 *(Zhang et al., 2022).*

After preprocessing the JP data, we built patterns that could be fed into the Prodigy tool that helps to train the model using the Named Entity Recognition (NER) approach with the JP data as input so we could improve the identifications of skills (Prodigy, 2024). The model was built using the 'EstSpaCy' model and the 'et_dep_ud_xlmroberta' based on 'xlm-roberta-base transformers library' (Müürisep, 2021) (Conneau et al., 2019). Finally, the model was evaluated by using a separate set of test data from job postings by validating the skills that were outputted by the model.

In preparing the data for constructing the model to identify occupations from CV-s, the initial step was to convert texts into tagged documents. This process involved transforming the text into a set of words. Following that, common words like articles and linking words were removed as stop words. Additionally, words with a length of 2 or fewer characters were omitted. Each document required a unique tag, and for this purpose, the index number of the document in the dataset was utilized as the tag. In the decision process of which ML technique to use we explored multiple options and existing models, including large language models (LLMs), JobBert (Zhang et al., 2022). However, considering their undeveloped capabilities in Estonian language and low reliability in providing adequate relations between input data and ESCO entities especially in Estonian language, we decided to use Gensim as a preliminary model to be used in the VCA (Řehůřek & Sojka, 2010). Gensim is an open-source Python library designed for natural language processing and topic modelling which is particularly well suited for tasks involving large text corpora, document similarity analysis, and unsupervised machine learning applications (Řehůřek & Sojka, 2010). The Gensim library offered the flexibility to incorporate other tags into the document. To build the model, we followed a similar approach to another research involving the assessment of commodity codes using HS code classification which is similar to ESCO (Spichakova & Haav, 2020). In our scenario, the ESCO code corresponding to the item description served as the second tag. During training, the distributed bag of words (PV-DBOW) was employed as the training algorithm, and the number of iterations (epochs) over the corpus was set to 40. The Doc2Vec model facilitated the discovery of the most similar documents to the original text (Le & Mikolov, 2014). We used 'altlabels' data for training and 'prefferdLabel' for testing the model. If we selected the original text as 'Kriminoloog' and aimed to identify texts resembling it, the process involved several steps. Initially, the original text underwent preprocessing, and it was transformed into a vector using a pretrained Doc2Vec model. Subsequently, the closest texts were determined by assessing cosine vector similarities between the vector representation of the original text and other texts within the vocabulary.

DISCUSSION

Our research has previously explored the situation of PES services in the EU and how implementation of AI can be pursued in the PES based on EUIF as an example. An important purpose of this paper was to experiment how occupational classification ESCO can be used with ML technology in order to recommend relevant JP to citizens. The paper lays down the preliminary foundation for constructing a digital career coach as a VCA utilizing the compulsory occupational classification for the EU member states. Figure 2 illustrates an approach how this could be achieved taking into account the data issues presented in both the EUIF and ESCO's Estonian version. The research shows which classification components of ESCO can be used to get started building the two separate models: (1) converting CV text into ESCO occupations and (2) identifying skills from text based JP. As scientific occupations in ESCO have the 122 occupations well described with a comprehensive list of labels, we could get high similarity scores with the chosen Doc2Vec model. The JP model's biggest challenge was the quality of the JP text often lacking essential description of the tasks and requirements. Hence, the preprocessing needed a lot of manual data annotation help which in our case was a resource intensive task. Following the logic established in this research we propose the proof of concept of how a VCA for recommending jobs to a citizen can be achieved in Figure 2.

