NLPnorth @ TalentCLEF 2025: Comparing Discriminative, Contrastive, and Prompt-Based Methods for Job Title and Skill Matching

Mike Zhang[1,*,†], Rob van der Goot[2,†]

[1]Aalborg University, Copenhagen, Denmark
[2]IT University of Copenhagen, Copenhagen, Denmark

Abstract

Matching job titles is a highly relevant task in the computational job market domain, as it improves e.g., automatic candidate matching, career path prediction, and job market analysis. Furthermore, aligning job titles to job skills can be considered an extension to this task, with similar relevance for the same downstream tasks. In this report, we outline NLPnorth's submission to TalentCLEF 2025, which includes both of these tasks: Multilingual Job Title Matching, and Job Title-Based Skill Prediction. For both tasks we compare (fine-tuned) classification-based, (fine-tuned) contrastive-based, and prompting methods. We observe that for Task A, our prompting approach performs best with an average of 0.492 mean average precision (MAP) on test data, averaged over English, Spanish, and German. For Task B, we obtain an MAP of 0.290 on test data with our fine-tuned classification-based approach. Additionally, we made use of extra data by pulling all the language-specific titles and corresponding *descriptions* from ESCO for each job and skill. Overall, we find that the largest multilingual language models perform best for both tasks. Per the provisional results and only counting the unique teams, the ranking on Task A is 5th/20 and for Task B 3rd/14.

Keywords

computational job market analysis, NLP for Human Resources, job title matching, job-skill matching, classification, contrastive learning, prompting, large language models

1. Introduction

The dynamic and rapidly evolving nature of labor markets, primarily driven by technological advancements [1, 2, 3, 4], global migration patterns [5], digitization [6], and economic shifts, has significantly increased the availability of detailed job advertisement data across various recruitment and employment platforms. These platforms actively leverage job postings to attract qualified candidates, generating rich and structured datasets that are highly valuable for labor market analysis [7, 8, 9]. Consequently, there has been substantial growth in research related to Natural Language Processing (NLP) applications for Human Resources [9, 10] and Computational Job Market Analysis [11, 12].

Specific research efforts in this area include skill extraction from job postings, a task crucial for identifying current workforce needs, emerging job roles, and skill shortages. Existing methods for skill extraction range from traditional rule-based and pattern-matching approaches to advanced machine learning and deep learning techniques, using both supervised and unsupervised methods [13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25].

Complementary research focuses on skill classification and matching, which aims to align extracted skills to other similar skills or to existing taxonomies such as ESCO [26]. This research area frequently explores innovative methodologies, such as leveraging self-supervised learning techniques, transformer-based language models, and semantic embeddings for skill representation and matching [27, 28, 29, 30, 17, 31, 32, 33, 34]. Similarly, job title classification and matching to, e.g., SOC [35] or ISCO [36] have

CLEF 2025 Working Notes, 9 – 12 September 2025, Madrid, Spain
*Corresponding author.
†These authors contributed equally.
jjz@cs.aau.dk (M. Zhang); robv@itu.dk (R. v. d. Goot)
https://jjzha.github.io/ (M. Zhang); https://robvanderg.github.io/ (R. v. d. Goot)
0000-0003-1218-5201 (M. Zhang); 0009-0003-1999-4156 (R. v. d. Goot)

Table 1
Example annotations for the job title "cider master" examples from the training data. For task B there are 330 skill matches, we only show a selection here.

Task A	Task B
master cider maker	analyse apple juice for cider production
cider production manager	analyse samples of food and beverages
cider manufacture manager	communicate with external laboratories
cider maker	display spirits
	...

been explored extensively, addressing challenges related to standardizing diverse job titles, which vary significantly across organizations and regions. Techniques include supervised classification, semantic matching algorithms, and transformer-based models such as JobBERT [37], which facilitate matching and categorization of job titles [38, 39, 40, 19, 41, 42]. These skills and job titles are also being further classified into their respective taxonomical counterparts to be used for further labor market demand analysis [43, 44].

Furthermore, research in career path prediction leverages historical job market data to anticipate future career trajectories, facilitating career counseling and workforce planning [45, 46]. Lastly, significant attention has been devoted to mapping jobs and skills onto standardized occupational and skill taxonomies, aiding systematic labor market research and policy-making decisions [47, 31, 48, 49, 50].

