European Commission

Machine Learning Assisted Mapping of Multilingual Occupational Data to ESCO

Report – October 2022

Employment, Social Affairs and Inclusion

Table of contents

1. Introduction

In our previous articles, we illustrated how maintenance and enrichment of the ESCO occupations pillar can be supported by artificial intelligence. The ability to link external data to an ESCO occupation concept is an essential building block of this process and it supports drafting new occupations and quality control on existing ones. In this report we touch upon a significant challenge that complicates this process. ESCO is currently supporting 28 languages which requires *multilingual* machine learning models to connect textual information to the ESCO occupations. This report discusses the multilingual mapping approach that the ESCO team established to support the maintenance of ESCO and applies it to different use-cases for illustrative purposes.

In Section 2 we present the use-case of mapping national occupational classifications to ESCO within the context of EURES. Section 3 focusses on the more general challenge of mapping free text to ESCO occupations. Section 4 provides insights in the internals of the multilingual modelling approach by visualising the multilingual embedding space and inspecting semantic relatedness between phrases and occupation titles. Possible extension to unseen languages, so-called zero-shot cross-lingual transfer, is investigated in Section 5.

2. Mapping multilingual occupational classifications to ESCO

EURES is a European cooperation network of employment services, designed to facilitate the free movement of workers[1]. As part of Regulation (EU) 2016/589 of the European Parliament and of the Council, Member States need to adopt the ESCO taxonomy at national level or map their national classifications to ESCO in order to exchange data via a standardised terminology of occupations, skills and competences with the EURES platform.

A software application was developed in order to support Member States in the process of mapping their national classification to ESCO. After importing the national classification in the tool, a Member State can start creating links between the ESCO taxonomy and the national

[1] https://ec.europa.eu/eures/

classification. Functionality is included to browse the source and target classifications and retrieve additional information. Also, potentially related ESCO concepts are suggested for a concept from the national classification through a TF-IDF-based approach to facilitate the manual mapping task. Some Member States worked outside the mapping platform and contracted external parties to support the mapping exercise. Several of these external parties reported the use of artificial intelligence to assist the mapping, but the applied methodology was not always made public. All this led to a very fragmented approach for what is essentially a common challenge faced by all Member States.

In addition, synergies were identified with the Europass project, where work history from Europass users is linked to the ESCO classification[2]. Given that the ESCO team is continuously working on updating the taxonomy and further optimising algorithms that support this process, it is beneficial to follow an integrated approach such that ESCO stakeholders as Europass and EURES can also benefit from the machine learning models that result from the maintenance tasks of the ESCO team.

Starting from the work that is performed within ESCO and considering all these stakeholder use-cases, the ESCO team decided to design a representation learning approach that supports efficient extension to various languages. The multilingual XLM-RoBERTa [3] model was selected and finetuned on labour market data such as ESCO, QDR qualifications [4] and EURES online job advertisements covering 28 ESCO languages. In this section we evaluate this finetuned model that is used for ESCO maintenance purposes on the mapping tables that gradually become available from the mapping efforts of the Member States to comply with regulation EU 2016/589 [5] and O*NET [6].

Table 1 presents occupation concepts selected from the national classifications of Latvia, Spain, Sweden, Italy and United States of America. The four Member States were selected for this comparison because we did not use EURES online job advertisements with the corresponding languages of these Member States to train the representation learning model. However, we did use the respective ESCO language variants for training the model such that the model at least learned about the languages of those countries while finetuning. Table 1 shows the top three suggestions and scores by the machine learning model for the source concept. The last column represents the expert validation as obtained from the mapping tables that the Member State provided. The types of matches are exact match, broader match, narrower match and close match[7]. In case the column is empty, the link between the source concept and the ESCO concept was not selected by the experts. As the results show,

[2] https://europa.eu/europass/
[3] Alexis Conneau, Kartikay Khandelwal, Naman Goyal, Vishrav Chaudhary, Guillaume Wenzek, Francisco Guzman, Edouard Grave, Myle Ott, Luke Zettlemoyer and Veselin Stoyanov, Unsupervised cross-lingual representation learning at scale, Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, 8440-8451, 2020.
[4] https://europa.eu/europass/qdr/
[5] https://esco.ec.europa.eu/en/use-esco/eures-countries-mapping-tables
[6] https://esco.ec.europa.eu/en/use-esco/other-crosswalks
[7] https://esco.ec.europa.eu/system/files/2021-07/425b7a5f-3048-4377-a816-5402c00e9a9505_A_Annex_Draft_ESCO_Implementation_manual.pdf

the multilingual model is able to suggest ESCO concepts that were selected by experts. Note: Table 6 in Appendix provides all information in English to ease interpretation.

Table 1: ESCO occupation suggestions (ESCO suggestion) for concepts (Source concept) from national occupational classifications (Source classification). Suggestion match Score and Expert validation feedback are included.

Source classification (Country)	Source concept	Score	ESCO suggestion	Expert validation
Latvijas Profesiju klasifikators (LV)	Bagarēšanas mašīnu operators (8342.01)	95	bagarēšanas mašīnu operators	exact
		91	ekskavatora vadītājs	
		89	buldozera vadītājs	
CO-SISPE 2011 (ES)	Entrevistadores/ Encuestadores (44301013)	93	encuestador/ encuestadora	exact
		87	encuestador de estudio de mercado/ encuestadora de estudio de mercado	broad
		77	responsable de estudios de campo	
Swedish Standard Classification of Occupations (SE)	Distributionschef , logistik och transporter (TN3u_buC_wJF)	90	distributionschef	narrow
		90	distributionschef, avfall och skrot	broad
		89	distributionschef, levande djur	broad
CP2011 (IT)	Conduttori di gru e di apparecchi di sollevamento (7.4.4.3.0)	96	gruista	close
		96	gruista di banchina	close
		95	operatore di gru semoventi/operatrice di gru semoventi	close
O*NET (US)	Aviation Inspectors (53-6051.01)	97	aviation inspector	exact
		92	avionics inspector	
		91	aircraft assembly inspector	broad

Giabelli et al. presented a study where they benchmarked a machine learning approach to map the Italian occupational classification to 426 level 4 ISCO concepts and reported a top 5 accuracy of 0.8 [8]. Table 2 presents benchmark results for the finetuned multilingual model for mapping five national classifications to ESCO occupation concepts based on the available mapping tables. Member State taxonomies were mapped to ESCO v1.0.7 (2,942 concepts) given that major version was used by the Member States, while O*NET was mapped to ESCO v1.1 (3,008 concepts) as this crosswalk was completed more recently. To compute the micro

Table 2: Benchmark results for suggesting ESCO occupations for concepts from different occupational classifications. Micro mean reciprocal rank (MRR) and Top k accuracy for mapping to different language variants (ESCO language) of ESCO and for different Input types (only title versus title and desription).

			MRR	Accuracy						
Country (# mapped occupations)	**Input**	**ESCO language**	**Micro**	**Top 1**	**Top 2**	**Top 3**	**Top 5**	**Top 10**	**Top 20**	**Top 30**
LV (2747)	Title	EN	0.5719	0.4536	0.5784	0.6400	0.7153	0.7961	0.8591	0.8890
		LV	0.6231	0.5169	0.6298	0.6837	0.7543	0.8231	0.8824	0.9057
ES (2180)	Title	EN	0.6150	0.4982	0.6284	0.6890	0.7560	0.8326	0.8872	0.9101
		ES	0.6358	0.5266	0.6422	0.7064	0.7729	0.8413	0.8945	0.9142
SE (2292)	Title	EN	0.6033	0.4838	0.6143	0.6824	0.7426	0.8229	0.8866	0.9149
		SV	0.6689	0.5620	0.6837	0.7365	0.7958	0.8687	0.9136	0.9402
IT (814)	Title	EN	0.6666	0.5725	0.6806	0.7260	0.7789	0.8391	0.8821	0.8956
		IT	0.6930	0.6093	0.7052	0.7506	0.7887	0.8342	0.8894	0.9029
	Title, Desc	EN	0.7141	0.6278	0.7224	0.7617	0.8145	0.8796	0.9214	0.9447
		IT	0.7308	0.6413	0.7494	0.7887	0.8280	0.8907	0.9361	0.9558
O*NET (940)	Title	EN	0.8671	0.8128	0.8809	0.9106	0.9362	0.9553	0.9734	0.9819
	Title, Desc	EN	0.8811	0.8351	0.8915	0.9138	0.9415	0.9564	0.9755	0.9787

[8] Anna Giabelli, Lorenzo Malandri, Fabio Mercorio and Mario Mezzanzanica, WETA: Automatic taxonomy alignment via word embeddings, Computers in Industry, Volume 138, 103626, 2022.

mean reciprocal rank and top *k* accuracy, the highest ranked ESCO suggestion as approved by the validators across all four match types was used. We also provide insights in the effect of mapping to a different language variant of ESCO, e.g. the Latvian occupations were separately mapped to the English and Latvian ESCO variant. In addition, for the Italian classification and O*NET we used two different types of input: only the title of occupation concept versus occupation concept title combined with description.

The results in Table 2 show that model suggestions consistently match better with expert validation when mapping to the ESCO variant having the same language as the source classification. Using more input information (i.e. title and description) further improves suggestion quality. We did not use examples of professions (e.g. job titles) as input in the experiments, but we suspect it should further improve the performance metrics. In conclusion, we observe that top 5 accuracy ranges between 0.75 and 0.83 for the Member State classifications in Table 2 and reaches 0.94 for O*NET. This illustrates that a single multilingual representation learning approach as established for ESCO maintenance can be used to support experts in mapping different national classifications to ESCO, thereby being a step forward compared to a simpler TF-IDF approach as currently available in the platform.

