Ontology-Aligned Embeddings for Data-Driven Labour Market Analytics

Heinke Hihn[1,2,*], Dennis A. V. Dittrich[2], Carl Jeske[3], Cayo Costa Sobral[3], Helio Pais[4] and Timm Lochmann[4]

[1] *IU International University of Applied Sciences, Department of Computer Science and Engineering, Berlin, Germany*
[2] *The Stepstone Group, AI Labs, Berlin, Germany*
[3] *The Stepstone Group, AI Labs, London, UK*
[4] *The Stepstone Group, AI Labs, Düsseldorf, Germany*

Abstract

The limited ability to reason across occupational data from different sources is a long-standing bottleneck for data-driven labour market analytics. Previous research has relied on hand-crafted ontologies that allow such reasoning but are computationally expensive and require careful maintenance by human experts. The rise of language processing machine learning models offers a scalable alternative by learning shared semantic spaces that bridge diverse occupational vocabularies without extensive human curation. We present an embedding-based alignment process that links any free-form German job title to two established ontologies - the German *Klassifikation der Berufe* and the *International Standard Classification of Education.* Using publicly available data from the German Federal Employment Agency, we construct a dataset to fine-tune a Sentence-BERT model to learn the structure imposed by the ontologies. The enriched pairs (job title, embedding) define a similarity graph structure that we can use for efficient approximate nearest-neighbour search, allowing us to frame the classification process as a semantic search problem. This allows for greater flexibility, e.g., adding more classes. We discuss design decisions, open challenges, and outline ongoing work on extending the graph with other ontologies and multilingual titles.

Keywords

Embedding Model, SentenceTransformers, Semantic Search, Contrastive Learning, Approximate k-Nearest-Neighbour Search, Labour Economics, KldB, ISCED

1. Introduction

Occupational data is one of the most versatile bits of information about a subject in quantitative data and provides a wide range of analytical use cases, such as socioeconomic indices, measures of workplace tasks, occupation-specific health risks, gender segregation, and occupational closure [1]. Naturally, knowledge-intensive organisations in the employment sector aim to take advantage of this rich information, raising the need for systematic and internationally comparable ontologies.

In the German labour market, job roles are codified by the *Klassifikation der Berufe* (KldB 2010) [2, 3], while educational attainments follow *Deutscher Qualifikationsrahmen für lebenslanges Lernen* (DQR) [4, 5]. The DQR is based on the *European Qualification Framework* [6] and is highly aligned with UNESCO's *International Standard Classification of Education* (ISCED 2011) [7]. Yet, these ontologies reside in separate silos and are rarely connected at scale, forcing practitioners to fall back on ad-hoc keyword heuristics that neither generalise nor explain their recommendations.

Free-form job titles aggravate the problem as the same occupation may appear in different forms, e.g. "Software Engineer" or "Software Developer". Recent studies have shown that dense vector representations help cluster these variants and predict plausible career paths, but do not align the results to standard taxonomies suitable for downstream reasoning [8, 9].

Workshop SIG Knowledge Management (FG WM) at KI2025, Potsdam, Germany

*Corresponding author.
heinke.hihn@iu.org (H. Hihn); dennisalexivalin.dittrich@stepstone.com (D. A. V. Dittrich); carl.jeske@stepstone.com (C. Jeske); cayo.costa-sobral@stepstone.com (C. Costa Sobral); helio.pais@stepstone.com (H. Pais); timm.lochmann@stepstone.com (T. Lochmann)
0000-0002-3244-3661 (H. Hihn); 0000-0002-4438-8276 (D. A. V. Dittrich); 0000-0002-3899-4894 (C. Costa Sobral); 0000-0003-2352-298X (H. Pais)

Table 1
Hierarchical structure of occupational classification of the KldB 2010 codes.

Level	Example Denomination
10 Occupational Areas	3: Occupations in construction, architecture, surveying and technical building services
37 Main groups (2-digit)	32: Occupations in building construction above and below ground
144 Groups (3-digit)	322: Civil engineering
702 Subgroups (4-digit)	3225: Canal and tunnel construction
1300 Types (5-digit)	32253: Canal and tunnel construction – complex specialist tasks
Specific job titles	32253-100 Canal Construction Foreman (Kanalmeister/in)

This paper closes that gap by introducing a lightweight method that

- fine-tunes Sentence-BERT (SBERT) [10] to predict 5-digit KldB 2010 codes with sub-second latency
- infers corresponding ISCED 2011 ranges via rule-based heuristics, linking occupational and educational hierarchies
- exposes the result as a lightweight graph for fast inference

The remainder of the paper reviews related work (Section 2), details data and methodology (Section 3), presents empirical results (Section 4), and concludes with a discussion and future research directions (Section 5).

2. Related Work

Classification of job titles into various occupational and educational taxonomies has been a long-standing challenge in labour market analytics. In the following, we summarise the state of the art, highlighting key contributions and gaps that our work addresses.