In this report, we investigate *multilingual job–title matching* (TalentCLEF Task A) and *job–skill matching* (Task B). Both tasks require mapping free-form labour-market text onto a structured knowledge base, yet they differ in language coverage, label granularity, and training-data volume. We show example annotations for both tasks in Table 1. We systematically compare three paradigms that dominate current NLP practice: (1) Classification, which treats ranking as binary relevance prediction and fine-tunes a multilingual encoder with cross-entropy loss. (2) Contrastive learning, which learns task-specific sentence embeddings via InfoNCE and relies on cosine ranking at inference. (3) Prompting, which directly exploits instruction-tuned LLM embedders in a zero-shot setting, requiring no task-specific updates.

We build all models on top of ESCO job titles and skills provided by the shared task organizers, for the contrastive models, we additionally exploit corresponding ESCO descriptions. Our experimental results show that contrastive learning excels on multilingual title matching, whereas discriminative fine-tuning leads on job–skill prediction. Surprisingly, zero-shot prompting with a 7B parameter model performs close to supervised systems on Task A, highlighting the rapid progress of instruction-tuned LLMs for Computational Job Market Analysis.

Contributions:

- We present the *first side-by-side comparison* of discriminative, contrastive, and prompting paradigms for both job–title and job–skill matching.
- We introduce a *unified ESCO-derived training corpus* (titles, alternative labels, and multilingual descriptions) and release preprocessing scripts to facilitate future research.[1]
- We achieve **5$^{\text{th}}$** place out of 20 teams on Task A and **3$^{\text{rd}}$** out of 14 on Task B by carefully tuning model size, negative-sampling strategy, and inference prompts, demonstrating that *model choice should be task-specific* rather than one-size-fits-all.

[1]Code is released at: https://github.com/jjzha/talentclef-nlpnorth

2. The Tasks

2.1. TalentCLEF Task A

This task challenges participants to build multilingual systems that, given a job title, rank related titles from a knowledge base [51]. Provided resources include:

- **Training Data**: 15,000 labeled pairs of related job titles per language (English, Spanish, German) support cross-lingual training.
- **Development Data**: Participants receive 100 manually annotated samples per language, each with a query job title and related titles. A language-specific knowledge base is also provided for ranking. This data also includes Chinese.
- **Test Data**: Participants receive 5,000 job titles, with evaluation based on a gold-standard subset of 100 titles per language. Titles include hidden annotations for industry and gender, allowing bias assessment alongside ranking performance. This data also includes Chinese.

2.2. TalentCLEF Task B

This task challenges participants to build models that, given a job title, return related skills from a knowledge base. All data is in English.

- **Training Data**: 2,000 job titles with associated professional skills, sourced from real job descriptions and semi-automatically curated for accuracy.
- **Development Data**: 200 job titles with related skills normalized to ESCO terminology. A skills knowledge base is also provided.
- **Test Data**: 500 job titles for which participants must predict related skills using the provided knowledge base.

3. Methodology

We compare three main categories of approaches to tackle the tasks. Below, we will outline each approach, and which variations within the approach we evaluated. For each approach, we first describe the setup for job title matching, and in the final paragraph describe how the model is adapted to perform skill matching. We train all models on the concatenation of the data for all languages.

3.1. Classification

Training Data. We reformulate the ranking task as binary classification. In Task A every provided positive instance consists of a query title and one related title. We create negative instances by pairing the same query with randomly sampled, unrelated titles. We experiment with positive–to–negative ratios of 1:1, 1:2, and 1:5 and select the best ratio (1:2) on the English development data. At prediction time, we use the output of the softmax to create a ranking (as opposed to using the binary labels directly). For classification, we only use TalentCLEF's training data and no extra descriptions as in the contrastive learning approach (Subsection 3.2).

Training. We train with MaChAmp [52], which places a single linear layer on top of a multilingual encoder and optimizes cross-entropy loss. We fine-tune several multilingual models given their domain representativeness and high performance on other classification tasks: `escoxlm-r` [53], `m-e5-large-instruct` [54], `m-e5-large` [54], `mdeberta-v3-base` [55], and `par-m-mpnet-base-v2` [56].

Hyperparameters. We started from the default hyperparameters from MaChAmp 0.4.2, which are a learning rate of 0.0001, batch size of 32, 20 epochs, and a slanted-triangular learning rate scheduler [57]. In our initial runs, we found early convergence, so we experimented with a lower learning rate and a smaller amount of epochs. We empirically saw similar performance with only 3 epochs, but the lower learning rate was not beneficial. Hence, we used all default settings except the number of epochs.

Adapting to Task B. For Task B we keep the architecture unchanged and simply replace related titles with related skills. Negatives are sampled from the entire skill vocabulary, the optimal negative ratio for taskB was 1:1.

3.2. Contrastive Learning

Training data. We construct pairs from ESCO. For each occupation we create (i) preferred title → description and (ii) preferred title → alternative title. Negatives come from the remaining descriptions or titles.