3. Mapping multilingual content to ESCO

As described above, establishing a crosswalk between different taxonomies is just one of the applications for a multilingual occupation mapping model. The ESCO team uses similar methodology to connect potentially new occupations to the classification or for investigating potentially new alternative labels[9].

Table 3 provides ESCO suggestions as obtained by the multilingual model for a number of input texts covering different languages and different input types. The first six examples are more straightforward cases as is reflected in the scores; they mainly contain job title-like terminology. The Swedish example (SV) is showing that a 13-year-old looking for an assistant is likely referring to a nanny although not clearly described. The last two Dutch (NL) examples come from the domain of supporting elderly and indicate the link between the specific task and the information that is in the descriptions of the ESCO occupations. I.e. someone who is a companion '*may do shopping activities as well as punctual transportation to doctor's appointments, etc*' while the more general '*caring for an older person*' results in the occupation residential home older adult care worker.

[9] https://esco.ec.europa.eu/en/about-esco/data-science-and-esco/leveraging-artificial-intelligence-update-esco-occupations-pillar

Table 3: ESCO occupation suggestions (ESCO suggestion) for input text (Input) in different languages. Suggestion match Score and English translation of the input are included for ease of interpretation.

Input (language)	Input – English translation	Score	ESCO suggestion
Sosiaalityöntekijä, Oulun kaupunki, Kiimingin lastensuojelu, Oulu (FI)	Social worker, City of Oulu, Kiiming child protection, Oulu	86	child welfare worker
Werkvoorbereider Tekenaar | AutoCad, Noordwijkerhout (NL)	Planner Draftsman | AutoCad, Noordwijkerhout	88	drafter
Майстер нігтьового сервісу (манікюр, педикюр) (UK)	Nail service master (manicure, pedicure)	93	manicurist
Sprzedawca-kasjer w stacji paliw 1/2 etatu (PL)	Salesperson-cashier at a gas station 1/2 time	91	fuel station specialised seller
Salvatore cerca una baby sitter a Fiano Romano (IT)	Salvatore is looking for a babysitter in Fiano Romano	90	babysitter
Djelatnik / ica za strojna čišćenja i hortikulturu (HR)	Employee for machine cleaning and horticulture	89	horticulture worker
Drømmer du om at blive selvstændig, og har du erfaring med enten rekruttering, salg, ledelse eller måske alle tre dele? (DA)	Do you dream of becoming self-employed and do you have experience in either recruitment, sales, management or perhaps all three parts?	80	recruitment consultant
Glad och busig kille på 13 år söker påhittig och ansvarfull assistent 75% (SV)	Happy and mischievous 13-year-old boy looking for resourceful and responsible assistant 75%	79	nanny
Boodschappen doen voor een oudere persoon (NL)	Grocery shopping for an older person	83	companion
Zorgen voor een oudere persoon (NL)	Caring for an older person	85	residential home older adult care worker

To conclude this part of the more qualitative analysis of the model, Figure 1 presents a multilingual version of the English marketing-related job titles figure that was published in a previous article. Figure 10 in Appendix shows translated labels for Figure 1 to ease interpretation.

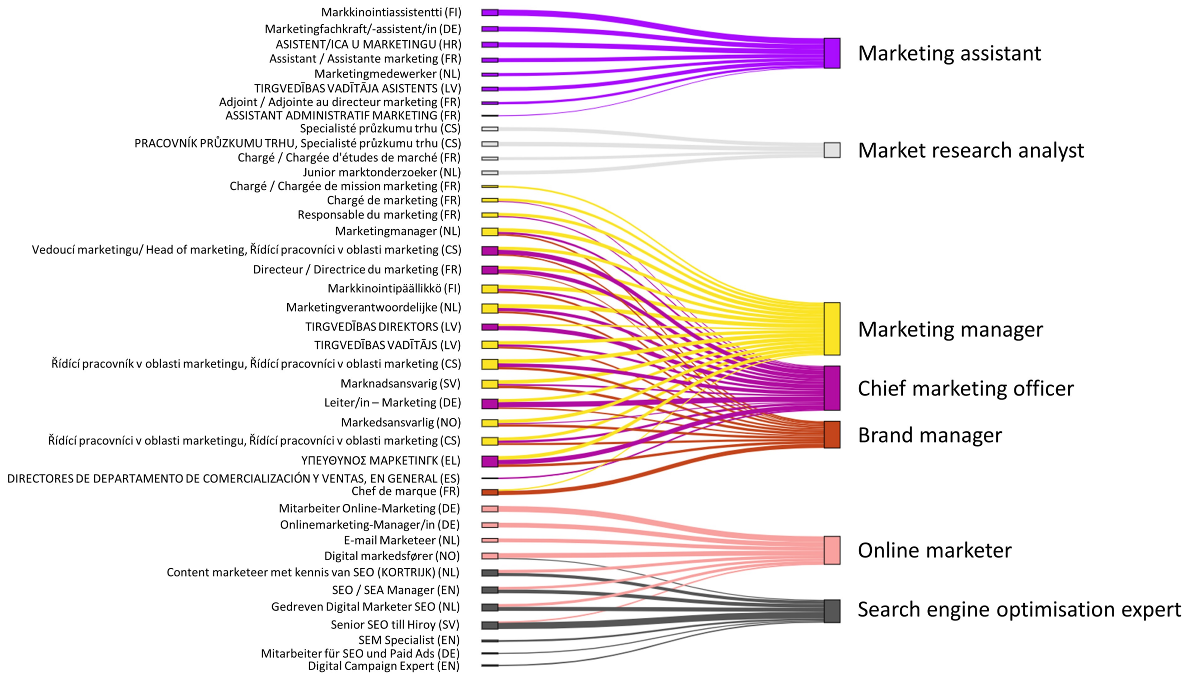

Figure 1: ESCO occupation suggestions from the marketing domain across different languages.

Recently, studies by Decorte et al.[10] and Zbib et al.[11] described approaches to compute job title similarities and perform job title normalisation. The authors highlight the limitations of a supervised method and present alternative approaches from representation learning. From our work on maintaining and updating ESCO, we also acknowledge the drawbacks of a purely supervised approach. ESCO contains over 3,000 occupation concepts in 28 languages and is constantly getting updated, which makes creating and maintaining supervised data sets for mapping to ESCO occupations challenging.

Thanks to the study by Decorte et al., a benchmark dataset of English job titles together with their corresponding ESCO occupation became available. The test dataset as published by Decorte et al. contains 15,463 job titles from a broad range of industry sectors and all the records have one of the 2,675 corresponding ESCO leaf occupations linked to them. The benchmark results from Decorte et al. and Zbib et al. are summarised in Table 4, together with the results from the ESCO multilingual model. ESCO results were obtained based on ESCO v1.0.7 and mapping to the English ESCO language variant.

[10] Jens-Joris Decorte, Jeroen Van Hautte, Thomas Demeester and Chris Develder, JobBERT: understanding job titles through skills, FEAST, ECML-PKDD 2021 Workshop, Proceedings, 2021.

[11] Rabih Zbib, Lucas Alvarez, Federico Retyk, Rus Poves, Juan Aizpuru, Hermenegildo Fabregat, Vaidotas Simkus and Emilia García-Casademont, Learning job titles similarity from noisy skill labels, Unpublished manuscript, 2022.

Table 4: Benchmark results for suggesting ESCO occupations by different approaches for the data set provided by Decorte et al. *Results from the multilingual model as presented in this report are indicated by ESCO. Macro, micro mean reciprocal rank (MRR) and Recall@k were only partially available from the report of Zbib et al.*

	MRR		Recall@1		Recall@5		Recall@10	
Method	**Macro**	**Micro**	**Macro**	**Micro**	**Macro**	**Micro**	**Macro**	**Micro**
Decorte et al.	0.3641	0.3092	0.2666	0.2248	0.4632	0.3865	0.5440	0.4604
Zbib et al.		0.3414				0.4595		0.5400
ESCO	0.3791	0.4113	0.2786	0.3138	0.4938	0.5214	0.5707	0.6025

ESCO team had an exchange with the authors of Decorte et al., where the authors explained that their skill extraction model had a significant impact on the final results and that by improving this step they were able to further improve the results, albeit unclear to which extent for us. Zbib et al. only report partial results as such these statistics are missing from Table 4.

While we very much appreciate the work by Decorte et al. and Zbib et al. and the release of a benchmark dataset, we also agree with Decorte et al. that a significant amount of cases from the test dataset are mislabelled, making these comparisons challenging. Also as we described in a previous news article, it is difficult to map to ESCO occupations by only taking the job title into account[12]. Due to the granularity of the classification it is beneficial to include a description of tasks, skills or knowledge (if available) for mapping purposes.

4. Gaining further insights into the model

In order to inspect how the multilingual model is representing text, we select a set of 34,904 unique job titles, covering 24 ESCO languages, from the EURES platform. As there is a dependency on what Member States provide and because of the respective size of the labour market, the number of records can vary significantly by language. For each of the job titles the model embedding is computed and reduced to two dimensions by UMAP (n_neighbors=20, min_dist=0.3, cosine distance metric). In addition, the model is used to map the multilingual job titles to the English variant of ESCO in order to simplify this process and, finally, only the highest ranked ESCO occupation suggestion is kept. Figure 2 visualises the embedding space for the set of 34,904 job titles together with the level 2 ISCO code as derived from the predicted ESCO occupation.

