

An analysis of Computer Science in OJAs with a dual-lingual ontology approach

Michael Tiemann ¹, Jens Dörpinghaus ², and Venkatesh Hariharapura Shivashankar³

Abstract: Increasing globalization of the Internet has led to a growing need for the processing of data in multiple languages. This has resulted in a significant increase in the amount of multilingual data available on the web, presenting a significant challenge for accessing, processing, and integrating data from different language sources. An ontology provides a shared and precise source for system interoperability and the reuse of knowledge bases. The objective of this study is to develop an efficient approach for mapping and enriching cross-domain, dual-lingual ontologies. In this case, we will combine the Computer Science Ontology (CSO) and DBpedia. The resulting taxonomy will be analyzed using a German online job advertisement dataset of 3,567,240 records to identify trends in the development and/or requirement of tools in the Computer Science domain. In order to achieve this, we employ a comparative analysis of two distinct approaches: a holistic, hierarchical approach and a precision-driven approach. The former considers the child-parent relation of the ontology, whereas the latter solely identifies those tools directly associated with a topic in the CSO.

Keywords: Ontology Mapping, dual-lingual ontologies, Labor Market Research, OJAs

1 Introduction

The increasing globalization of the internet has led to a growing need for the processing of data in multiple languages. This has resulted in a significant increase in the amount of multilingual data available on the web, which presents a significant challenge for accessing, processing, and integrating data from different language sources [Da14]. Analysing data from various sources always presents researchers with challenges of varied understandings of concepts but also opportunities for knowledge discovery in various domains. In such contexts, the processing of multilingual content presents significant challenges. Firstly, content is frequently semi-structured or unstructured, making it challenging to gather and integrate from different data sources. Additionally, user-generated content lacks a consistent structural organization, necessitating the collection of relevant information from multiple sites. Furthermore, different data source platforms vary in their organization and structure, further complicating content integration. The objective of this study is to develop an efficient approach for mapping and enriching cross-domain, dual-lingual ontologies. The Computer Science Ontology (CSO) and Wikidata will be combined.

¹ Federal Institute for Vocational Education and Training (BIBB), Germany ,
tiemann@bibb.de,  <https://orcid.org/0000-0001-7136-2744>

² Federal Institute for Vocational Education and Training (BIBB), Germany, University of Koblenz, Germany,
doerpinghaus@uni-koblenz.de,  <https://orcid.org/0000-0003-0245-7752>

³ University of Koblenz, Germany, venkateshhs@uni-koblenz.de

It is of paramount importance for individuals, companies, and governments to understand the developments in the labor market and how tools associated with occupations have changed over time. Regularly, consultants warn about high numbers of workers needing to shift occupational categories and pick up new skills by the year 2030 as a result of automation and other technological advancements [Ma17]. The contents of jobs undergo constant change as new positions are created and existing ones are reinterpreted. A World Economic Forum report indicates that workers should undergo retraining and skill enhancement to remain aligned with the evolving demands of the labor market. Furthermore, recognizing emerging employment opportunities and the abilities essential for success in them necessitates an understanding of how tools associated with jobs have evolved. Those who can adopt new tools and skills will be better positioned for success as the employment landscape transforms due to challenges as decarbonisation, demographic change, technological change, and de-globalisation.

The skills required for digitalization and computer science can be analyzed among different stakeholders, including educational institutions, employers, employees, and financiers. In this context, continuing education advertisements can complement data from curricula, job advertisements, and labor market surveys. The German labor market is dynamic, with technical innovations and social changes leading to new skills needed by employees. Vocational education and training (VET), retraining, and continuing vocational education and training (CVET) are essential for responding to these new demands [Do19; He16]. Machine Learning (ML) as part of the development and use of Artificial Intelligence (AI) are two key domains in current technological progress. Throughout this paper we will focus on these.

The objective of this paper is to address the following research questions:

1. How can we efficiently enrich the CSO ontology with other data from Wikidata and DBpedia to create a dual-language ontology enriched with tools and entities to analyze German OJAs?
2. What is the demand of different areas within CSO, namely AI and ML? How can we analyze this demand, and how do they develop in different occupations?

The remainder of this work is organized as follows: After the introduction, we will discuss related work, followed by a discussion of methods and data used to create a dual-language ontology and tools extraction. We will then present results on OJAs and conclude with a discussion and outlook.