Training. We further fine-tune sentence embedding models with InfoNCE [58]. Given a batch of N aligned pairs (u_i, v_i), we treat every other v_j ($j \neq i$) as a hard negative for u_i and vice versa:

$$\ell_i = -\log \frac{\exp(\mathrm{sim}(u_i, v_i))}{\sum_{j=1}^{N} \exp(\mathrm{sim}(u_i, v_j))} \tag{1}$$

We use cosine similarity for $\mathrm{sim}(\cdot)$. The overall loss is the average of the ℓ_i.

Hyperparameters. We systematically vary the number of in-batch negatives $k \in \{1, 2, 5, 10, 15, 16, 20, 32\}$, batch size $\in \{16, 32, 64\}$, and learning rate $\in \{1 \times 10^{-4}, 5 \times 10^{-5}, 2 \times 10^{-5}, 2 \times 10^{-6}\}$. The optimal configuration uses $k = 16$, batch size 32, and learning rate 2×10^{-6}.

Adapting to Task B. We build three pair types: (i) job → skill, (ii) job → alt_skill, and (iii) alt_job → skill. We sample negatives from the complementary pool (all skills or all jobs, respectively).

3.3. Prompting

Finally, we test instruction-tuned LLM embedders in a zero-shot setting. Specifically, we embed text with `multilingual-e5-large-instruct` (560M), `Linq-Embed-Mistral` (7.11B) and `gte-Qwen2-7B-instruct` (7.61B). In this setup, we embed the query with a task description prefix, and we embed the candidate jobs/skills separately. Then, we rank candidates by cosine similarity. We use the following prefixes:

Task A: "`Given a job title, find the most relevant job titles.`"

Task B: "`Given a job title, find the most relevant skills.`"

4. Results

4.1. Task A: Multilingual Job Title Matching

For task A, the contrastive models achieves the highest performance on the validation data, and prompting on the test data, but their results are quite close on bath datasplits. It should be noted that the prompting model is 14 times larger, but did not require fine-tuning. The classification based models perform worse. The trends across different language models are consistent, larger models perform better, and m-e5-large provides a strong performance across approaches (except for prompting, due

[2]We only obtained these results after the deadline, hence we submitted the m-e5-large-instruct model for the test data.

Table 2
Results Task A. MAP scores for classification, similarity, and prompting-based methods. For the test predictions, we select the best performing model from each category.

		Validation					Test				
Classification (fine-tuned)											
	#P	en	de	es	zh	**avg**	en	de	es	zh	**avg**
TalentCLEF Baseline	?M	0.499	0.377	0.284	0.437	0.399	0.408	0.348	0.324	0.380	0.365
escoxlm-r	561M	0.550	0.376	0.421	0.505	0.463					
mdeberta-v3-base	276M	0.550	0.414	0.429	0.505	0.475[2]					
m-e5-large-instruct	560M	0.555	0.401	0.421	0.512	**0.472**	0.490	0.421	0.406	0.444	0.440
m-e5-large	560M	0.548	0.369	0.416	0.500	0.458					
par-m-mpnet-base-v2	278M	0.530	0.312	0.418	0.496	0.439					
Contrastive (fine-tuned)											
	#P	en	de	es	zh	**avg**	en	de	es	zh	**avg**
escoxlm-r	561M	0.578	0.405	0.473	0.516	0.493					
m-e5-large	560M	0.612	0.446	0.481	0.539	**0.520**	0.511	0.446	0.484	0.477	0.480
par-m-mpnet-base-v2	278M	0.549	0.368	0.448	0.480	0.461					
Prompting (zero-shot)											
	#P	en	de	es	zh	**avg**	en	de	es	zh	**avg**
gte-Qwen2-7B-instruct	7.61B	0.624	0.408	0.473	0.565	**0.518**	0.537	0.496	0.442	0.495	**0.493**
Linq-Embed-Mistral	7.11B	0.583	0.352	0.431	0.512	0.470	0.489	0.467	0.466	0.487	0.477
m-e5-large-instruct	560M	0.532	0.368	0.422	0.535	0.464					

Table 3
Cross-lingual Task A Test Results. We report mean-average precision (MAP) for each run on all fully in-language pairs (en–en, es–es, de–de, zh–zh) and on the three cross-language pairs that involve English (en–es, en–de, en–zh). The macro average across English, Spanish, and German is given in the second column.