Large Language Models (LLMs) LLMs are increasingly combined with structured graphs to offer both neural adaptability and symbolic transparency [11, 12]. However, they come with high computing costs, but the need for efficient methods is pressing: The Stepstone Group's job board hosts more than 600,000 open vacancies per year [13], each with a free-form job title that must be classified into various occupational and educational taxonomies. This requires a lightweight, fast, and explainable solution that can be integrated into real-time recommender systems [14], such as pay-band calibration [15], labour force demand and supply analyses, job title normalisation [16], or skill-gap detection [17].

Automated occupation coding. Early attempts to normalise free-text job titles for statistics relied on rule bases or TF-IDF similarity. Transformer models now dominate the task. Baskaran and Müller fine-tuned a Sentence-BERT variant on vacancy titles and achieved 0.86 accuracy at the 5-digit KldB 2010 level, Safikhani et al. compared BERT and GPT-3 on survey responses and showed that hierarchical loss functions boost KldB coding by 15 pp over traditional baselines.

Embedding-enriched representations. JobBERT introduced by Decorte et al. augments BERT with co-occurring ESCO skill labels to learn dense job-title embeddings that outperform generic sentence encoders on ESCO normalisation.

Skills and occupation knowledge graphs. Beyond classification, several authors develop labour market knowledge graphs. de Groot et al. fuse job–skill relations from postings with ESCO/O*NET to drive skills-based matching and career-path search. Seif et al. add temporal layers to capture how demand evolves, stressing graph update mechanisms. Neither graph incorporates formal education codes or German-specific taxonomies.

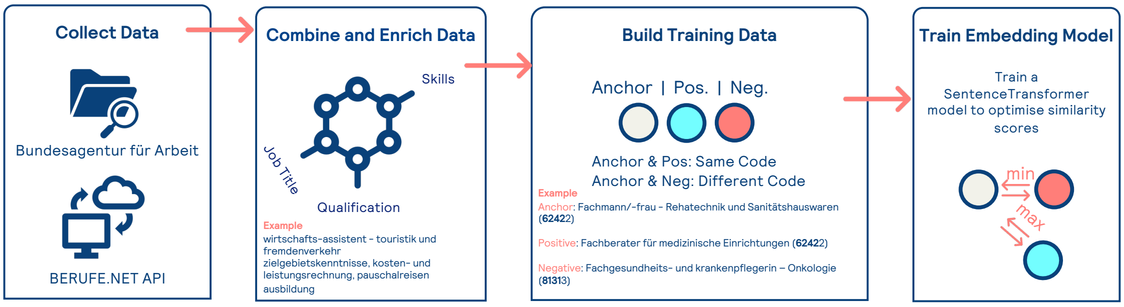

Figure 1: We collect publicly available data from the *Bundesagentur für Arbeit*, enrich it, build a training data set, and fine-tune an embedding model. Anchor-Positive-Negative triplets are determined by matching KldB 2010 codes. In this example, we link to the anchor with KldB 62422 another job title from the KldB subgroup 6262 ("sales occupations (retail) selling medical supplies and healthcare goods") as the positive and with a job title from KldB subgroup 8318 ("Occupations in nursing specialised in a particular branch of nursing") as the negative sample.

Linking occupations to education. Seif et al. proposed large-scale projections of qualification supply and already use the fifth KldB 2010 digit (requirement level) as a proxy for educational attainment, but do not have an explicit ISCED 2011 mapping. Our approach bridges between KldB 2010 levels and ISCED ranges to address this gap.

Positioning of this work. The state-of-the-art covers (i) KldB 2010 or ISCO classification, (ii) skill-centric knowledge graphs, and (iii) exploratory LLM-graph integrations, each in isolation. We propose a method that combines occupational and educational codes, and exposes the result as a queryable graph fit for downstream LLM applications, such as pay-band calibration or skill-gap detection. By integrating occupational, educational, and managerial semantics in a single embedding space, our work extends prior efforts and addresses an unmet need in knowledge-based workforce analytics.

3. Data and Methodology

This section details the resources, preprocessing steps, and modelling choices that enable the alignment of free-form German job titles with KldB 2010 and ISCED 2011.

3.1. The German Classification of Occupations: KldB

The *Klassifikation der Berufe* (KldB) is a five-digit code that uniquely identifies occupations based on tasks, skills, industry, and complexity developed and maintained by the German Federal Employment Agency (*Bundesagentur für Arbeit (BAfA)*). Each digit defines a level: areas, main groups, groups, sub groups, and types. The last digit serves as an indicator for the complexity of the job: 1 identifies helper jobs that require no vocational training, 2 is for specialist activities, 3 are complex specialist activities, and 4 are highly complex activities. Table 1 gives an example. The fourth digit is an indicator for management duties: for requirement levels 2 and 3 it indicates supervisors, and for requirement level 4 it indicates managers. The distinction helps to distinguish "blue-collar" and "white-collar" occupations.

We use an extended data set that contains not only official job titles, but also terms related to job titles. The data contains all indexed key words supported by the BAfA job site. In total, it contains 525.207 terms.