[12] https://esco.ec.europa.eu/en/about-esco/data-science-and-esco/role-contextual-information-when-connecting-data-esco-occupations-pillar-using-artificial

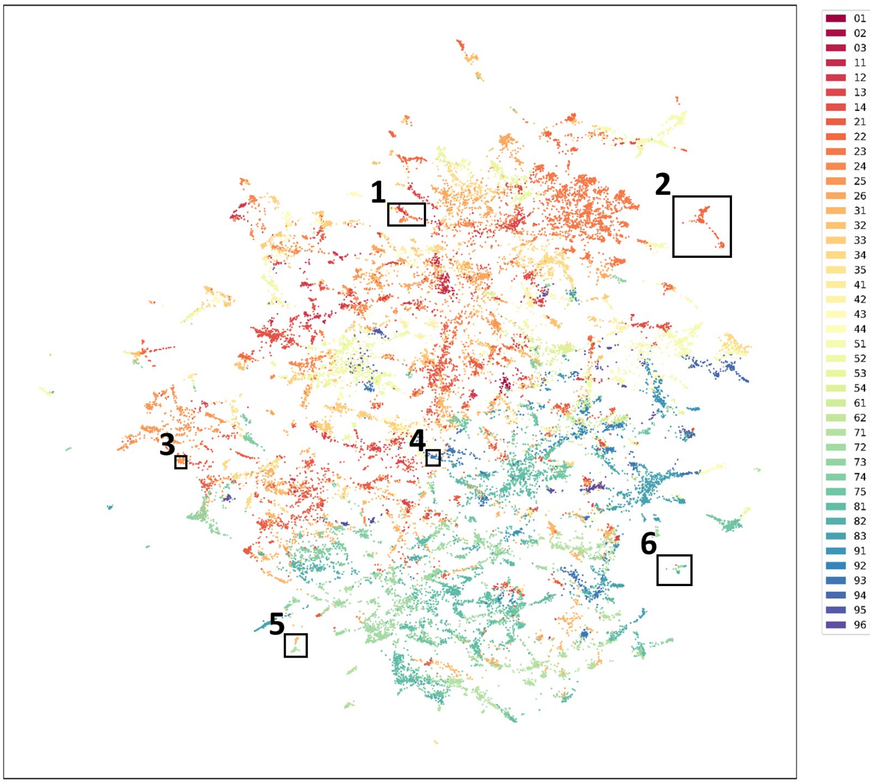

Figure 2: UMAP reduced embedding space for 34,904 multilingual job titles from the EURES portal. Colours represent level 2 ISCO code for top 1 suggested ESCO occupation as obtained by the machine learning model on the job title.

In Figures 3 – 8 we focus on different areas of the embedding space, thereby covering different parts of the European labour market. In contrast to Figure 2 the colour coding reflects the predicted ESCO occupation instead of ISCO groups. The labels in the figures denote the job titles together with their language code between parentheses. From these figures it is clear that closely related job titles from different languages are close together in the embedding space. There does not seem to be apparently unrelated job titles from a single language clustered together. Also, seniority levels appear to be grouped together across the different languages. These results tend to suggest there is alignment between the different ESCO languages for the finetuned model. Note: Figures 11 – 16 in Appendix present English translations for the original labels in Figures 3 – 8 in order to make the results easier to interpret.

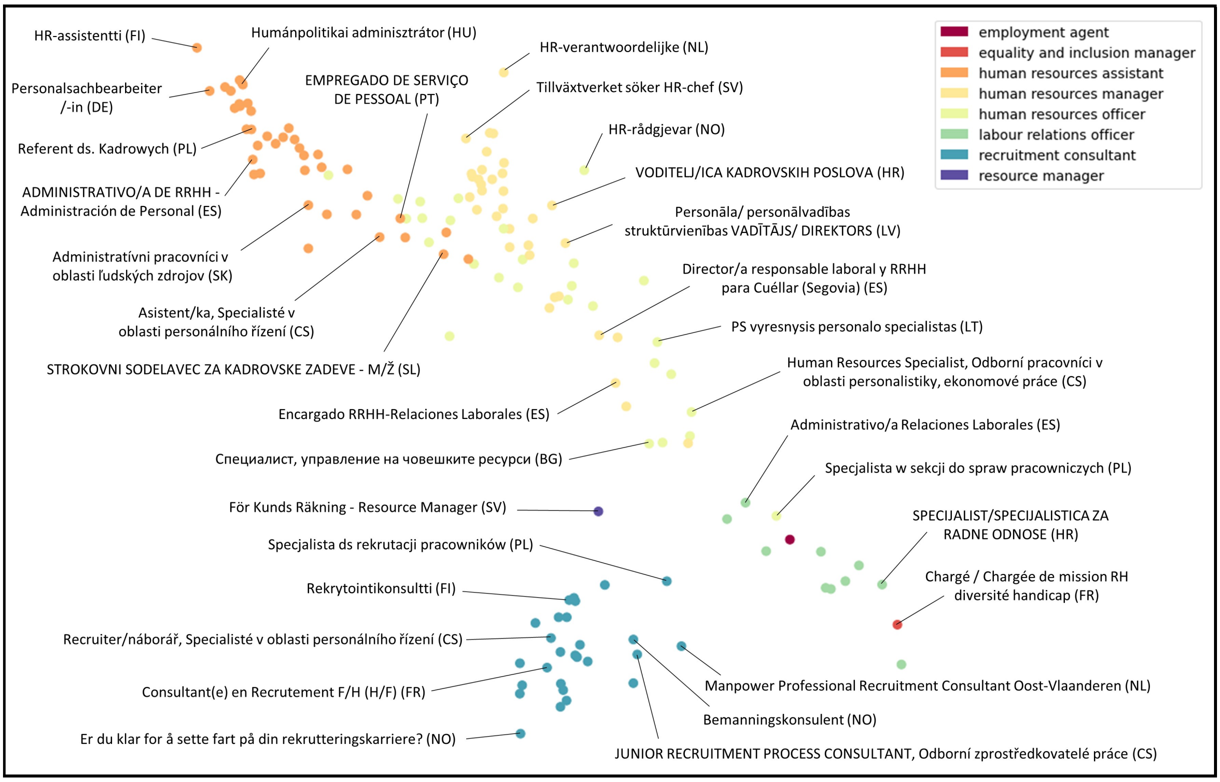

Figure 3: Detailed embedding space from area 1 where colours represent the top 1 suggested ESCO occupation as obtained by feeding the job title to the machine learning model.

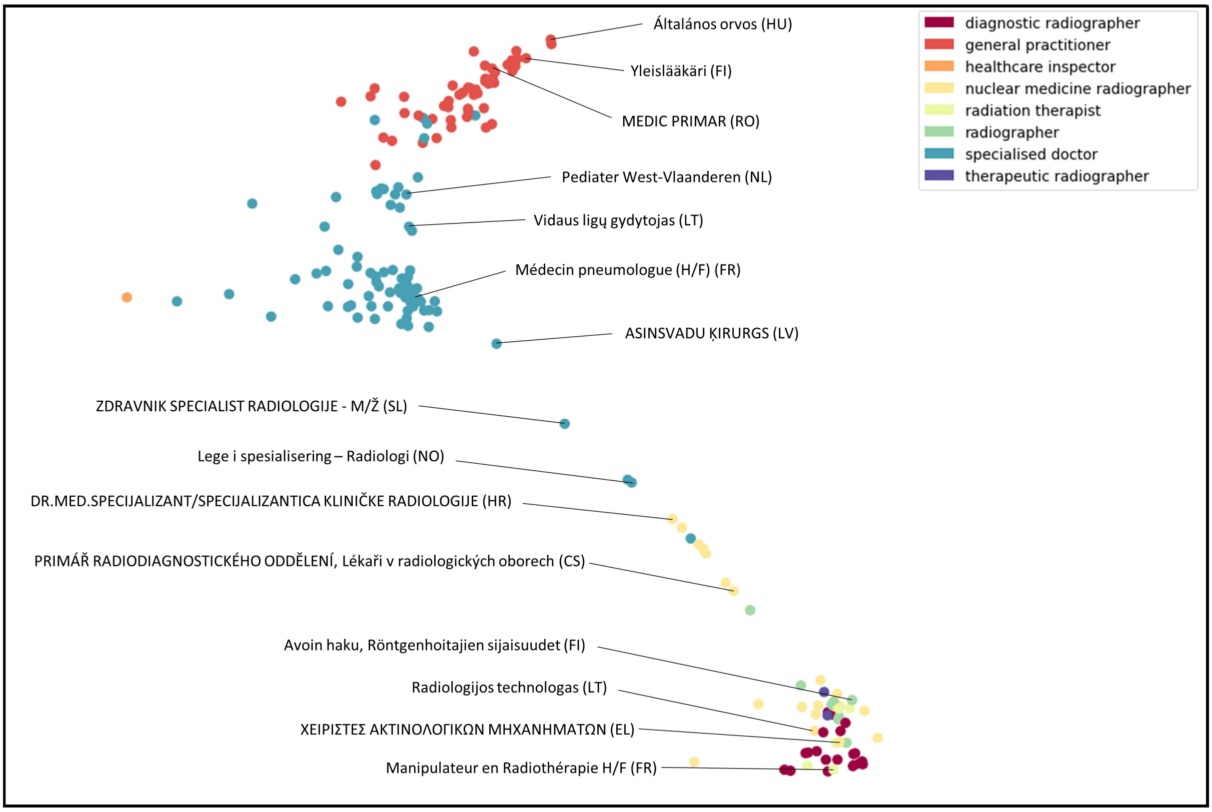

Figure 4: Detailed embedding space from area 2.