2 Related Work

To analyze labor market data like OJAs (Online Job Advertisements), social media data or (vocational) education data we find an increasing interest in mining data from educational databases, advertisements and information systems, see [DIH17; MT13; RV07]. Over the

last decade, there has been an increasing interest in mining these data [DSH23; DT23]. Supporting decision making and process management in education is key. However, there are still several challenges on data and data integration, see [KKM20]. Skill concepts have been widely used for the analysis of online job advertisements (OJAs), their visualization and provide a good starting point for matching open positions with corresponding employees [De18; Kr18]. OJAs are usually published in an online database, e.g. Monster, Stepstone or official organizations like the federal employment agency (BA) in Germany. They contain various data about the hiring company, the position and the requirements for the employee. The proposed technologies range from automated mapping of search terms to classification of skills, see [Zi12]. OJAs are a well-studied topic, especially in the English language [BP21; CJR14; KKG21; NPU20], and even historical advertisements have been studied [Ro20]. So far, there is little research on German OJAs [Bu22; GC20], although some work has been done, in particular focusing on qualification development [HS16].

In education research and pedagogy, several approaches towards competencies and skills exist which provide divergent resources: Not only with different meaning but also with different synonyms. While in English Competences, Skills and Knowledge are often used as synonyms, this does not hold for German language: Some words refer to multiple concepts, while some concepts are labeled with multiple words, see [GT23; KM22].

Only a few taxonomies for skills in the German language exist. Most important is the European Classification of Skills, Competences, Qualifications and Occupations (ESCO). It provides a multilingual hierarchy of skills and competences (and additionally qualifications) mapped to occupations with a full text description, scope notes and examples [GG20]. Some work has been done on semantic interoperability between skills and labor market documents [DVP15; Vr14]. However, ESCO is not the only German-language skills system available [KM22]. For example, the BA in Germany provides its own resources for structuring skills and linking them to occupations: the BERUFENET which is directly linked to other data, e.g. KURSNET [FD24]. Other approaches include contextual extraction and representation methods based on knowledge graphs, ontologies and taxonomies, see [Dö22a; Dö22b; DS19; DWB22; FGB22].

Ontology gaps are regions or aspects that are missing from or inadequately covered by the present ontology models. Limitations, shortcomings, or missing elements in the current knowledge and understanding of a given subject area are represented by these gaps [BSM09; GF10; KH04]. Hailu et. al. conclude that automatic translation of biomedical ontologies, specifically GO terms, is feasible and achieves high adequacy and fluency scores for the terms associated with most concepts [HCH14]. The methods using DBpedia and Google Translate API were effective in translating GO terms from English to German. Other studies disagree due to the coverage of entities in Wikidata and DBpedia [Bo20]. However, several studies indicate that the availability of multilingual information significantly improves the performance of the matching algorithms [FBO10; SHC11].

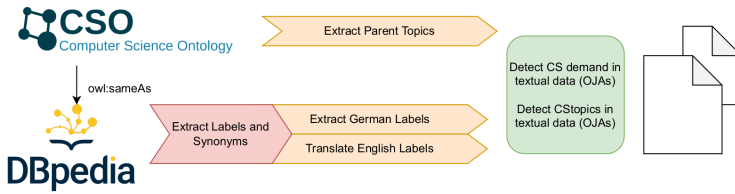


Fig. 1: The proposed workflow

Since we can only rely on very limited previous work, we will continue with a detailed discussion of the methods to encourage further research in this field.

3 Method

Figure 1 depicts the proposed workflow for enriching the CSO ontology with data from DBpedia. This process will result in a dual-language ontology, which will be enhanced with tools and entities for analyzing German OJAs. The following sections will present details on the data and the process, respectively.

3.1 Data

In this study, we utilized version v.3.2 of CSO. For the purpose of this research, we specifically employed the `owl:sameAs` relation between CSO and DBpedia. By employing this relation, we were able to identify equivalent resources for each topic in CSO within DBpedia. This correlation proved instrumental in facilitating dual-language translation and accessing relevant tools for each topic. A total of 5,193 topics in CSO had `sameAs` links to DBpedia, forming the foundation for conducting dual-lingual translation and retrieving tools and concepts associated with each topic.