3.2. The International Standard Classification of Education: ISCED

The International Standard Classification of Education (ISCED), maintained by the UNESCO Institute for Statistics [7], is the globally recognised taxonomy for describing programmes and attainment

Table 2
ISCED 2011 Levels [30] define an encoding of educational attainments that makes different qualifications and requirements internationally comparable. Note, that codes 35_2 and 35_3 were added to distinguish between forms of vocational training and are not part of the original ISCED 2011 levels. Note, that we only use a subset of the levels, excluding 01, 02, 10, 44, and 45.

Level	Educational Level	Example occupation
01, 02	Pre-School	Not used in our approach
10	Elementary School	Not used in our approach
24	Lower secondary education	Demolition Worker
35_2	Two-year dual vocational training	Motor-Vehicle Service Assistant
35_3	Three-year dual vocational training	Automotive Mechatronics Technician – Vehicle Comm. Tech.
44, 45	Post-secondary non-tertiary education	Not used in our approach
55	Certified Vocational Specialist	Vocational Specialist – Foreign-Language Communication
64	Bachelor's degree or eq.	Software Engineer
65	Master Craftsperson or eq.	Master Craftsperson for Upholstery Technology
74	Master's degree or eq.	Data Scientist
75	Master Professional or eq.	Certified IT Technical Engineer
84	Doctorate (Ph.D.)	University Professor

across all education systems - see Table 2. First introduced in 1976 and revised most recently in 2011, ISCED organises learning opportunities into a hierarchical code scheme (levels 0 to 8) that captures the progressive complexity and specialisation of study, from early childhood to doctoral research. Because each code is defined by internationally agreed content-based criteria rather than by national nomenclature, ISCED enables robust cross-country comparison, longitudinal tracking of educational expansion, and seamless integration of qualification data into labour market and demographic statistics.

3.3. Training

To train the embedding model, we select a baseline SentenceTransformer tuned on German from HuggingFace's sentence similarity model repository [23] by evaluating their performance on a fraction of the data. The training data set is constructed in the contrastive learning setting of triplets [24]: Anchor - Positive - Negative. Positives and negatives are randomly assigned by KldB 2010 codes in relation to the anchor sample, where positives share a KldB 2010 code, and negatives have different codes. We create two data sets: one by the first four digits of the KldB 2019nd one by the last digit and then de-duplicate them. For each term in the base data file, we randomly select three positives and negatives, resulting in roughly 1.5M training triplets.

The model is trained on a single GPU making use of several memory saving techniques [25, 26] to push the batch size as high as possible, as this has been proven effective in contrastive learning and similar tasks [24, 27]. We use Multiple Negatives Ranking Loss [28] as a base loss and Matryoshka Loss [29] as a meta-loss to encourage the model to learn embeddings on different dimensions (64 to 1024). Figure 1 gives an overview of the training process.

3.3.1. Query Structure

To encourage the model to learn distinguishable parts of the embeddings, we define a query structure based on special separator tokens. The query structure is defined as:

```
query = "[JOB_TITLE_SEP] {job_title} [QUALIFICATION_SEP] {
    qualification} [SKILL_SEP] {skills}",
```

where the separator tokens are special tokens defined in the tokenizer and are thus not part of the vocabulary on which the model is trained.

Table 3
Mapping of original occupational qualification values to consolidated groups for simplified grouping during training.

Consolidated group	Original categories mapped to it
Weiterbildung	Andere Weiterbildungen; Kaufmännische Weiterbildungen; Tätigkeiten nach Weiterbildung; Weiterbildungen (bedingen Hochschulstudium)
Ausübungsformen	Ausübungsformen
Bamtenausbildung	Beamtenausbildungen; Tätigkeiten nach Beamtenausbildung
helfer-/anlerntätigkeiten	Helfer-/Anlerntätigkeiten
Meister	Meister
soldatenausbildung	Soldatenausbildungen; Tätigkeiten nach Soldatenausbildung
Ausbildung	Sonstige Ausbildungen; Tätigkeiten nach Ausbildung; Duale Ausbildung, Ausbildungen für Menschen mit Behinderungen; Berufsfachschulausbildungen (rechtlich geregelt)
Studium	Studienfächer/-gänge; Tätigkeiten nach Studium
Techniker	Techniker

Skills and qualification are extracted through the publicly available BERUFENET.API [31] which provides a RESTful interface to the BAfA's occupational database. An example of a set of skills is shown here (translated into English):

Job Title: Finance Assistant | **KldB**: 76911 | **Skills**:
international business; construction/mortgage financing; home-savings business; credit assessment; financial-services consulting; financial planning; costing; lending operations; customer consulting/support; marketing; retail banking; savings and investment services; tax law; insurance business; insurance law; contract processing/administration; securities business; field service/external sales; foreign-trade financing; banking and stock-exchange law; banking and capital-markets law; controlling; electronic banking; corporate banking; account management; cost and performance accounting; claims processing; sustainable investments; online banking; auditing/internal audit; clerical processing; contract law; payment transactions; competence group bank products; competence group office communication (MS-Office); competence group insurance lines

To simplify training, we use only a subset of the available qualifications, namely: Helper Jobs, Vocational Training, Additional Vocational Training, University Degree, Civil Servants, and Armed Forces Personnel. The remaining qualifications are assigned to one of them, e.g., "Study Programme" is assigned to University Degree - see Table 3 for the full mapping we use. In cases where the job title is associated with management duties, as indicated by the fourth digit being 9, we add "with management duties" to the qualification, for example, "university degree with management duties".