	Avg. (en,es,de)	In-language				Cross-lingual		
		en–en	es–es	de–de	zh–zh	en–es	en–de	en–zh
TalentCLEF Baseline	0.340	0.408	0.324	0.348	0.380	0.335	0.345	
gte-Qwen2-7B-instruct	**0.493**	**0.537**	**0.496**	0.442	**0.495**	**0.492**	0.461	**0.494**
Linq-Embed-Mistral	0.477	0.489	0.467	0.466	0.487	0.455	0.468	0.454
m-e5-large (contrastive)	0.480	0.511	0.446	**0.484**	0.477	0.434	**0.469**	0.480

to its size). Performance on the languages also shows a highly similar trend. English gets the best performance, followed by Chinese, Spanish, and German. This is somewhat surprising, as Chinese is more distant to the other languages, and also completely unseen during training. We also note that the performance on the test data is lower compared to the validation data for all models. This could be a sign of overfitting, but after inspecting the data, we also saw that there is a larger overlap of job titles (~20% versus ~0%) with the train data. We hope details about the data creation process can shed more light on these differences.

Table 3 breaks the scores down by language pair for the prompt-based models (which are the only one we submitted for the crosslingual track). `gte-Qwen2-7B-instruct` achieves the strongest English–Spanish (0.492) and English–Chinese (0.494) transfer, and ties for the best English–German performance (0.461). In-language results show the same trend: It tops English (0.537) and Spanish (0.496), while remaining competitive on German (0.442). These findings suggest that (i) prompting benefits most from the larger pre-training signal in English and Spanish ESCO descriptions, and (ii) cosine-similarity scoring is robust across both monolingual and cross-lingual retrieval scenarios.

Table 4
Results Task B. We show three methods for prediction; classification, similarity, and prompting-based methods. For the test predictions, we select the best performing model from each category.

		Valid.	Test
Classification (fine-tuned)			
	#P	en	en
TalentCLEF	?	0.187	0.196
escoxlm-r	561M	0.255	
mdeberta-v3-base	276M	**0.267**	**0.290**
m-e5-large-instruct	560M	0.266	
m-e5-large	560M	0.264	
par-m-mpnet-base-v2	278M	0.247	

		Valid.	Test
Contrastive (fine-tuned)			
	#P	en	en
escoxlm-r	561M	0.204	
m-e5-large	560M	0.214	
par-m-mpnet-base-v2	278M	**0.250**	0.253
Prompting (zero-shot)			
	#P	en	en
gte-Qwen2-7B-instruct	7.61B	**0.222**	0.283
Linq-Embed-Mistral	7.11B	0.209	
m-e5-large-instruct	560M	0.178	

Table 5
Task A: Corpus mapping coverage by language in the validation set.

	English	Spanish	German
Total corpus titles	2619	4661	4729
Mapped titles	591	569	585
Unmapped titles (%)	77.4%	87.8%	87.6%

4.2. TaskB: Job Title-Based Skill Prediction

For task B (Table 4) the performance of the models is swapped. The classification based models outperform the other models on both the validation and the test data. Overall, the MAP scores are also much lower compared to task A, showing that the mapping of skills to jobs is a more challenging task. Therefore, we hypothesize that the direct supervised training signal is the key to the higher performance. Interestingly, models size has a less clear impact compared to task A, for both the contrastive and the classification models a the smallest language model performs best.

5. Analysis

Per-category Performance. For analysis, we investigate in which ESCO job title major group the models perform best for Task A. We take the three best-performing models in each method category (i.e., classification, contrastive, prompting) from Task A and map each data point from the validation set to their respective ESCO major group. In Table 5, we report the amount of job titles which can be mapped to a specific ESCO code. For English, there are 77.4% unmapped titles and for both Spanish and German this is around 87%.

We report the results on the mapped job titles in Table 6. We observe that for all models, job titles in categories such as “managers”, “professionals”, “technicians and associate professionals”, “clerical support workers”, and “service and sales workers” are the most difficult to predict. In contrast, we see job titles from categories such as “armed forces occupations”, “craft and related trades”, “plant and machine operators” and “elementary occupations” being often predicted correctly. In the case of unmapped job titles, we see that most models do not perform well.

Table 6
Mean average precision score by ESCO major group across languages and model variants.