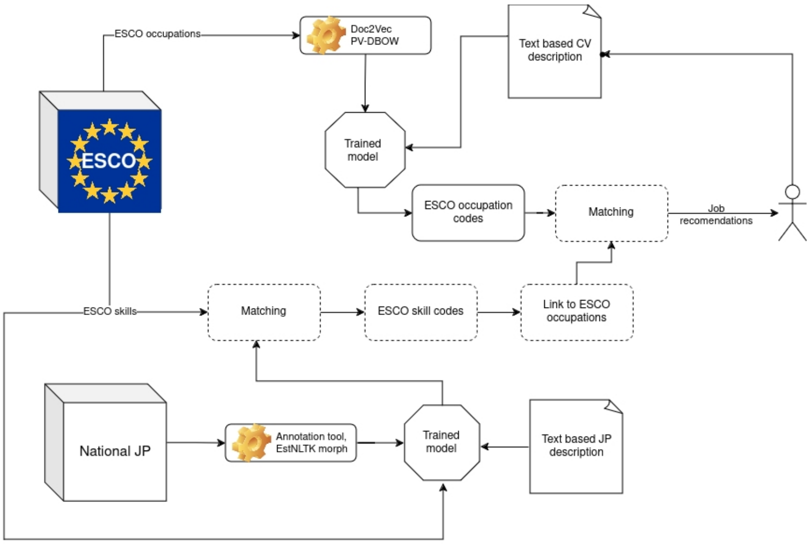

Figure 2: POC for an AI-based VCA.

The concept involves the two models outputting specific ESCO conceptURI-s and ESCO occupational codes with the limitation of describing specific techniques to match the outputs together. Specific matching techniques require further investigation which was not in the scope of this research and are illustrated as dotted-line boxes in Figure 2.

CONCLUSION

In conclusion, many developments the European PES sector are leaning towards the adoption of AI tools. However, these tend to lean towards making the backend tasks more efficient and not so much designing user-friendly services. Our research lays the groundwork for a personalized recommendation tool VCA, making ESCO the default classification. Using tools like EstNLTK, Prodigy, SpaCy and Doc2Vec, show promising outcomes. However, data quality, especially in ESCO's 'altLabels' and JP descriptions need improvement from both the labor market and ESCO. Facilitating experimentation with such models and tools is crucial to make the services offered by governments modern, efficient and based on the needs of the citizen. However, given the closure of access to Estonian JP data by EUIF during the collection phase of this research and not making specific dataset linked to classification identifiers available for training purposes clearly illustrated room for improvement in that area. In essence, our study offers practical insights on how to initiate designing of the AI-based VCA in the evolving landscape of PES. In the future, the adoption of VCA in a country's PES will enable citizens to navigate the complexities of job markets and personal development with greater efficiency. Continuous enhancements in VCA's algorithms ensure it remains at the forefront of personalized career assistance, adapting to evolving job trends and user feedback to deliver increasingly accurate and valuable recommendations.

ACKNOWLEDGMENT

This work in the project 'ICT programme' was supported by the European Union through European Social Fund.

REFERENCES

Conneau, A. K. K. G. N. C. V. W. G. G. F. et al., 2019. *(2019), Unsupervised Cross-lingual Representation Learning at Scale, CoRR*. Online, Association for Computational Linguistics.

Desiere, S. & Struyven, L., 2021. Using Artificial Intelligence to classify Jobseekers: The Accuracy-Equity Trade-off. *Journal of Social Policy,* 50(2), pp. 367–385.

e-Estonia, 2024. *Interoperability services - X-Road*. [Online] Available at: https://e-estonia.com/solutions/interoperability-services/x-road/ [Accessed 28 02 2024].

Estonian Unemployment Insurance Fund, 2020. *Job postings API of The Estonian Unemployment Insurance Fund*. [Online] Available at: https://www.tootukassa.ee/et/api/toopakkumised [Accessed 27 02 2024].

Estonian Unemployment Insurance Fund, 2024. *E-service "e-töötukassa"*. [Online] Available at: https://www.tootukassa.ee/etootukassa/eng/logi-sisse [Accessed 28 02 2024].

European Commission, 2024. *How is ESCO used by Public Employment Services in Europe and in EURES?*. [Online] Available at: https://esco.ec.europa.eu/en/node/171 [Accessed 20 02 2024].