3.3.2. Enriching with ISCED

The BAfA provides high-level qualification levels, but does not map them to ISCED codes. Regulated occupations and academic degrees are classified within the German Qualifications Framework for Lifelong Learning [4, 5] into a set of DRQ levels that can be directly mapped to ISCED levels. We therefore define a set of rules to map the KldB codes to ISCED levels. The mapping is based on a combination of the requirement level, the qualification provided by the BAA, the job title, and, if it exists, its corresponding DQR level and in some cases the KldB code itself. For example, if the job title contains an education level such as "Bachelor of Science" or "staatlich geprüfter Betriebswirt", we map it to the corresponding ISCED levels 64 and 65 - see Table 4 for more examples of the mapping.

3.4. Inference

Inference is performed by a k-Nearest-Neighbour (k-NN) search on the embedding space. The k-NN search is performed using the HNSW library [32, 33] which allows for an efficient approximate nearest-

Table 4
Examples of how qualification, requirement level, and keywords are mapped to ISCED 2011 levels.

ISCED	Qualification	Req. Level	Example Keywords
24	Helper jobs	1	Trainee (Auszubildende/r)
35_2	Vocational Training	1, 2	official job titles
35_3	Vocational Training	2, 3	official job titles
55	Additional Vocational Training	3	Vocational Specialist (Berufsspezialist/in)
64	University Degree	3	✗
65	Additional Vocational Training	3	Master Craftsperson (Meister/in)
74	University Degree	4	✗
75	Additional Vocational Training	4	Certified IT Technical Engineer
84	University Degree	4	Professor

neighbour search in logarithmic time complexity. The final decision on the k-NN search is made by a majority vote of the top-k-nearest neighbours, where k is a hyperparameter that can be tuned according to the desired level of granularity. The majority vote determines the KldB 2010 code and the ISCED 2011 level separately.

3.4.1. Inference Examples

For the query Construction Supervisor (Baustellenleitung) we get the following results:

k-**nearest**	**ISCED**	**Type**	**CS**
Construction Supervisor (Bauaufseher)	74, 75	31194	0.945
Construction Supersivor (Bauaufseherin)	74, 75	31194	0.942
Contruction Manager (Baustellen-Manager)	74, 75	31194	0.939,

where CS is the cosine similarity between the query and the item. Combing them with a majority vote results in ISCED 2011 74, 75 (Master Degree, Master Professional) and KldB 31194 (Managers in construction scheduling and supervision, and architecture). The requirement level 4 of the nearest job titles result in high educational levels. The fourth digit being 9 indicates jobs with managerial duties.

Adding skills "Bathroom planning, Customer Management, Sanitary Engineering, Personnel Responsibility" ("Badplanung, Kundendienst, sanitärtechnik, personalverantwortung") specifies the query further and results in:

k-**nearest**	**ISCED**	**Type**	**CS**
Sanitary Engineering Manager (Sanitärtechnikmanager)	65, 75	34293	0.931
Sanitary Engineering Manager (Sanitärtechnikmanagerin)	65, 75	34293	0.928
Installation Supervisor (Assembly) (Installationsleiterin (montage))	35_3, 55, 65, 75	31193	0.921

Again, combining them with a majority vote result in ISCED 2011 65, 75 (Bachelor Professional, Master Professional), and KldB 34293 (supervisors in sanitation, heating, ventilation, and air conditioning). Adding specific skills uncovers jobs relevant to the task of sanitary engineering and management. The requirement level 3 indicates a specialist job, which requires additional specialist vocational training. This reclassification also shifts the title from "Manager" to "Supervisor", underscoring its hands-on nature as a functional specialist lead role that relies on advanced vocational training rather than purely managerial duties.

4. Results

4.1. Model Evaluation

A comparative evaluation is particularly difficult as there is no common evaluation data set or procedures for KldB classification. To facilitate a quantitative and standardised evaluation we compare our method