	English			Spanish			German		
ESCO major group	*clas. E5-inst*	*contr. E5-con*	*prompt GTE*	*clas. E5-inst*	*contr. E5-con*	*prompt GTE*	*clas. E5-inst*	*contr. E5-con*	*prompt GTE*
Unmapped Titles	0.553	0.551	0.617	0.433	0.478	0.477	0.431	0.462	0.426
0 Armed forces occupations	1.000	1.000	0.958	1.000	1.000	0.833	0.861	1.000	0.667
1 Managers	0.788	0.840	0.823	0.553	0.595	0.633	0.582	0.669	0.591
2 Professionals	0.647	0.668	0.702	0.569	0.585	0.639	0.545	0.573	0.575
3 Technicians & assoc. profs.	0.703	0.693	0.775	0.574	0.621	0.675	0.611	0.659	0.557
4 Clerical support workers	0.774	0.857	0.729	0.537	0.581	0.533	0.468	0.542	0.374
5 Service & sales workers	0.764	0.728	0.729	0.772	0.961	0.898	0.698	0.707	0.531
6 Skilled agri./forestry/fishery workers	—	—	—	—	—	—	—	—	—
7 Craft & related trades	0.917	1.000	0.889	0.678	0.747	0.700	0.765	0.697	0.869
8 Plant & machine operators	1.000	0.833	1.000	0.875	0.479	0.875	1.000	1.000	1.000
9 Elementary occupations	0.822	1.000	0.989	0.441	0.793	0.552	0.770	0.815	0.433

6. Conclusion

In this paper, we report our methods for the 2025 TalentCLEF shared task. We demonstrate that prompting is effective for multilingual job title matching (Task A) and classification approaches excel in predicting job-related skills (Task B). However for task B, prompting-based methods, despite their lower performance, show promising results in a zero-shot scenario, suggesting potential avenues for further exploration.

Acknowledgments

MZ is supported by the research grant (VIL57392) from VILLUM FONDEN.

Declaration on Generative AI

During the preparation of this work, the author(s) used GPT-4o in order to: Check grammar and spelling check. After using these tool(s)/service(s), the author(s) reviewed and edited the content as needed and take(s) full responsibility for the publication's content.

References

[1] *H. Lasi, P. Fettke, H.-G. Kemper, T. Feld, M. Hoffmann, Industry 4.0, Business & information systems engineering 6 (2014) 239–242.*

[2] *K. Schwab, The fourth industrial revolution, Currency, 2017.*

[3] *European Commission, Industry 5.0: Towards more sustainable, resilient and human-centric industry, https://research-and-innovation.ec.europa.eu/news/all-research-and-innovation-news/industry-50-towards-more-sustainable-resilient-and-human-centric-industry-2021-01-07_en, 2021. Accessed: 2023-10-27.*

[4] *T. Eloundou, S. Manning, P. Mishkin, D. Rock, Gpts are gpts: An early look at the labor market impact potential of large language models, ArXiv preprint abs/2303.10130 (2023). URL: https://arxiv.org/abs/2303.10130.*

[5] *D. H. Autor, D. Dorn, The growth of low-skill service jobs and the polarization of the us labor market, American economic review 103 (2013) 1553–1597. URL: https://www.aeaweb.org/articles?id=10.1257/aer.103.5.1553.*

[6] *D. H. Autor, F. Levy, R. J. Murnane, The skill content of recent technological change: An empirical exploration, The Quarterly journal of economics 118 (2003) 1279–1333. URL: https://academic.oup.com/qje/article-abstract/118/4/1279/1925105?login=false.*

[7] E. Brynjolfsson, A. McAfee, Race against the machine: How the digital revolution is accelerating innovation, driving productivity, and irreversibly transforming employment and the economy, Brynjolfsson and McAfee, 2011.

[8] E. Brynjolfsson, A. McAfee, The second machine age: Work, progress, and prosperity in a time of brilliant technologies, WW Norton & Company, 2014.

[9] K. Balog, Y. Fang, M. De Rijke, P. Serdyukov, L. Si, et al., Expertise retrieval, Foundations and Trends® in Information Retrieval 6 (2012) 127–256.

[10] N. Otani, N. Bhutani, E. Hruschka, Natural language processing for human resources: A survey, ArXiv preprint abs/2410.16498 (2024). URL: https://arxiv.org/abs/2410.16498.

[11] E. Senger, M. Zhang, R. van der Goot, B. Plank, Deep learning-based computational job market analysis: A survey on skill extraction and classification from job postings, in: E. Hruschka, T. Lake, N. Otani, T. Mitchell (Eds.), Proceedings of the First Workshop on Natural Language Processing for Human Resources (NLP4HR 2024), Association for Computational Linguistics, St. Julian's, Malta, 2024, pp. 1–15. URL: https://aclanthology.org/2024.nlp4hr-1.1/.

[12] M. Zhang, Computational Job Market Analysis: with Natural Language Processing, 2024.

[13] L. Sayfullina, E. Malmi, J. Kannala, Learning representations for soft skill matching, in: International Conference on Analysis of Images, Social Networks and Texts, 2018, pp. 141–152.