The analysis will be conducted using the job advertisement data collected from various online job advertisement portals. This data set contains 3,567,240 OJAs. The data were collected over the period from 2015 to 2022. In total, there are nine distinct sets of job advertisements, labelled following ISCO-08 codes: 431 (Numerical Clerks), 7121 (Roofers), 7126 (Plumbers and Pipe Fitters), 8311 (Locomotive Engine Drivers), 2511 (Systems Analysts), 2512 (Software Developers), 2513 (Web and Multimedia Developers), 2514 (Applications Programmers), and 2519 (Software and Applications Developers and Analysts Not Elsewhere Classified). These occupations have been chosen since changes should be visible in the software occupations, where ML and AI are developed. The other occupations are test cases: Do we see demand for ML or AI related skills in them, as current discussions (see above) indicate such skills are becoming ever more important across the whole occupational landscape.

3.2 Creating a dual-lingual Ontology

In this study, we utilize CSO in the extraction process of tools, inspired by previous research that has highlighted the pivotal role of CSO as a knowledge hub, particularly its extensive connections to external resources such as DBpedia, Wikipedia, Freebase, YAGO, Cyc, and Microsoft Academic. This method encompasses the following steps:

1. **Names Translation.** Having classified each entity extracted from DBpedia as a tool, concept, or other, the next challenge is to obtain the German translation for each of these tool names. As identified through our literature review, various methods exist for translating text, and in this case, the language tags present in each DBpedia resource are utilized. It is important to note that not all DBpedia entities have German language tags associated with them. Furthermore, it is crucial to consider instances where the tool names are identical in both German and English. For instance, "Microsoft Teams" remains unaltered in both languages. In the absence of language tags, we employ the Google Translate API to obtain the requisite translations. The integrated approach of combining both methods ensures the acquisition of translations for all tools and mitigates the impact of lacking language tags on DBpedia resources.
2. **Synonym Extraction.** The subsequent stage in our process is the extraction of synonyms for each of the classified tools. The extraction of synonyms allows us to identify alternative names for each tool, which can be invaluable for NLP tasks. This comprehensive list of tool names, along with their alternative names, can serve as input for text extraction algorithms, enabling the identification of skills required for various job advertisements. To extract alternative labels for each tool, we utilize an inherent property in DBpedia called "is dbo:wikiPageRedirects of". This property provides a list of synonyms along with their corresponding hyperlinks.
3. **Retrieve Domains for a Parent Topic.** We implemented an algorithm traversing and extracting domain information associated with a given tool name. This method utilizes both direct matches within the dataset and semantic relations like "same as" in CSO. The significance of discovering various paths and, subsequently, top domains becomes apparent through a practical example involving the tool "TensorFlow". In our data, TensorFlow is extracted within the DBpedia resource "Speech recognition", corresponding to the CSO node "Speech Recognition". The objective is to identify all feasible paths where "Speech Recognition" appears, accomplished recursively using the "SuperTopicOf" relation.

3.3 Tools Extraction

We employ natural language processing (NLP) technologies for the purpose of entity recognition. In this process, we construct dictionaries comprising synonyms and perform string matching in order to map the entities to the respective taxonomy or ontology. To this

end, we utilise the SpaCy library, employing the `PhraseMatcher` approach. This methodology is frequently referenced in the literature, see [DD23; DSH23; DT23].

4 Results

4.1 Dual-lingual Ontology

The data pre-processing phase commenced with 6,879 unique topics from the CSO. To ensure that these topics had corresponding pages in DBpedia, the `SSame As` relation was employed. This resulted in the identification of 5,193 unique topics with corresponding DBpedia pages. However, not all topics in CSO had dedicated pages in DBpedia, reflecting the evolving nature of its coverage. The subsequent phase involved the extraction of entities from DBpedia for the aforementioned 5,193 topics. This resulted in the identification of 20,859 entities, which encompassed a diverse range of information, including tools, concepts, and other data.

4.2 Analysis of OJAs: Precision-Driven

To further validate the methodology for analyzing the progression of tools over time within various job sectors, we focus on the domain of AI, a rapidly emerging topic. This will enable us to not only track the evolution of AI tools but also to delve into ML, a significant subset of AI. By applying our extracted tools and their corresponding translations, we aim to conduct a thorough analysis of the trends and patterns in tool adoption. In Precision-Driven Extraction (PDE), the focus was on the direct extraction of tools specifically related to AI and ML topics, without considering CSO's hierarchical structure. The results obtained through this approach offer a more targeted perspective on the tools explicitly associated with AI and ML, without incorporating sub-domains or parent-child relationships within the ontology. This approach aims to provide a precise snapshot of tools directly linked to AI and ML, addressing potential challenges associated with ontology precision and generalization.