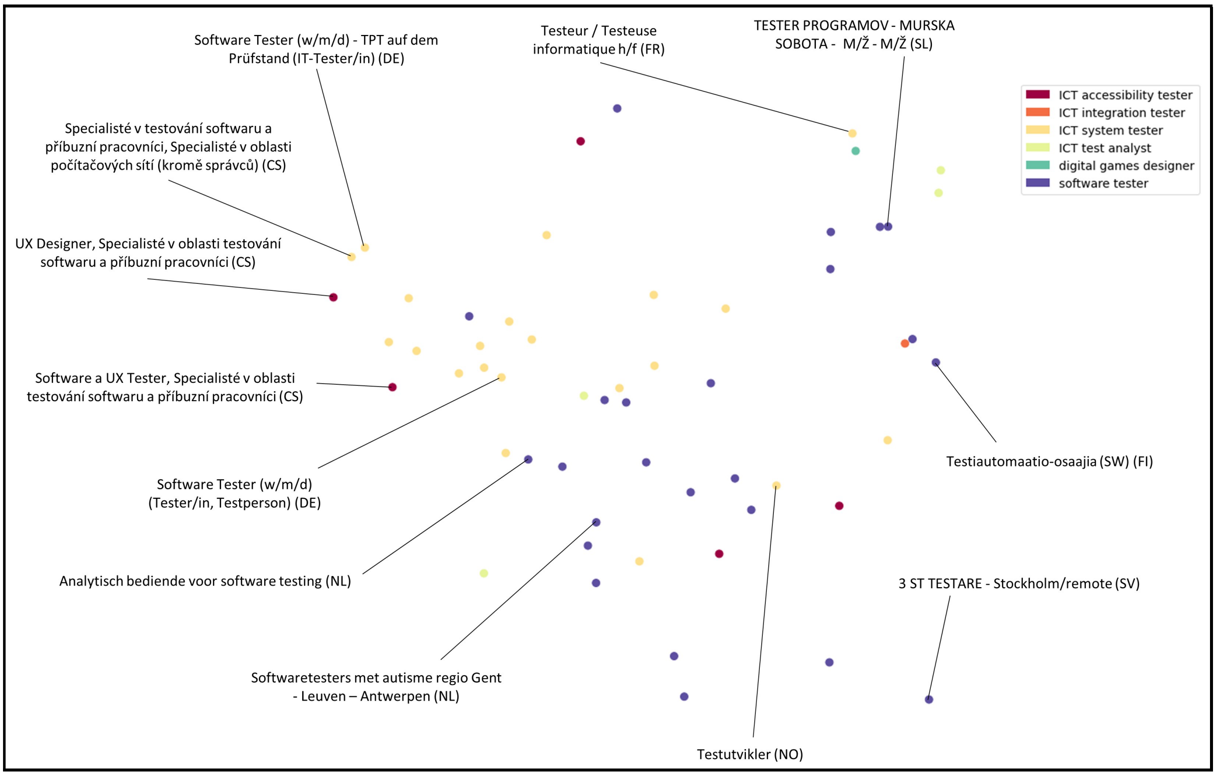

Figure 5: Detailed embedding space from area 3.

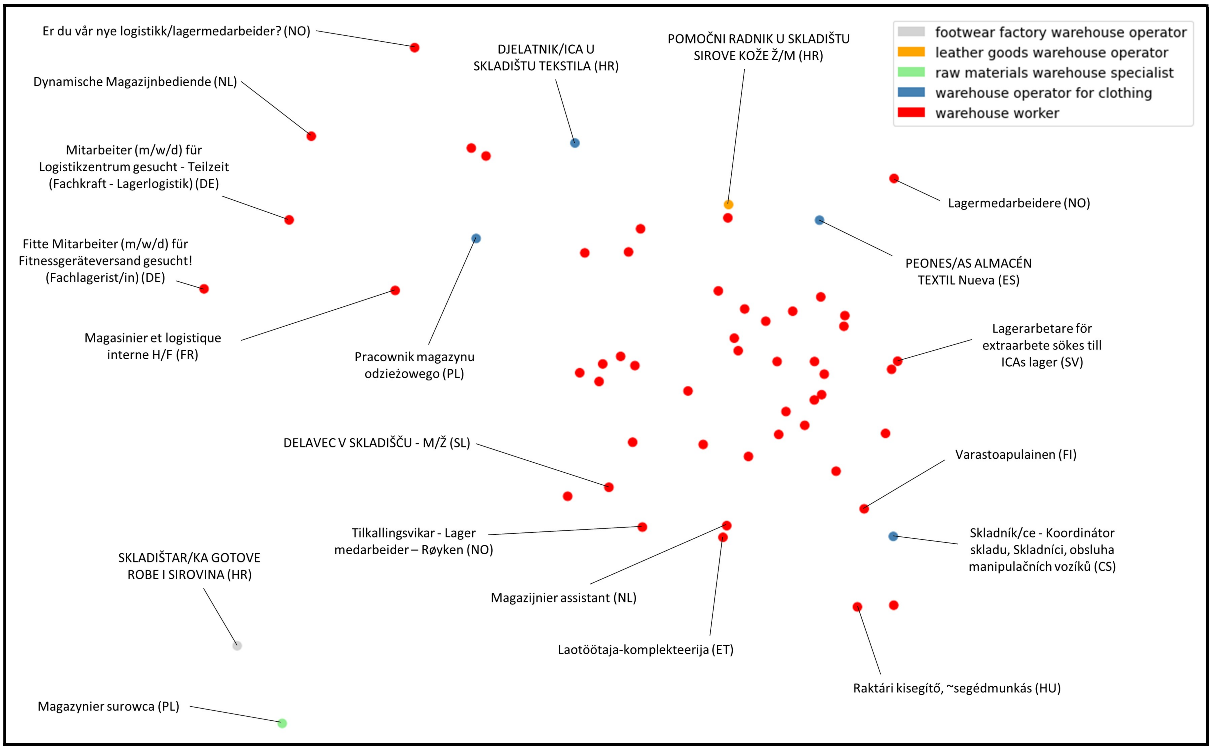

Figure 6: Detailed embedding space from area 4.

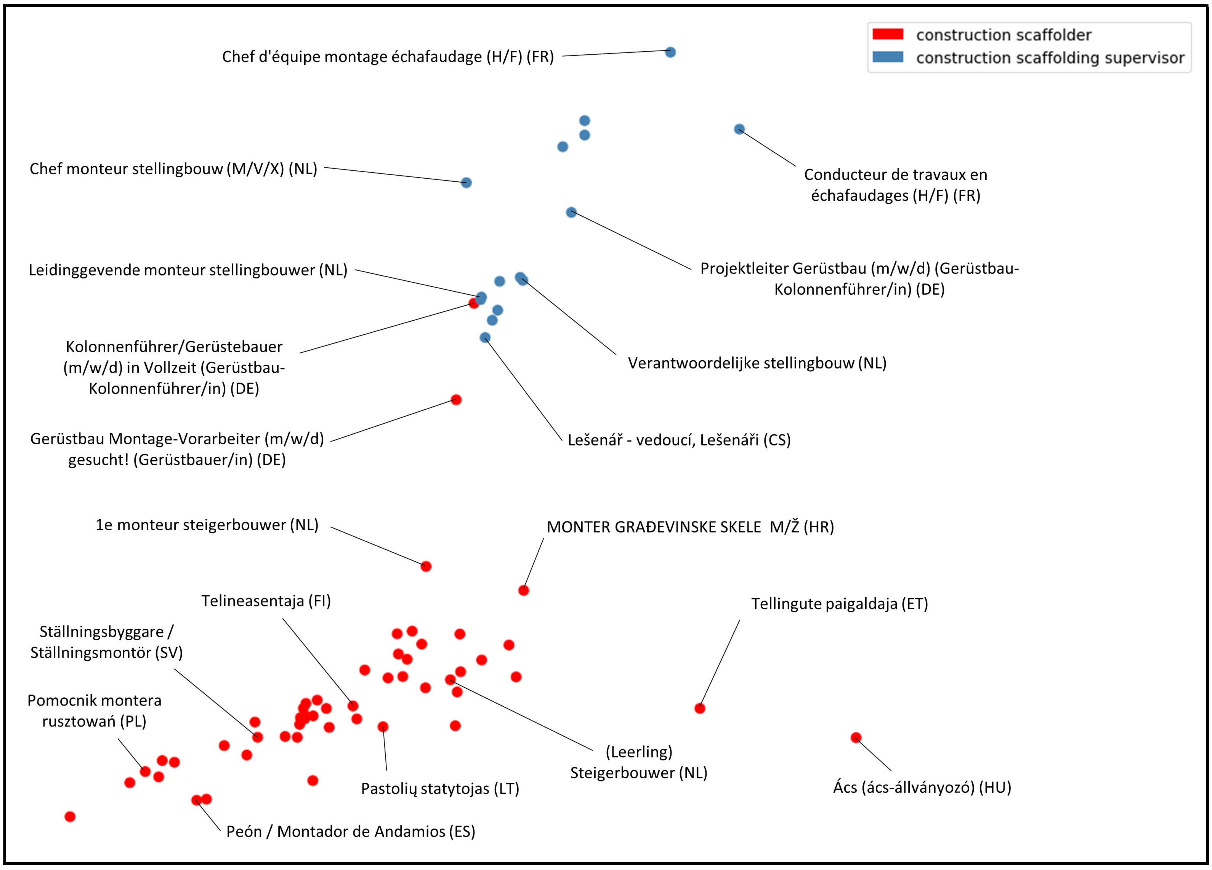

Figure 7: Detailed embedding space from area 5.