[14] A. Bhola, K. Halder, A. Prasad, M.-Y. Kan, Retrieving skills from job descriptions: A language model based extreme multi-label classification framework, in: D. Scott, N. Bel, C. Zong (Eds.), Proceedings of the 28th International Conference on Computational Linguistics, International Committee on Computational Linguistics, Barcelona, Spain (Online), 2020, pp. 5832–5842. URL: https://aclanthology.org/2020.coling-main.513/. doi:10.18653/v1/2020.coling-main.513.

[15] I. Khaouja, G. Mezzour, I. Kassou, Unsupervised skill identification from job ads, in: 2021 IEEE 22nd International Conference on Information Reuse and Integration for Data Science (IRI), IEEE, 2021, pp. 147–151.

[16] M. Zhang, K. Jensen, S. Sonniks, B. Plank, SkillSpan: Hard and soft skill extraction from English job postings, in: M. Carpuat, M.-C. de Marneffe, I. V. Meza Ruiz (Eds.), Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Association for Computational Linguistics, Seattle, United States, 2022, pp. 4962–4984. URL: https://aclanthology.org/2022.naacl-main.366/. doi:10.18653/v1/2022.naacl-main.366.

[17] M. Zhang, K. N. Jensen, B. Plank, Kompetencer: Fine-grained skill classification in danish job postings via distant supervision and transfer learning, in: Proceedings of the Language Resources and Evaluation Conference, European Language Resources Association, Marseille, France, 2022, pp. 436–447. URL: https://aclanthology.org/2022.lrec-1.46.

[18] M. Zhang, K. N. Jensen, R. van der Goot, B. Plank, Skill extraction from job postings using weak supervision, in: Proceedings of RecSysHR'22, RecSysHR'22, 2022.

[19] T. Green, D. Maynard, C. Lin, Development of a benchmark corpus to support entity recognition in job descriptions, in: Proceedings of the Language Resources and Evaluation Conference, European Language Resources Association, Marseille, France, 2022, pp. 1201–1208. URL: https://aclanthology.org/2022.lrec-1.128.

[20] A.-S. Gnehm, E. Bühlmann, S. Clematide, Evaluation of transfer learning and domain adaptation for analyzing german-speaking job advertisements, in: Proceedings of the Language Resources and Evaluation Conference, European Language Resources Association, Marseille, France, 2022, pp. 3892–3901. URL: https://aclanthology.org/2022.lrec-1.414.

[21] K. Nguyen, M. Zhang, S. Montariol, A. Bosselut, Rethinking skill extraction in the job market domain using large language models, in: Proceedings of the First Workshop on Natural Language Processing for Human Resources (NLP4HR 2024), Association for Computational Linguistics, St. Julian's, Malta, 2024, pp. 27–42. URL: https://aclanthology.org/2024.nlp4hr-1.3.

[22] A. Herandi, Y. Li, Z. Liu, X. Hu, X. Cai, Skill-llm: Repurposing general-purpose llms for skill extraction, ArXiv preprint abs/2410.12052 (2024). URL: https://arxiv.org/abs/2410.12052.

[23] L. Vásquez-Rodríguez, S. M. Bertrand Audrin, S. Galli, J. Rogenhofer, J. N. Cusa, L. van der Plas, Hardware-effective approaches for skill extraction in job offers and resumes (2024).

[24] M. Zhang, R. v. d. Goot, M.-Y. Kan, B. Plank, NNOSE: Nearest neighbor occupational skill extraction, in: Proceedings of the 18th Conference of the European Chapter of the Association for Computational Linguistics (Volume 1: Long Papers), Association for Computational Linguistics, St. Julian's, Malta, 2024, pp. 589–608. URL: https://aclanthology.org/2024.eacl-long.35.

[25] H. Kavas, M. Serra-Vidal, L. Wanner, Multilingual skill extraction for job vacancy–job seeker matching in knowledge graphs, in: Proceedings of the Workshop on Generative AI and Knowledge Graphs (GenAIK), 2025, pp. 146–155.

[26] M. le Vrang, A. Papantoniou, E. Pauwels, P. Fannes, D. Vandensteen, J. De Smedt, Esco: Boosting job matching in europe with semantic interoperability, Computer 47 (2014) 57–64.

[27] A. Giabelli, L. Malandri, F. Mercorio, M. Mezzanzanica, Graphlmi: A data driven system for exploring labor market information through graph databases, Multimedia Tools and Applications (2020) 1–30.

[28] M. de Groot, J. Schutte, D. Graus, Job posting-enriched knowledge graph for skills-based matching, 2021. arXiv:2109.02554.