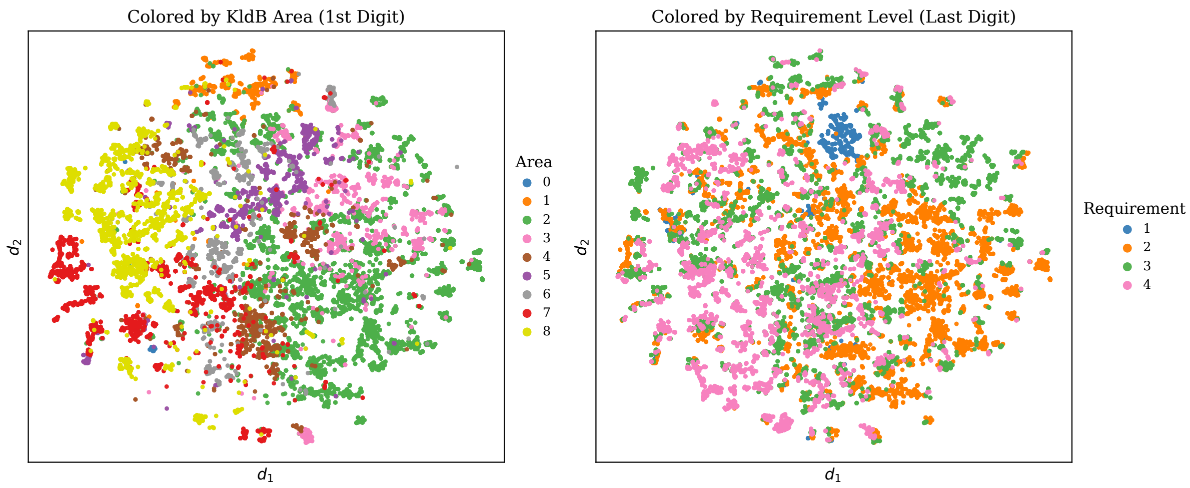

Figure 2: 2D -projection using t-SNE [36] of the embedding of the test dataset learned by our model, showing the structure it has learned for KldB 2010 areas (left) and requirement levels (right). We can see clusters evolving - especially in the requirement level groupings.

to a k-NN classifier fit on a variety of pre-trained models on occupational data. We acknowledge that the comparison may not be fair, as only our model has been explicitly trained on KldB data, but as the results suggest training on ESCO and even a simple German corpus provides a decent baseline. Furthermore, we compare against an TF-IDF embedding [34] both as a basis for a k-NN classifier and as features for a logistic regression classifier. We also evaluate the retrieval accuracy of the embedding models by inspecting their respective *mAP@3* and *mAP@5* metrics.

In the requirement-level classification, we take a closer look at the predictions on a class level and compare our results with one other method. Note, that the model we use for comparison was trained to classify full job ads [35], so it's reported performance here may not reflect its capabilities on the full information provided by the complete job ad.

The data set is divided into 80% for training and 20% for validation and testing. The metrics are computed per class and aggregated by the macro-average, giving equal weight to the minority and majority classes. For testing when using evaluation data from official sources, we have found that $k = 1$ is optimal. The results of the evaluation on the test/train split with $k = 1$ are shown in Table 5 and we show a low-dimensional visualization of the embedding space in Figure 2.

4.2. Ablation Experiments

To gauge how much the classifier relies on surface forms rather than true semantic content, we conducted four targeted ablation studies. Each study starts from a *Original* test slice and rewrites the titles according to a specific perturbation rule. All rewritten titles remain plausible German phrases, so no out-of-domain noise is injected. Table 7 summarises the macro metrics.

Management nouns (Leiter / Leiterin / Leiter/in → Leitung). Replacement of the gendered or slash-notation head noun with its gerundive form (*Leitung*) affects 765 titles. The subgroup–level accuracy drops by 1.7 pp (0.968 → 0.951), while the requirement digit loses only 0.4 pp. Thus, the embedding is largely insensitive to the noun morphology often used to encode gender.

Gender Inflection On 3,324 neutral titles, we generated masculine and feminine variants. Performance declines by less than 0.6 pp at the subgroup level and remains virtually unchanged for the requirement digit, confirming that the model handles gender-specific suffixes (–*er* vs. –*in*) robustly.

Table 5
Evaluation metrics (macro-F_1, rounded to three decimals). Baselines use TF–IDF or pretrained embeddings on occupational ontologies [37, 38]. The best result in a setting is indicated by a **bold** typeface.

Classification Evaluation

Level	LogReg +	k-NN +				Our Approach
	TF–IDF‡‡	TF–IDF‡‡	gBERT†	CareerBERT‡	JobBERT‡	64 dim.
Area	0.910	0.960	0.955	0.950	0.940	**0.976**
Main Group	0.898	0.942	0.943	0.944	0.917	**0.968**
Group	0.850	0.923	0.906	0.914	0.884	**0.948**
Sub-Group	0.717	0.888	0.848	0.866	0.846	**0.926**
Type	0.632	0.858	0.778	0.792	0.780	**0.886**
Requirement	0.852	0.923	0.900	0.903	0.928	**0.961**

Retrieval Evaluation

Level	mAP@3			mAP@5		
	CareerBERT‡	JobBERT‡	Ours (64 dim.)	CareerBERT‡	JobBERT‡	Ours (64 dim.)
Area	0.967	0.951	**0.981**	0.960	0.940	**0.979**
Main Group	0.953	0.929	**0.971**	0.944	0.918	**0.968**
Group	0.938	0.884	**0.960**	0.928	0.899	**0.956**
Sub-Group	0.917	0.888	**0.946**	0.905	0.876	**0.939**
Type	0.888	0.861	**0.929**	0.874	0.847	**0.919**
Requirement	0.947	0.946	**0.970**	0.936	0.937	**0.964**

† 1024-dimensional sentence embeddings trained on a German retrieval corpus.
‡ 768-dimensional sentence embeddings trained on the ESCO ontology[39].
‡‡ Uses $(5, 3)$-grams.