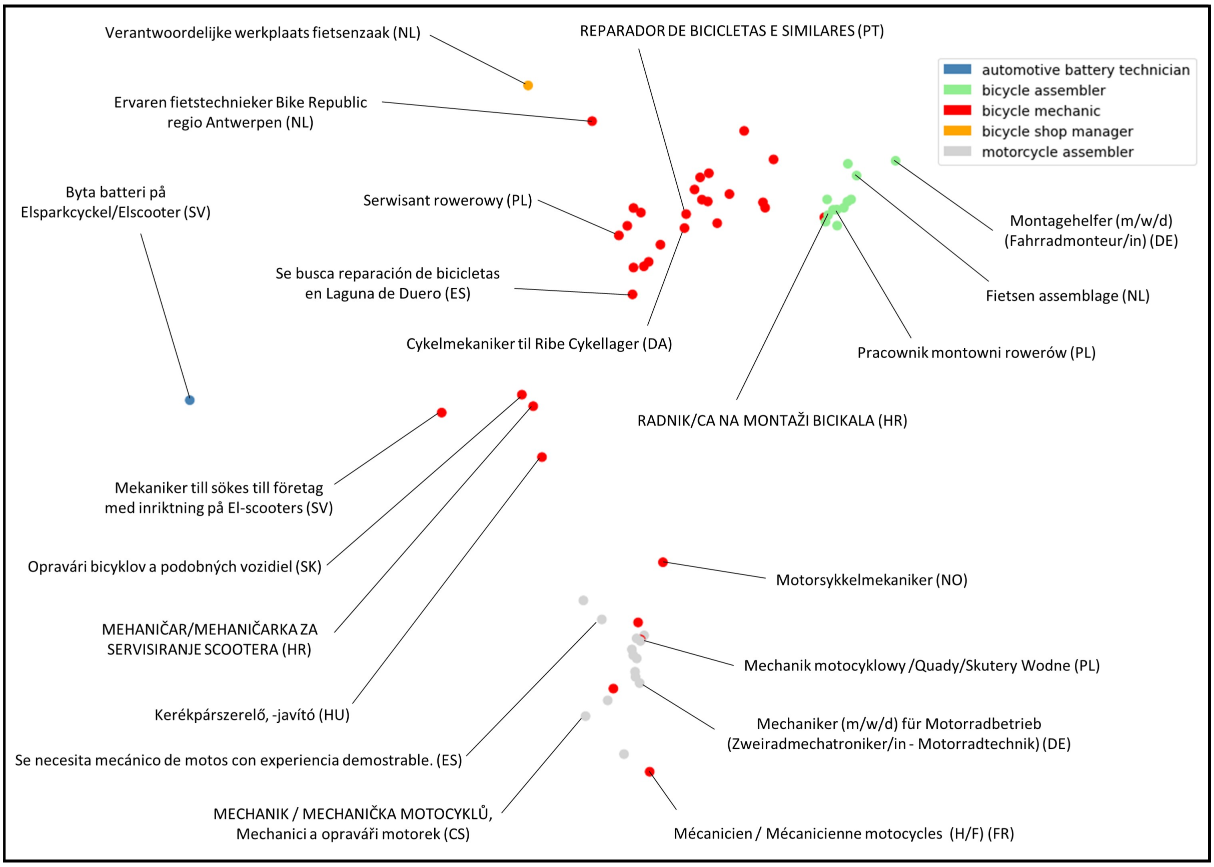

Figure 8: Detailed embedding space from area 6.

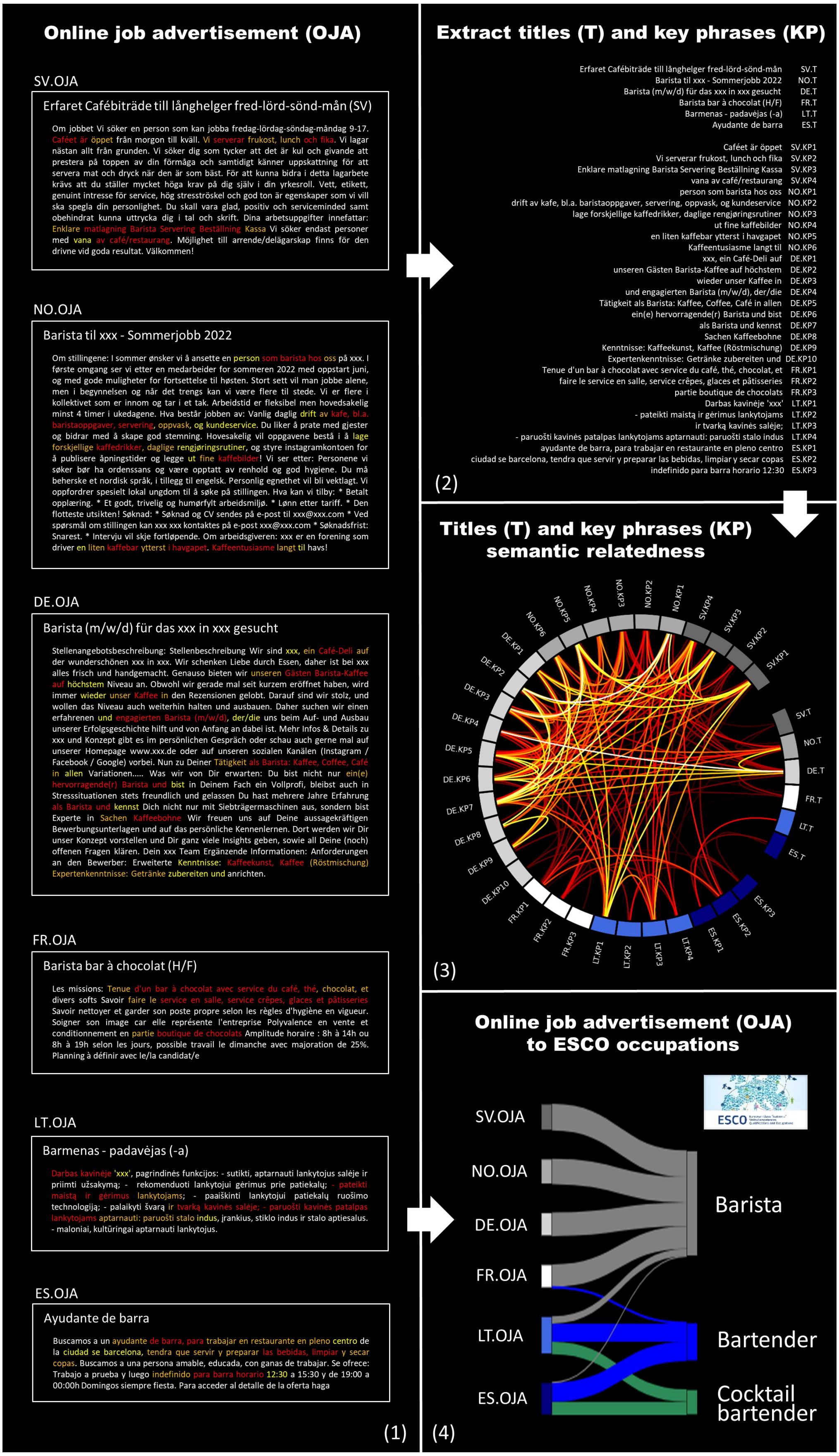

Figure 9: (1) Online job advertisements with key phrases. (2) Extracted key phrases with identifiers. (3) Semantic relatedness between key phrases and online job advertisement titles. (4) Suggested ESCO occupations.

The representation learning model is finetuned on multilingual ESCO skills, ESCO occupations, QDR qualifications and EURES online job advertisements without a filtering step in contrast to the approach by Decorte et al. and Zbib et al. This means that the model had to learn from raw text (such as full descriptions of online job advertisements) the parts that were most relevant. To understand to which extent this has happened we extracted six online job advertisements from EURES, corresponding to the following languages: Swedish, Norwegian, German, French, Lithuanian and Spanish. Figure 9 panel (1) shows the titles and descriptions for the online job advertisements and colours the parts of the descriptions that have the largest contribution to the predicted ESCO occupation suggestion (i.e. so-called key phrases). The underlying continuous score contributions are mapped to a discrete four-level scale (white, yellow, orange, red) ranging from minimal to maximal contribution. These key phrases, together with the online job advertisement titles, are also listed in panel (2) and identifiers are added for ease of interpretation. The circle diagram in panel (3) visualises the semantic relatedness between all key phrases and the titles. The results show cross-lingual patterns originating from the alignment of the multilingual embedding space. A higher level of semantic relatedness exists between key phrases and titles for the Swedish, Norwegian, German and French online job advertisements, which are all for barista roles. The Lithuanian and Spanish online job advertisements are for a bartender position and show, for example, relations between the Lithuanian title and the Spanish key phrases. Finally, panel (4) presents the ESCO occupation suggestions by feeding complete online job advertisements to the model.

5. Mapping from unseen languages

As explained in Section 3, an important aspect when establishing the methodology was to not rely on purely supervised techniques. Another significant factor was to opt for methodology that can be further improved for existing ESCO languages without difficulty (e.g. when future labour market data become available), but also methodology that can be extended to other languages with minimal effort. This guarantees that ESCO can support implementers all across Europe and can have maximum impact. For example, recently a Ukrainian variant of ESCO was created[13]. This means that mapping free text in Ukrainian to ESCO can now be done in a more efficient way by directly mapping Ukrainian input to Ukrainian ESCO terminology. From the mapping methodology point of view it meant that one can align the Ukrainian variant of ESCO directly with the other language variants of ESCO and extend the mapping algorithm. Additionally, this alignment over 28 languages enables languages that are less developed in terms of ESCO content (e.g. some languages have no/fewer alternative terms compared to others), can possibly take advantage of mapping to terminology of languages with more extensive ESCO content. However, it is also possible to connect input from a non-ESCO language to the ESCO taxonomy by mapping to any of the existing ESCO language variants.