European Commission, 2017. *ESCO: harnessing the power of big data analysis for a functional labour market,* Luxembourg: Publications Office of the European Union.

European Commission, 2018. Commision implementing decision (EU) 2018/1020 of 18 July 2018 on the adoption and updating of the list of skills, competences and occupations of the European classification for the purpose of automated matching through the EURES common IT platform. *Official Journal of the European Union.*

European Union, 2024. *Europass (February 27, 2024).* [Online] Available at: https://europa.eu/europass/en [Accessed 27 02 2024].

Laur, S., Orasmaa, S., Särg, D. & Tammo, P., 2020. *EstNLTK 1.6: Remastered Estonian NLP Pipeline.* Marseille, European Language Resources Association.

Leinuste, K., 2021. *Decision support tool "OTT" for employment counsellors in the Estonian PES. Development, implementation and future plans,* s.l.: OECD.

Le, Q. & Mikolov, T., 2014. *Distributed representations of sentences and documents.* s.l., International conference on machine learning.

Liutkevicius, M. & Ben Yahia, S., 2022. Research roadmap for designing a Virtual Competence Assistant for the European labour market. *Procedia Computer Science,* Volume 207, pp. 2404–2413.

Liutkevicius, M. & Erlenheim, R., 2021. *Validating the usage of occupational classification systems in the process of creating a national virtual competency assistant within the EU labor market.* Athens, Proceedings of the 14th International Conference on Theory and Practice of Electronic Governance (ICEGOV 2021).

Ministry of Economic Affairs and Communications of Estonia, 2024. *Personaalne riik,* Tallinn: Ministry of Economic Affairs and Communications of Estonia.

Müürisep, K., 2021. *SpaCy pipelines for Estonian language.* [Online] Available at: https://github.com/EstSyntax/EstSpaCy/ [Accessed 28 02 2024].

Niittylä, H., 2020. *Job Market Finland.* [Online] Available at: https://tyomarkkinatori.fi/en/blogit/vieraskyna-niittyla-tekoalyn-hyodyntaminen-esco-osaamisten-haussa [Accessed 15 02 2024].

OECD, 2022. *Harnessing digitalisation in Public Employment Services to connect people with jobs,* Paris: Policy Brief on Active Labour Market Policies.

Printsmann, P. E., 2023. *Automated tool for impact evaluation of active labour market measures: Linking administrative data,* Paris: OECD.

Prodigy, 2024. *Radically efficient machine teaching. An annotation tool powered by active learning.* [Online] Available at: https://prodi.gy [Accessed 27 02 2024].

Řehůřek, R. & Sojka, P., 2010. *Software framework for topic modelling with large corpora.* Valletta, LREC Workshop on New Challenges for NLP Frameworks.

SkillLab, 2022. *ESCO and the Creation of a Data-Rich Environment, An Opportunity for the EU Labour Market,* Amsterdam: SkillLab.

Spichakova, M. & Haav, H.-M., 2020. *Using Machine Learning for Automated Assessment of Misclassification of Goods for Fraud Detection.* Tallinn, Databases and Information Systems (DB&IS).

Zhang, M., Jensen, K. N. & Plank, B., 2022. *Kompetencer: Fine-grained Skill Classification in Danish Job Postings via Distant Supervision and Transfer Learning. Language Resources and Evaluation Conference,.* Marseille, LREC 2022.

Zhang, M., Jensen, K. N., Sonniks, S. D. & Plank, B., 2022. *SkillSpan: Hard and Soft Skill Extraction from English Job Postings. Association for Computational Linguistics,*. Seattle, Association for Computational Linguistics.

Zhang, M., Jensen, K. N., van der Goot, R. & Barbara, P., 2022. *Skill Extraction from Job Postings using Weak Supervision.* Aalborg, CEUR Workshop Proceedings.