Table 6
Evaluation metrics for the requirement level (5th digit) of the KldB 2010 codes. We compare our results with the classification model `oja_reqlevel_de` (OJRD) [40, 35] and the averaged results.

Requirement Level	Precision		Recall		F1	
	Ours	OJRD	Ours	OJRD	Ours	OJRD
1: Helper Jobs	**0.963**	0.319	**0.963**	0.536	**0.963**	0.400
2: Specialist Activities	**0.983**	0.551	**0.979**	0.799	**0.981**	0.652
3: Complex Specialist Activities	**0.951**	0.605	**0.960**	0.441	**0.956**	0.512
4: Highly Complex Activities	**0.959**	0.675	**0.942**	0.448	**0.946**	0.533
	Ours	OJRD				
Macro Averaged F1	**0.961**	0.525				

Word-Order Reversal For 1,005 two-token titles we swapped the token order (*chemietechnischer Assistent* $\mapsto$ *Assistent chemietechnischer*). Because German occupational compounds are mainly head-final, this permutation removes a strong lexical cue. Accuracy drops by 1.9 pp at the subgroup level and 1.4 pp at requirement, but stays above 0.85, indicating that contextual embedding still recovers much of the semantics.

Across the three scenarios, the classifier retains at least 0.85 accuracy at the subgroup level and more than 0.98 at higher aggregation levels. The greatest vulnerability arises when the syntactic head is moved to sentence initial position, reinforcing the intuition that the model encodes head nouns more strongly than modifiers. Still, the modest degradation confirms that the embedding space has learned a degree of permutation invariance that is indispensable for noisy inputs in the real world.

Table 7
Macro-averaged performance (Sub-Group = 4-digit, Requirement = 5th digit) under three perturbation scenarios. Δ denotes the absolute change relative to the original slice.

Perturbation	**Sub-Group (4-digit)**			**Requirement (5th digit)**		
	Acc.	F1	ΔAcc. [pp]	Acc.	F1	ΔAcc. [pp]
Management noun	0.951	0.952	-1.7	0.982	0.981	-0.4
Gender variant	0.866	0.863	-0.6	0.950	0.949	-0.3
Word order	0.858	0.862	-1.9	0.881	0.882	-1.4

5. Discussion and Future Work

We have presented a lightweight, and embedding-based knowledge representation system that aligns free-form German job titles with the KldB 2010 and ISCED 2011 ontologies, enabling data-driven labour market analytics. Designed for real-time recommenders, the system is fast, memory efficient and fully explainable.

Across the KldB 2010 hierarchy the system attains high macro-accuracy and F1 scores, with peak performance at the area and main-group levels. Performance gradually degrades toward the type level, which is expected given the greater lexical variability of job titles and the smaller number of cases available at the more granular classification levels. Despite this, the model performs strongly on the requirement dimension, showing that it still captures much of the semantic complexity of job roles.

5.1. Future Work

Currently, the system uses the KldB 2010 codes to infer ISCED 2011 levels, but this could be inverted to improve the KldB predictions. This could be achieved by introducing a two-step prediction process: first, predicting the ISCED 2011 level based on the job title and then using the predicted ISCED level to refine the KldB 2010 prediction. This would allow the system to take advantage of the educational hierarchy to improve the accuracy of KldB predictions, especially for job titles that are ambiguous or have multiple interpretations.

The system is easily extensible: new ontologies, languages, or task-specific modules can be plugged in with minimal engineering effort. The planned extensions include (i) adding ESCO/ISCO codes and multilingual job titles to widen the global applicability, (ii) replacing the current rule-based ISCED 2011 mapper with a learned component, and (iii) revisiting the training algorithm to make the KldB 2010 hierarchy explicit, which should bolster performance at the type level.

5.2. Evaluation Challenges and Opportunities

Reliable evaluation remains difficult because there is no publicly available, gold-standard set of free-form titles mapped to all 1300 KldB 2010 codes. Although we can benchmark on the subset of titles already present in official datasets, truly open-ended job titles still require expert annotation, which is costly and prone to subjective disagreement. A promising alternative is controlled access to social security contribution data, where employers have already supplied both job titles and KldB 2010 codes. Such administrative data would provide a much larger quasi-gold corpus against which the accuracy and fairness of the system could be measured.

Acknowledgments

The authors thank the German Federal Employment Agency (BAfA) for providing the data and BERUFENET.API, which enables us to access the occupational database. The authors also thank the Stepstone Group for supporting this research and providing the necessary resources. The authors acknowledge the contributions of our colleagues at Stepstone Group AI Labs, who provided valuable feedback and insights during the development of this work. We also thank the contributors to the HuggingFace model repository for providing the baseline Sentence-BERT model that we used for fine-tuning. Lastly, the authors thank Kai Krüger from the Federal Institute of Vocational Training and Education for providing additional information about the `oja_reqlevel_de` model used for evaluation.