[13] https://esco.ec.europa.eu/en/news/esco-now-available-ukrainian-language

The representation learning model that we developed is based on XLM-RoBERTa, which is pretrained on 100 languages. This model is finetuned to map text to ESCO occupations in 28

Table 5: ESCO occupation suggestions (ESCO suggestion) for input text (Input) in different unseen languages. Suggestion match Score and English translation of the input are included for ease of interpretation.

Input (language)	Input – English translation	Score	ESCO suggestion
Šef smene u skladištu logistike mleka i mlečnih proizvoda (SR)	Head of the shift in the logistics warehouse of milk and dairy products	89	dairy products and edible oils distribution manager
Vozač drumskog vozila u domaćem i međunarodnom drumskom saobraćaju (SR)	Driver of a road vehicle in domestic and international road traffic	90	cargo vehicle driver
Zamenik poslovođe maloprodajnog objekta (SR)	Deputy manager of a retail store	87	department store manager
Per EKSELENTET! Intership/praktike: Kontabilitet/Fiskalitet VLORE (SQ)	For EXCELLENCES! Internship/practice: Accounting/Fiscal VLORE	78	accounting assistant
Zhvillues Software, Sektori i Zhvillimit IT (SQ)	Software Developer, IT Development Sector	83	software developer
ofroj vend pune Inxhinier Mjedisi ne degen e Energjetikes (SQ)	I offer a job as an Environmental Engineer in the Energy department	87	energy engineer
마케팅전문가 (KO)	marketing expert	92	marketing consultant
제조업 사무직 자재구매 및 입출고 사무업무 (매입,매출전표관리) 거래처 발주 접수 및 거래명세서 작성 (KO)	Manufacturing office worker Material purchasing and warehouse/shipping office work (purchase and sales slip management) Receiving customer orders and preparing transaction details	81	purchaser
화학제품생산직(기계조작) 표면처리 약품 제조 회사 입니다. 약품 원료를 가지고 배합하여 제조 및 말통 포장 하는 업무. 주 5 일 근무 9 시-18 시 (KO)	Chemical product production job (machine operation) We are a manufacturer of surface treatment chemicals. Manufacturing and packaging of pharmaceutical raw materials by mixing them. Working 5 days a week 9:00 - 18:00	90	chemical mixer

languages at the moment. For some languages (e.g. Ukrainian) the finetuning is only based on ESCO data, while for other languages it is based on combinations of ESCO, qualifications, national classifications and online job advisements. As expected, the amount of language specific training data that are available has an impact on the mapping accuracy. To extend our model to new languages, the optimal approach would be to further finetune the model with labour market data from the corresponding languages.

In case no training data is available for a language for which we would like to map to ESCO, we are essentially dealing with zero-shot cross-lingual transfer learning. It was reported by Huang et al. that XLM-RoBERTa does not align well across languages although it learns language encoders having a shared multilingual contextual embedding space[14]. Huang et al. described that the multilingual encoders fail to capture similarity when the source and target languages are less similar at levels of morphology, syntax, and semantics. While here we further finetuned the XLM-RoBERTa shared embedding space for a selected number of languages, we also investigated mapping for languages that were not in the training set. Table 5 contains suggested ESCO occupations for mapping Serbian (SR) and Albanian (SQ) input text to the English variant of ESCO. A more extreme test was performed for Korean (KO). These examples show that transfer to these languages holds to some extent. While the examples were selected based on their higher mapping score (indicating higher confidence in the suggestion), we also found that extending to this zero-shot setting by focussing on unseen languages remains challenging in general. Further finetuning the model with labour market data from an unseen language represents the only viable approach.

6. Summary

This report illustrated based on different use-cases the approach that the ESCO team is following for multilingual modelling to support the maintenance of the occupations pillar. Representation learning techniques are used to represent free text originating from the 28 ESCO languages, thereby aligning the embedding space for the different languages. The approach is flexible in terms of extending the model to new languages. Because of the alignment in the multilingual embedding space, it is possible to map between different languages. This has the advantage that ESCO languages with less terminology (e.g. alternative titles) can benefit from ESCO languages with more extensive content. The model is based on XLM-RoBERTa and was finetuned on labour market data (ESCO, qualifications and online job advertisements) from 28 languages and we benchmarked it for different alternative use-cases. Results on supporting mapping Member State classifications to ESCO and suggesting potentially relevant ESCO occupations for job titles showed the approach is promising. Finally, insights are provided to inspect some of the internals of the methodology. Two-dimensional

[14] Kuan-Hao Huang, Wasi Ahmad, Nanyun Peng and Kai-Wei Chang, Improving zero-shot cross-lingual transfer learning via robust training, Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing, 1684-1697, 2021.

visualisations of the embedding space show that multilingual job titles cluster together and semantic relatedness between job titles and online job advertisement description phrases are presented.

At the moment, the ESCO team is already working on the continuous update of the model because of additional feedback coming from the task of maintaining ESCO and the availability of more labour market data. While the original training dataset consisted of tens of millions of records, the coverage for some languages was rather low at the time. Continuously more data became available which should have a substantial impact on the results for some low-resource languages. In addition, ESCO team aims to integrate the model in the mapping platform as a more powerful alternative compared to the existing TF-IDF approach and build APIs such that other stakeholders (e.g. Europass) can also take advantage of our expertise in this area.

If you are an ESCO implementer and want to share your feedback, please get in touch via email at EMPL-ESCO-SECRETARIAT@ec.europa.eu or use our hashtag #ESCO_EU.

7. Appendix

Table 6: ESCO occupation suggestions (ESCO suggestion) for concepts (Source concept) from national occupational classifications (Source classification). Suggestion match Score and Expert validation feedback are included (English version).

Source classification (Country)	Source concept	Score	ESCO suggestion	Expert validation
Latvijas Profesiju klasifikators (LV)	Dredger operator (8342.01)	95	dredge operator	exact
		91	excavator operator	
		89	bulldozer operator	
CO-SISPE 2011 (ES)	Interviewers/ Enumerators (44301013)	93	survey enumerator	exact
		87	market research interviewer	broad
		77	field survey manager	
Swedish Standard Classification of Occupations (SE)	Distribution manager, logistics and transport (TN3u_buC_wJF)	90	distribution manager	narrow
		90	waste and scrap distribution manager	broad
		89	live animals distribution manager	broad
CP2011 (IT)	Operators of cranes and lifting equipment (7.4.4.3.0)	96	tower crane operator	close
		96	container crane operator	close
		95	mobile crane operator	close
O*NET (US)	Aviation Inspectors (53-6051.01)	97	aviation inspector	exact
		92	avionics inspector	
		91	aircraft assembly inspector	broad

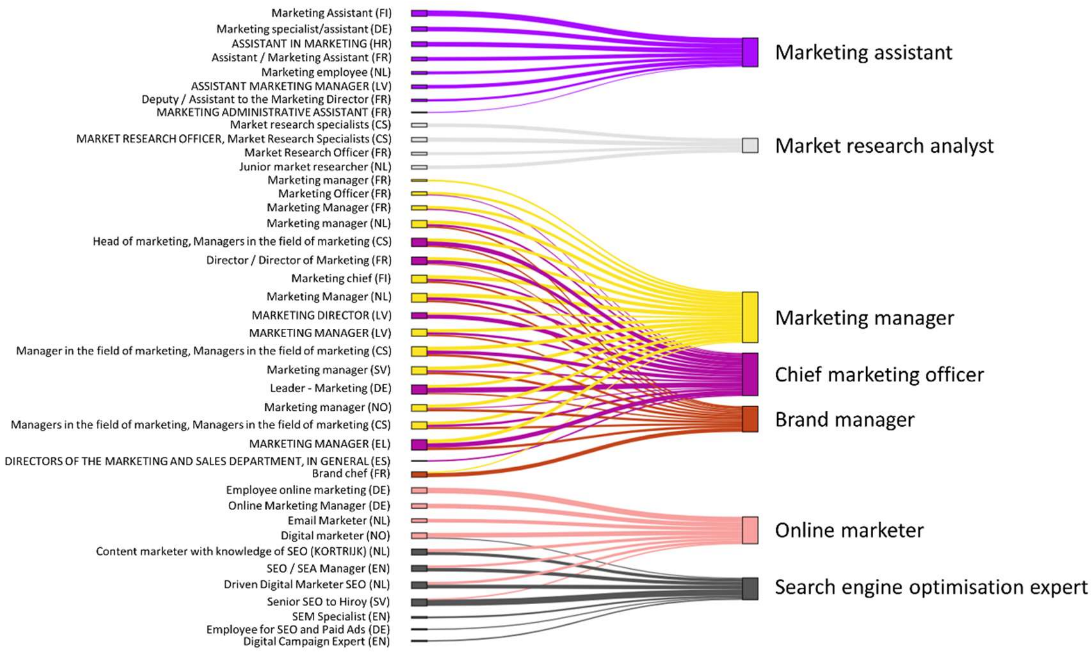

Figure 10: ESCO occupation suggestions from the marketing domain across different languages (English version).