[29] R. Yazdanian, R. L. Davis, X. Guo, F. Lim, P. Dillenbourg, M.-Y. Kan, On the radar: Predicting near-future surges in skills' hiring demand to provide early warning to educators, Computers and Education: Artificial Intelligence (2021) 100043.

[30] J.-J. Decorte, J. Van Hautte, J. Deleu, C. Develder, Demeester, Design of negative sampling strategies for distantly supervised skill extraction, ArXiv preprint abs/2209.05987 (2022). URL: https://arxiv.org/abs/2209.05987.

[31] B. Clavié, G. Soulié, Large language models as batteries-included zero-shot esco skills matchers, ArXiv preprint abs/2307.03539 (2023). URL: https://arxiv.org/abs/2307.03539.

[32] J.-J. Decorte, J. V. Hautte, T. Demeester, C. Develder, Skillmatch: Evaluating self-supervised learning of skill relatedness, ArXiv preprint abs/2410.05006 (2024). URL: https://arxiv.org/abs/2410.05006.

[33] A. Magron, A. Dai, M. Zhang, S. Montariol, A. Bosselut, JobSkape: A framework for generating synthetic job postings to enhance skill matching, in: E. Hruschka, T. Lake, N. Otani, T. Mitchell (Eds.), Proceedings of the First Workshop on Natural Language Processing for Human Resources (NLP4HR 2024), Association for Computational Linguistics, St. Julian's, Malta, 2024, pp. 43–58. URL: https://aclanthology.org/2024.nlp4hr-1.4/.

[34] M. Gavrilescu, F. Leon, A.-A. Minea, Techniques for transversal skill classification and relevant keyword extraction from job advertisements, Information 16 (2025) 167.

[35] P. Elias, M. Birch, et al., Soc2010: revision of the standard occupational classification, Economic & Labour Market Review 4 (2010) 48–55.

[36] P. Elias, Occupational Classification (ISCO-88): Concepts, Methods, Reliability, Validity and Cross-National Comparability, Technical Report, OECD Publishing, 1997.

[37] J.-J. Decorte, J. Van Hautte, T. Demeester, C. Develder, Jobbert: Understanding job titles through skills, ArXiv preprint abs/2109.09605 (2021). URL: https://arxiv.org/abs/2109.09605.

[38] F. Javed, Q. Luo, M. McNair, F. Jacob, M. Zhao, T. S. Kang, Carotene: A job title classification system for the online recruitment domain, in: 2015 IEEE First International Conference on Big Data Computing Service and Applications, IEEE, 2015, pp. 286–293.

[39] F. Javed, M. McNair, F. Jacob, M. Zhao, Towards a job title classification system, ArXiv preprint abs/1606.00917 (2016). URL: https://arxiv.org/abs/1606.00917.

[40] F. Javed, P. Hoang, T. Mahoney, M. McNair, Large-scale occupational skills normalization for online recruitment, in: S. P. Singh, S. Markovitch (Eds.), Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence, February 4-9, 2017, San Francisco, California, USA, AAAI Press, 2017, pp. 4627–4634. URL: http://aaai.org/ocs/index.php/IAAI/IAAI17/paper/view/14922.

[41] F. Retyk, L. Gasco, C. P. Carrino, D. Deniz, R. Zbib, Melo: An evaluation benchmark for multilingual entity linking of occupations, ArXiv preprint abs/2410.08319 (2024). URL: https://arxiv.org/abs/2410.08319.

[42] X. Liu, Y. Wang, Q. Dong, X. Lu, Job title prediction as a dual task of expertise prediction in open source software, in: Joint European Conference on Machine Learning and Knowledge Discovery in Databases, Springer, 2024, pp. 381–396.

[43] E. Malherbe, M. Aufaure, Bridge the terminology gap between recruiters and candidates: A multilingual

skills base built from social media and linked data, in: R. Kumar, J. Caverlee, H. Tong (Eds.), 2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, ASONAM 2016, San Francisco, CA, USA, August 18-21, 2016, IEEE Computer Society, 2016, pp. 583–590. URL: https://doi.org/10.1109/ASONAM.2016.7752295. doi:10.1109/ASONAM.2016.7752295.

[44] E. M. Sibarani, S. Scerri, C. Morales, S. Auer, D. Collarana, Ontology-guided job market demand analysis: a cross-sectional study for the data science field, in: Proceedings of the 13th International Conference on Semantic Systems, 2017, pp. 25–32.

[45] J.-J. Decorte, J. Van Hautte, J. Deleu, C. Develder, T. Demeester, Career path prediction using resume representation learning and skill-based matching, in: RecSys in HR2023: the 3rd Workshop on Recommender Systems for Human Resources, in conjunction with the 17th ACM Conference on Recommender Systems, volume 3490, CEUR, 2023.