The percentages of AI tools extracted using PDE across various job datasets show a gradual increase over the years as shown in Figure 2. Notably, Job ID 2511 exhibits a steady rise from 0.53% in 2015 to 6.37% in 2022, indicating a growing demand for AI-related tools in the field of Systems Analysis. Similarly, Job ID 2512 experiences a substantial increase, reaching 12.40% in 2022. Job IDs 2513, 2514, and 2519 also depict a noticeable upward trend, suggesting a consistent demand for AI tools within these domains. Numerical Clerks (Job ID 431) saw an increase from 1% to just over 2%. Roofers (Job ID 7121) had a significant jump from nearly 0% to 8%. Plumbers and Pipe Fitters (Job ID 7126) grew modestly from 1% to 3%, while Locomotive Engine Drivers (Job ID 8311) remained steady at around 1%.

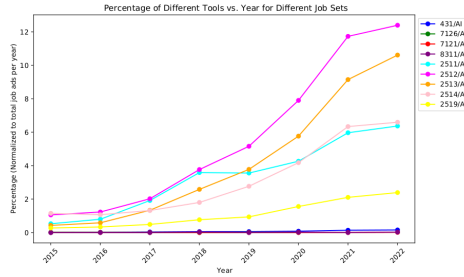


Fig. 2: Percentage of different AI tools vs. year using Precision-Driven extraction

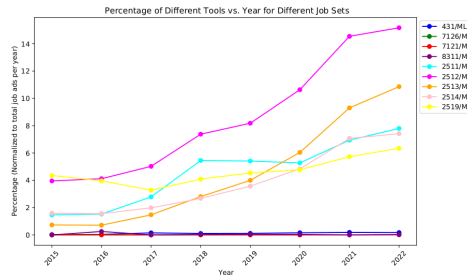


Fig. 3: Percentage of different ML tools vs. year using Precision-Driven extraction

The analysis of ML tool percentages across various job datasets using PDE is as shown in Figure 3. Job ID 431 (Numerical Clerks) demonstrates a gradual increase from 0.03% in 2015 to 0.17% in 2022. Job ID 7126 (Plumbers and Pipe Fitters) exhibits a relatively consistent but low percentage throughout the years, reaching 0.03% in 2019 but decreasing subsequently. Job ID 2519 demonstrates a consistent rise in ML tool percentages, reaching 6.34% in 2022, signifying a growing presence of ML tools in professions related to Software and Applications Developers and Analysts Not Elsewhere Classified.

4.3 Analysis of OJAs: Holistic approach

The analysis of AI tools using holistic hierarchical extraction (HE) as shown in Figure 4, involves a comprehensive consideration of the hierarchical structure of the CSO, provides valuable insights into the evolution of AI tools across various job datasets and years.

Notably, the sub directories related to specific job IDs exhibit distinctive patterns in AI tool adoption. For instance, within the Numerical Clerks category (Job ID 431), there is a significant increase in AI tool usage from 2.34% in 2015 to 3.94% in 2022, reflecting a steady integration of AI technologies in this particular professional domain. In contrast, Plumbers and Pipe Fitters (Job ID 7126) showcase a fluctuating trend, with AI tool usage peaking at 5.13% in 2018 and then gradually decreasing. Similarly, Roofers (Job ID 7121) display a notable surge in AI tool adoption, particularly from 2.66% in 2020 to 10.16% in

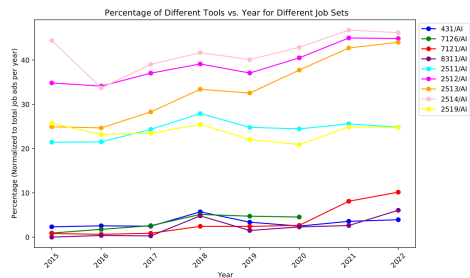


Fig. 4: Percentage of different AI tools vs. year using holistic hierarchical extraction