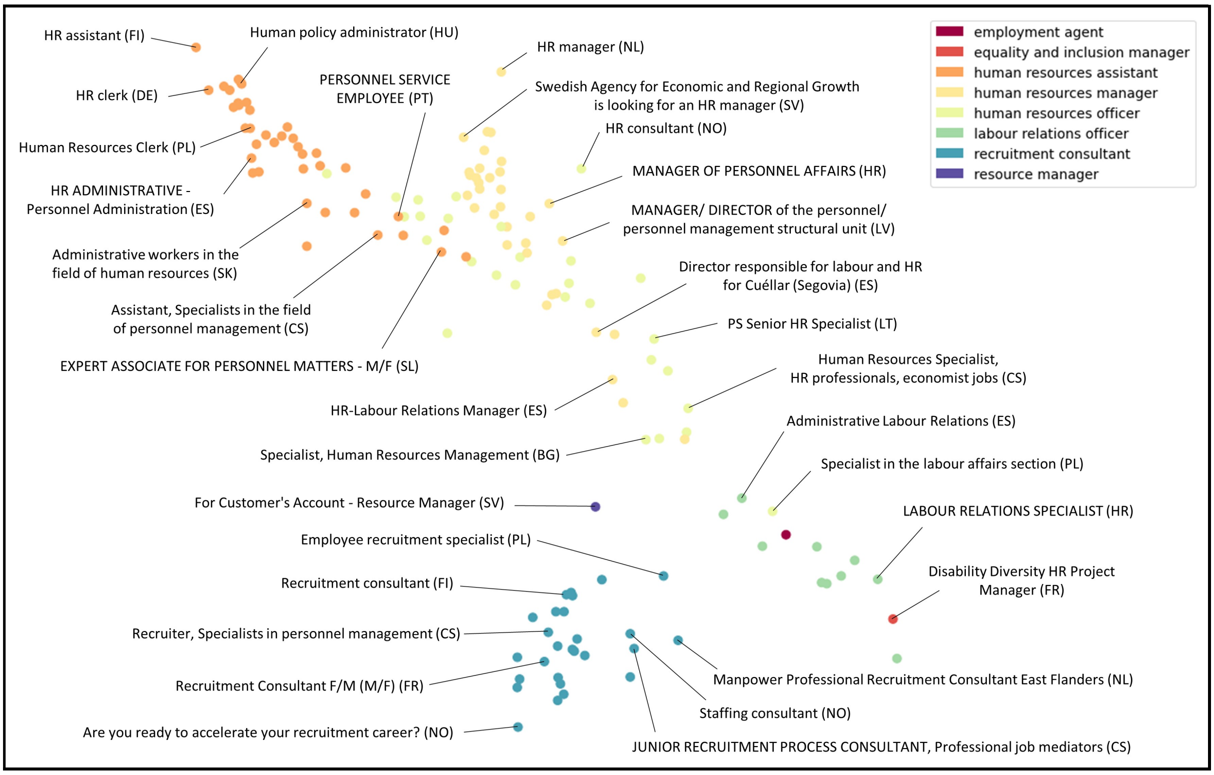

Figure 11: Detailed embedding space from area 1 (English version).

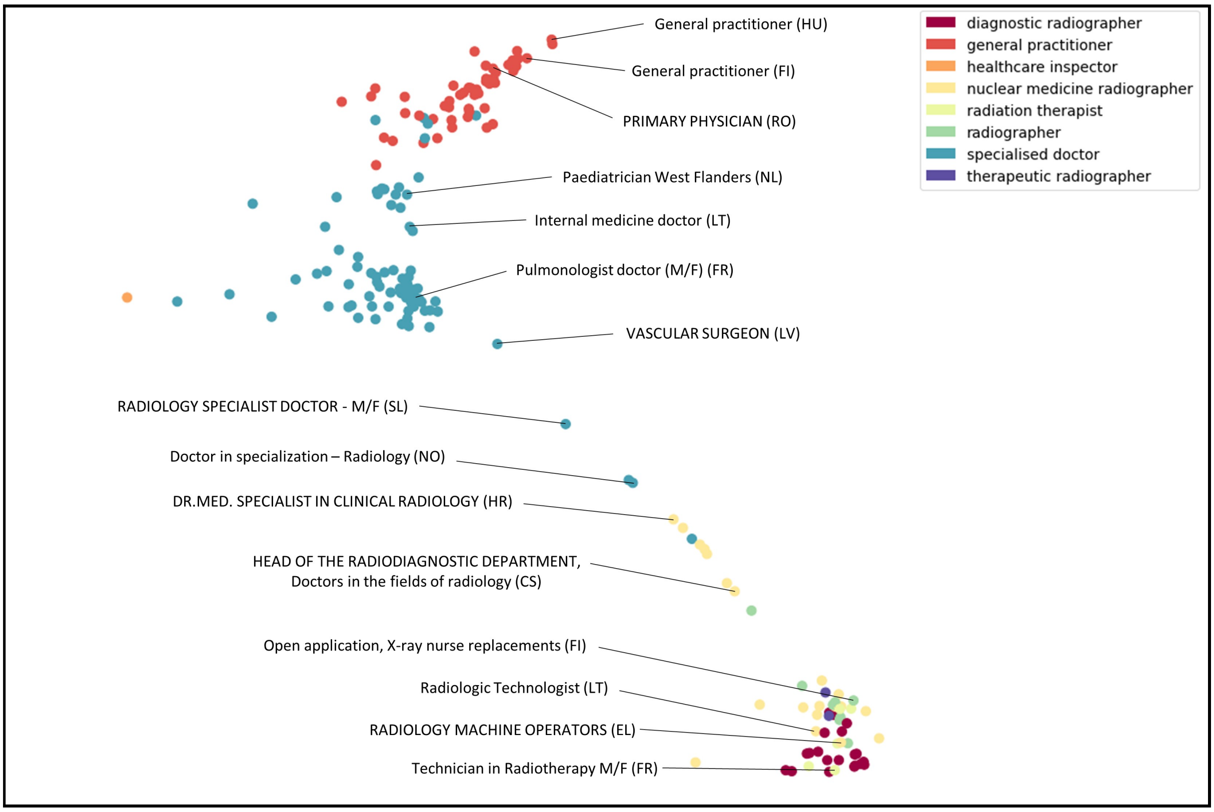

Figure 12: Detailed embedding space from area 2 (English version).

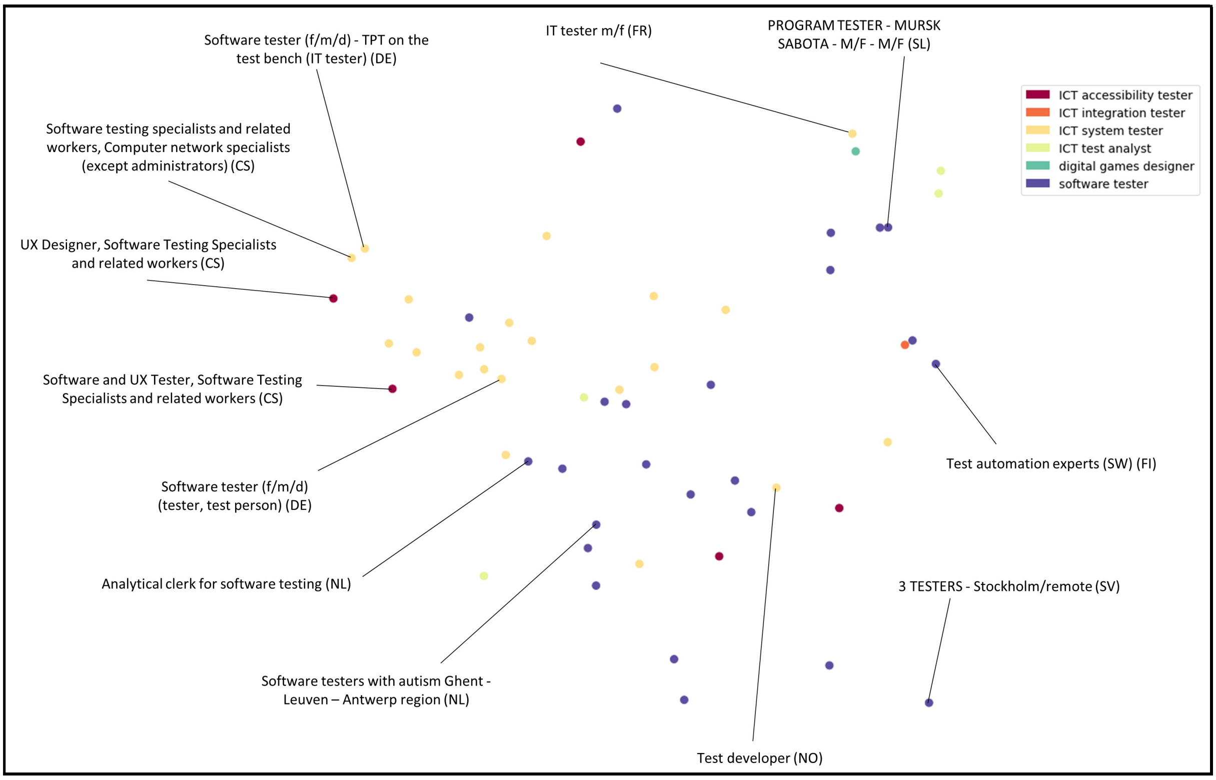

Figure 13: Detailed embedding space from area 3 (English version).