Declaration on Generative AI

During the preparation of this work, the authors used LLMs for grammar and spelling checks, summarising, and rephrasing. After using these tools/services, the authors reviewed and edited the content as needed and take full responsibility for the publication's content.

References

[1] B. Christoph, B. Matthes, C. Ebner, Occupation-based measures—an overview and discussion, Kölner Zeitschrift für Soziologie und Sozialpsychologie 72 (2020) 41–78. doi:10.1007/s11577-020-00673-4.

[2] Bundesagentur für Arbeit, Klassifikation der Berufe 2010 - überarbeitete Fassung 2020 Band 1: Systematischer und alphabetischer Teil mit Erläuterungen, Bundesagentur für Arbeit, Nürnberg, 2021.

[3] J. Dörpinghaus, J. Binnewitt, S. Winnige, K. Hein, K. Krüger, Towards a German labor market ontology: Challenges and applications, Applied ontology 18 (2023) 343–365. doi:10.3233/ao-230027.

[4] Handbuch zum Deutschen Qualifikationsrahmen, Bund-Länder-Koordinierungsstelle für den Deutschen Qualifikationsrahmen für lebenslanges Lernen, Berlin, 2013. URL: https://www.dqr.de/dqr/shareddocs/downloads/media/content/dqr_handbuch_01_08_2013.pdf?__blob=publicationFile&v=2.

[5] Deutscher Qualifikationsrahmen für lebenslanges Lernen: Liste der zugeordneten Qualifikationen, Bund-Länder-Koordinierungsstelle für den Deutschen Qualifikationsrahmen für lebenslanges Lernen, 2024. URL: https://www.dqr.de/dqr/shareddocs/downloads/media/content/2024_dqr_liste_zugeordnete_qualifik_01082024.pdf?__blob=publicationFile&v=2.

[6] P. Méhaut, C. Winch, The european qualification framework: skills, competences or knowledge?, European educational research journal 11 (2012) 369–381.

[7] UNESCO Institute for Statistics, International Standard Classification of Education (ISCED) 2011, UNESCO Institute for Statistics, 2012. doi:10.15220/978-92-9189-123-8-en.

[8] D. Zhang, J. Liu, H. Zhu, Y. Liu, L. Wang, P. Wang, H. Xiong, Job2Vec: Job title benchmarking with collective multi-view representation learning, in: Proceedings of the 28th ACM International Conference on Information and Knowledge Management (CIKM '19), ACM, Beijing, China, 2019, pp. 2763–2772. URL: https://doi.org/10.1145/3357384.3357825. doi:10.1145/3357384.3357825.

[9] J. Liu, Y. C. Ng, Z. Gui, T. Singhal, L. T. M. Blessing, K. L. Wood, K. H. Lim, Title2Vec: A contextual job title embedding for occupational named entity recognition and other applications, Journal of Big Data 9 (2022). URL: https://journalofbigdata.springeropen.com/articles/10.1186/s40537-022-00649-5. doi:10.1186/s40537-022-00649-5.

[10] N. Reimers, I. Gurevych, Sentence-bert: Sentence embeddings using siamese bert-networks, in: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), 2019, pp. 3982–3992.

[11] A. Khan, T. Wu, X. Chen, LLM+KG@VLDB'24 Workshop Summary: Data Management Opportunities in Unifying Large Language Models and Knowledge Graphs, in: Proceedings of the VLDB 2024 Workshops, 2024. URL: https://arxiv.org/abs/2410.01978, arXiv:2410.01978.

[12] R. S. Dehal, M. Sharma, E. Rajabi, Knowledge Graphs and Their Reciprocal Relationship with