[46] E. Senger, Y. Campbell, R. van der Goot, B. Plank, KARRIEREWEGE: A large scale career path prediction dataset, in: O. Rambow, L. Wanner, M. Apidianaki, H. Al-Khalifa, B. D. Eugenio, S. Schockaert, K. Darwish, A. Agarwal (Eds.), Proceedings of the 31st International Conference on Computational Linguistics: Industry Track, Association for Computational Linguistics, Abu Dhabi, UAE, 2025, pp. 533–545. URL: https://aclanthology.org/2025.coling-industry.46/.

[47] ESCO, Machine Learning Assisted Mapping of Multilingual Occupational Data to ESCO (Part 1), 2022. URL: https://esco.ec.europa.eu/en/about-esco/data-science-and-esco/machine-learning-assisted-mapping-multilingual-occupational-data-esco-part-1.

[48] J. Vrolijk, D. Graus, Enhancing plm performance on labour market tasks via instruction-based finetuning and prompt-tuning with rules, ArXiv preprint abs/2308.16770 (2023). URL: https://arxiv.org/abs/2308.16770.

[49] M. Zhang, R. v. d. Goot, B. Plank, Entity linking in the job market domain, in: Findings of the Association for Computational Linguistics: EACL 2024, Association for Computational Linguistics, St. Julian's, Malta, 2024, pp. 410–419. URL: https://aclanthology.org/2024.findings-eacl.28.

[50] J. Rosenberger, L. Wolfrum, S. Weinzierl, M. Kraus, P. Zschech, Careerbert: Matching resumes to esco jobs in a shared embedding space for generic job recommendations, Expert Systems with Applications 275 (2025) 127043.

[51] L. Gasco, H. Fabregat, L. García-Sardiña, P. Estrella, D. Deniz, A. Rodrigo, R. Zbib, Overview of the TalentCLEF 2025 Shared Task: Skill and Job Title Intelligence for Human Capital Management, in: International Conference of the Cross-Language Evaluation Forum for European Languages, Springer, 2025.

[52] R. van der Goot, A. Üstün, A. Ramponi, I. Sharaf, B. Plank, Massive choice, ample tasks (MaChAmp): A toolkit for multi-task learning in NLP, in: D. Gkatzia, D. Seddah (Eds.), Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: System Demonstrations, Association for Computational Linguistics, Online, 2021, pp. 176–197. URL: https://aclanthology.org/2021.eacl-demos.22/. doi:10.18653/v1/2021.eacl-demos.22.

[53] M. Zhang, R. van der Goot, B. Plank, ESCOXLM-R: Multilingual Taxonomy-driven Pre-training for the Job Market Domain, in: Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Association for Computational Linguistics, Toronto, Canada, 2023, pp. 11871–11890. URL: https://aclanthology.org/2023.acl-long.662. doi:10.18653/v1/2023.acl-long.662.

[54] Z. Zhang, Y. Gao, J.-G. Lou, e^5: Zero-shot hierarchical table analysis using augmented LLMs via explain, extract, execute, exhibit and extrapolate, in: K. Duh, H. Gomez, S. Bethard (Eds.), Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers), Association for Computational Linguistics, Mexico City, Mexico, 2024, pp. 1244–1258. URL: https://aclanthology.org/2024.naacl-long.68/. doi:10.18653/v1/2024.naacl-long.68.

[55] P. He, J. Gao, W. Chen, Debertav3: Improving deberta using electra-style pre-training with gradient-disentangled embedding sharing, 2021. arXiv:2111.09543.

[56] N. Reimers, I. Gurevych, Making monolingual sentence embeddings multilingual using knowledge distillation, in: Proceedings of the 2020 Conference on Empirical Methods in Natural Language

Processing (EMNLP), Association for Computational Linguistics, Online, 2020, pp. 4512–4525. URL: https://aclanthology.org/2020.emnlp-main.365. doi:10.18653/v1/2020.emnlp-main.365.

[57] J. Howard, S. Ruder, Universal language model fine-tuning for text classification, in: I. Gurevych, Y. Miyao (Eds.), Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Association for Computational Linguistics, Melbourne, Australia, 2018, pp. 328–339. URL: https://aclanthology.org/P18-1031/. doi:10.18653/v1/P18-1031.

[58] A. v. d. Oord, Y. Li, O. Vinyals, Representation learning with contrastive predictive coding, ArXiv preprint abs/1807.03748 (2018). URL: https://arxiv.org/abs/1807.03748.