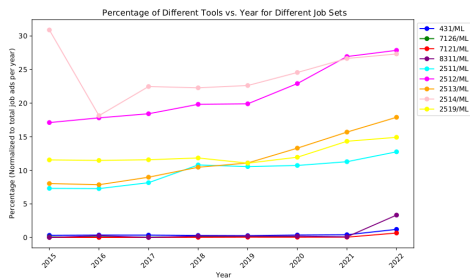


Fig. 5: Percentage of different ML tools vs. year using holistic hierarchical extraction

2021. It is noteworthy that the levels of AI tool adoption are so different: While it could be expected for Numerical Clerks to have a higher AI tools usage than around four percent, it seems astonishing in comparison to see Roofers peaking at just over ten percent. This can possibly be attributed to what is explicitly stated in OJAs: Skills or requirements that are already an established part of occupational contents tend not to be named, whereas what is new or out-of-the-ordinary will be explicated.

Using HE for analyzing the integration of ML tools as shown in Figure 5 across various professions, there is a clear trend of increasing reliance on ML technologies from 2015 to 2022. In IT-related jobs, Systems Analysts (Job ID 2511) saw an increase in ML tool usage from 7.31% to 12.76%, reflecting a steady adoption of ML for analytical purposes. Software Developers (Job ID 2512) experienced a robust growth from 17.11% to 27.87%, showcasing the critical role of ML in modern software engineering. Web and Multimedia Developers (Job ID 2513) also nearly doubled their use of ML tools, with an increase from 8.02% to 17.90%, pointing to ML’s growing impact on digital content creation. Applications Programmers (Job ID 2514) started with a high adoption rate at 30.90% and exhibited some fluctuation, stabilizing to 27.33%, which indicates a sustained significance of ML in programming. Software and Applications Developers and Analysts (Job ID 2519) showed a steady increase from 11.55% to 14.92%, suggesting a growing incorporation of ML in diverse developmental roles.

PDE offers an exquisite lens through which to view the explicit demand for these technologies,

with its precise extraction of particular AI and ML tools. It provides an objective and quantifiable pattern of how particular tools become more or less common over time, enabling the labor market to be analyzed with greater focus. This accuracy is especially helpful in determining which AI and ML technologies are becoming more popular and which are becoming less relevant. Both methodologies offer a micro and macro perspective of the labor market, making them essential for research. While HE provides a comprehensive understanding of the overall influence of AI and ML, PDE offers the accuracy required to trace the development of certain technologies. Together, they provide a comprehensive methodological framework for comprehending the dynamic interaction between new technologies and the changing nature of the workforce. This framework highlights the significance of both targeted and broad analysis in the context of technological advancement and its effects on the labor market.

5 Conclusions and Outlook

In the current era of extensive internet usage and global connectivity, the surge in multilingual data on the web has posed significant challenges, demanding innovative solutions. This study proposes an approach, a dual-lingual ontology translation method leveraging the CSO, to navigate the complex landscape of processing diverse and semi-structured multilingual content. The early stages of this involved the extraction of tools and concepts from the extensive DBpedia data set. This data set constituted the foundation for subsequent analyses, offering a comprehensive understanding of the intricate relationships between technological entities and their roles in the evolving job market.

The analysis of job advertisements, particularly in the context of AI and ML tool adoption, offers valuable insights into the evolving landscape of the job market. The examination of AI and ML tool trends, containing both percentage matches and the distinct count of tools, provides a comprehensive understanding of the diversity and specificity of technological requirements sought by employers. The prevalence of IT-related roles, such as systems analysts and software developers, in adopting AI and ML tools underscores the technology-centric nature of these professions. While this can be seen especially in the holistic approach it is noteworthy that there are relevant proportions of specific AI and ML tools being mentioned in non-IT jobs. Scrutinising those mentionings we find that these often relate to tools that are indirectly using AI, such as cloud computing solutions or word processing. One limitation of extracting tools from DBpedia is that only 22% of the 5,193 topics in DBpedia had at least one tool associated with them. This limitation hinders the effective extraction of tools from DBpedia. Another potential limitation of the proposed methodology is the ambiguity of bilingual names, exemplified by instances where tool names coincide with common names in the target language. Consequently, future work may entail improving the NLP algorithms to overcome the possible drawback of dual-lingual name ambiguity. In particular, work may be done to improve the algorithm's ability to take contextual cues into account and lessen false positives brought on by tool-name matches in irrelevant circumstances.