Large Language Models, Machine Learning and Knowledge Extraction 7 (2025) 38. URL: https://doi.org/10.3390/make7020038. doi:10.3390/make7020038.
[13] The Stepstone Group, The stepstone group — we help companies to hire the right talent, 2025. URL: https://www.thestepstonegroup.com/english/, accessed 2025-06-11.
[14] N. Dawson, M.-A. Williams, M.-A. Rizoiu, Skill-driven recommendations for job transition pathways, PloS one 16 (2021) e0254722. doi:10.1371/journal.pone.0254722.
[15] T. Clemens, C. Dibben, A method for estimating wage, using standardised occupational classifications, for use in medical research in the place of self-reported income, BMC medical research methodology 14 (2014) 59. doi:10.1186/1471-2288-14-59.
[16] R. Zbib, L. A. Lacasa, F. Retyk, R. Poves, J. Aizpuru, H. Fabregat, V. Simkus, E. García-Casademont, Learning job titles similarity from noisy skill labels, arXiv [cs.IR] (2022). arXiv:2207.00494.
[17] G. Brunello, P. Wruuck, Skill shortages and skill mismatch: A review of the literature, Journal of Economic Surveys 35 (2021) 1145–1167. doi:10.1111/joes.12424.
[18] R. Baskaran, J. Müller, Classification of German Job Titles in Online Job Postings Using the KldB 2010 Taxonomy, Technical Report, &effect data solutions GmbH, 2023. URL: https://www.and-effect.com/publications/2022-11-21_technical_report_kldb.pdf, version 1, last updated 07 Nov 2023.
[19] P. Safikhani, H. Avetisyan, D. Föste-Eggers, D. Broneske, Automated occupation coding with hierarchical features: A data-centric approach to classification with pre-trained language models, Discover Artificial Intelligence 3 (2023). doi:10.1007/s44163-023-00050-y.
[20] J.-J. Decorte, J. V. Hautte, T. Demeester, C. Develder, Jobbert: Understanding job titles through skills, arXiv preprint arXiv:2109.09605 (2021). URL: https://arxiv.org/abs/2109.09605.
[21] M. de Groot, J. Schutte, D. Graus, Job posting-enriched knowledge graph for skills-based matching, arXiv preprint arXiv:2109.02554 (2021). URL: https://arxiv.org/abs/2109.02554.
[22] A. Seif, S. Toh, H. K. Lee, A dynamic jobs–skills knowledge graph, in: Proceedings of the 4th Workshop on Recommender Systems for Human Resources (RecSys in HR 2024), CEUR-WS, 2024. URL: https://recsyshr.aau.dk/wp-content/uploads/2024/10/RecSysHR2024-paper_1.pdf.
[23] D. T. A. May, Philipp, German bert large paraphrase cosine, 2025. URL: https://huggingface.co/deutsche-telekom/gbert-large-paraphrase-cosine, huggingFace repository, accessed 2025-12-06.
[24] A. Jaiswal, A. R. Babu, M. Z. Zadeh, D. Banerjee, F. Makedon, A survey on contrastive self-supervised learning, Technologies 9 (2020) 2.
[25] E. Rojas, A. N. Kahira, E. Meneses, L. B. Gomez, R. M. Badia, A study of checkpointing in large scale training of deep neural networks, arXiv preprint arXiv:2012.00825 (2020).
[26] D. Soydaner, A comparison of optimization algorithms for deep learning, International Journal of Pattern Recognition and Artificial Intelligence 34 (2020) 2052013.
[27] H. Hihn, D. A. Braun, Online continual learning through unsupervised mutual information maximization, Neurocomputing 578 (2024) 127422.
[28] M. Henderson, R. Al-Rfou, B. Strope, Y.-H. Sung, L. Lukács, R. Guo, S. Kumar, B. Miklos, R. Kurzweil, Efficient natural language response suggestion for smart reply, arXiv preprint arXiv:1705.00652 (2017).
[29] A. Kusupati, G. Bhatt, A. Rege, M. Wallingford, A. Sinha, V. Ramanujan, W. Howard-Snyder, K. Chen, S. Kakade, P. Jain, et al., Matryoshka representation learning, Advances in Neural Information Processing Systems 35 (2022) 30233–30249.
[30] S. L. Schneider, The international standard classification of education 2011, in: Class and stratification analysis, volume 30, Emerald Group Publishing Limited, 2013, pp. 365–379.
[31] A. Fischer, BERUFENET.API: Dokumentation zur berufenet-api der bundesagentur für arbeit, 2025. URL: https://github.com/AndreasFischer1985/berufenet-api, gitHub repository, accessed 2025-12-06.
[32] Y. A. Malkov, D. A. Yashunin, Efficient and robust approximate nearest neighbor search using hierarchical navigable small world graphs, IEEE transactions on pattern analysis and machine intelligence 42 (2018) 824–836.
[33] Y. A. Malkov, D. A. Yashunin, hnswlib: Header-only C++/python library for fast approximate

nearest neighbors, 2023. URL: https://github.com/nmslib/hnswlib, gitHub repository, accessed 2025-06-11.
[34] K. Sparck Jones, A statistical interpretation of term specificity and its application in retrieval, Journal of Documentation 28 (1972) 11–21. URL: https://doi.org/10.1108/eb026526. doi:10.1108/eb026526.
[35] K. Krüger, Using explainable ai for robustness checks in requirement level classification for german online job advertisements, Springer Special Issue (forthcoming). In press; publication details pending.
[36] L. van der Maaten, G. Hinton, Visualizing data using t-SNE, Journal of Machine Learning Research 9 (2008) 2579–2605. URL: http://www.jmlr.org/papers/v9/vandermaaten08a.html.
[37] 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 (forthcoming), Expert Systems With Applications (2025).
[38] J. Decorte, J. V. Hautte, T. Demeester, C. Develder, Jobbert: Understanding job titles through skills, CoRR abs/2109.09605 (2021). URL: https://arxiv.org/abs/2109.09605. arXiv:2109.09605.
[39] R. Varga, K. Reuter, H.-J. Bürger, S. Broeckx, ESCO: Towards a Semantic Web for the European Labor Market, in: Proceedings of the 1st Workshop on Open Data and Knowledge Graphs in Practice (LDK 2014), volume 1268 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2014. URL: http://ceur-ws.org/Vol-1268/paper-04.pdf.
[40] K. Krüger, Kkrueger/oja_reqlevel_de, 2025. URL: https://huggingface.co/KKrueger/oja_reqlevel_de, huggingFace repository, accessed 2025-22-06.