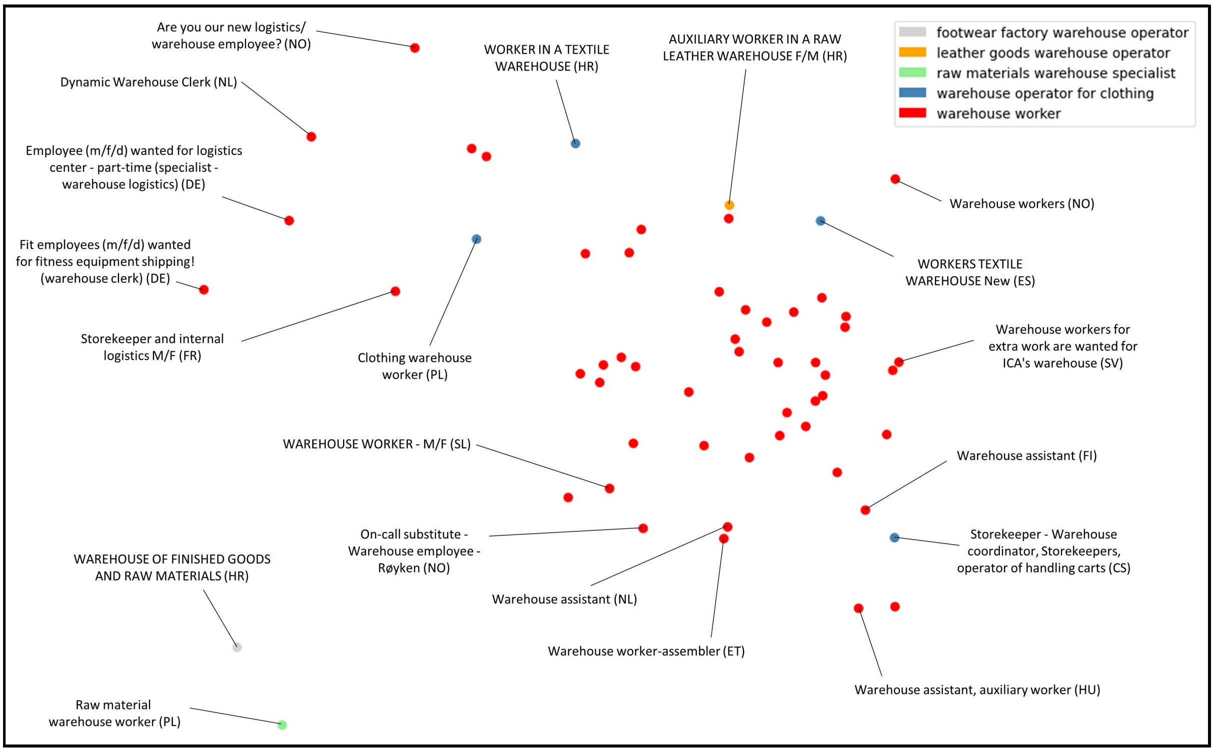

Figure 14: Detailed embedding space from area 4 (English version).

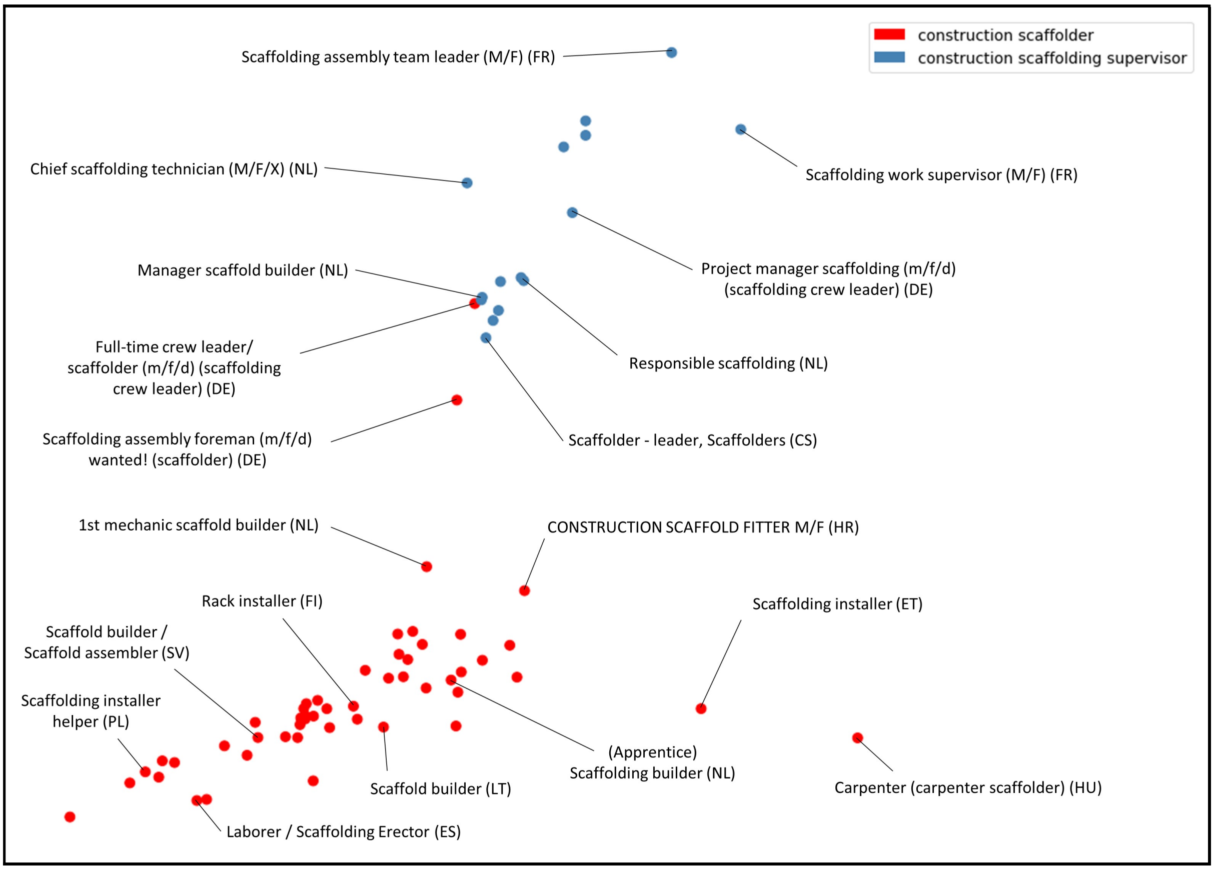

Figure 15: Detailed embedding space from area 5 (English version).

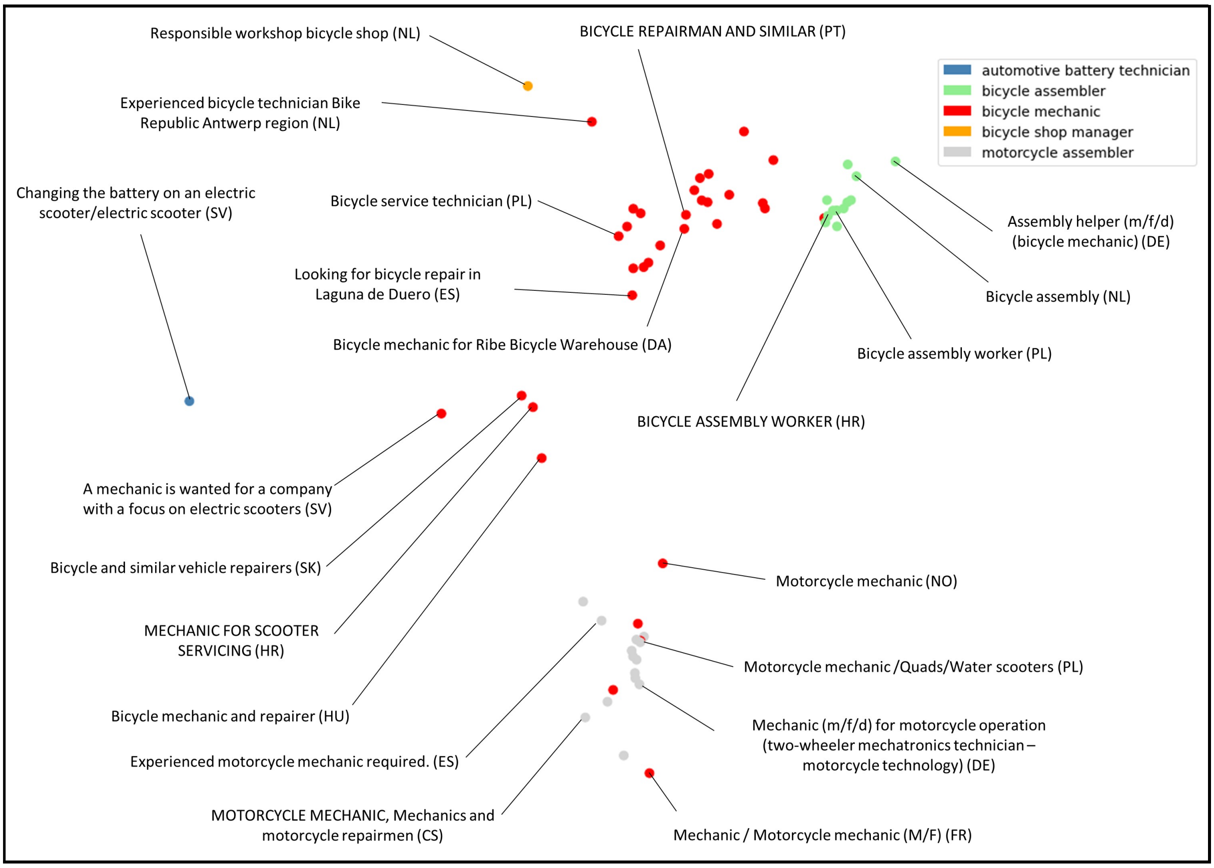

Figure 16: Detailed embedding space from area 6 (English version).