References

- [Bo20] Bouscarrat, L. et al.: Multilingual enrichment of disease biomedical ontologies. arXiv preprint arXiv:2004.03181/, 2020.
- [BP21] Beręsewicz, M.; Pater, R.: Inferring job vacancies from online job advertisements. Publications Office of the European Union, 2021.
- [BSM09] Biffl, S.; Sunindyo, W.D.; Moser, T.: Bridging Semantic Gaps Between Stakeholders in the Production Automation Domain with Ontology Areas. In: SEKE. Citeseer, pp. 233–239, 2009.
- [Bu22] Buchmann, M. et al.: Swiss Job Market Monitor: A Rich Source of Demand-Side Micro Data of the Labour Market. *European Sociological Review* 38/6, pp. 1001–1014, 2022.
- [CJR14] Carnevale, A. P.; Jayasundera, T.; Repnikov, D.: Understanding online job ads data. Georgetown University, Center on Education and the Workforce, Technical Report (April)/, 2014.
- [Da14] Dang, Y. et al.: An integrated framework for analyzing multilingual content in Web 2.0 social media. *Decision Support Systems* 61/, pp. 126–135, 2014.
- [DD23] Derksen, F.; Dörpinghaus, J.: Digitalization and Sustainability in German Continuing Education. In: *INFORMATIK 2023 - Designing Futures: Zukünfte gestalten*. Gesellschaft für Informatik e.V., Bonn, pp. 1945–1953, 2023.
- [De18] Degenhardt, S.: Kompetenzen für eine digitalisierte Arbeitswelt–Anforderungen an Aus- und Weiterbildung. In: *Digitaler Wandel in der Sozialwirtschaft*. Nomos Verlagsgesellschaft mbH & Co. KG, pp. 259–272, 2018.
- [DIH17] Dutt, A.; Ismail, M. A.; Herawan, T.: A systematic review on educational data mining. *Ieee Access* 5/, pp. 15991–16005, 2017.
- [Do19] Dobischat, R. et al.: Digitalisierung und die Folgen: Hype oder Revolution? *Bildung 2.1 für Arbeit 4.0?*, pp. 9–24, 2019.
- [Dö22a] Dörpinghaus, J. et al.: Context mining and graph queries on giant biomedical knowledge graphs. *Knowledge and Information Systems* 64/5, pp. 1239–1262, 2022.
- [Dö22b] Dörpinghaus, J. et al.: From social networks to knowledge graphs: A plea for interdisciplinary approaches. *Social Sciences & Humanities Open* 6/1, p. 100337, 2022.
- [DS19] Dörpinghaus, J.; Stefan, A.: Knowledge Extraction and Applications utilizing Context Data in Knowledge Graphs. In: *2019 Federated Conference on Computer Science and Information Systems (FedCSIS)*. IEEE, pp. 265–272, 2019.
- [DSH23] Dörpinghaus, J.; Samray, D.; Helmrich, R.: Challenges of Automated Identification of Access to Education and Training in Germany. *Information* 14/10, p. 524, 2023.

- [DT23] Dörpinghaus, J.; Tiemann, M.: Vocational Education and Training Data in Twitter: Making German Twitter Data Interoperable. *Proceedings of the Association for Information Science and Technology* 60/1, pp. 946–948, 2023.
- [DVP15] De Smedt, J.; le Vrang, M.; Papantoniou, A.: ESCO: Towards a Semantic Web for the European Labor Market. In: *Ldow@* www. 2015.
- [DWB22] Dörpinghaus, J.; Weil, V.; Binnewitt, J.: Towards the Analysis of Longitudinal Data in Knowledge Graphs on Job Ads. In: *The Workshop on Computational Optimization*. Springer, pp. 52–70, 2022.
- [FBO10] Fu, B.; Brennan, R.; O'sullivan, D.: Cross-Lingual Ontology Mapping and Its Use on the Multilingual Semantic Web. *MSW* 571/, pp. 13–20, 2010.
- [FD24] Fischer, A.; Dörpinghaus, J.: Web Mining of Online Resources for German Labor Market Research and Education: Finding the Ground Truth? *Knowledge* 4/1, pp. 51–67, 2024.
- [FGB22] Fettach, Y.; Ghogho, M.; Benatallah, B.: Knowledge graphs in education and employability: A survey on applications and techniques. *IEEE Access* 10/, pp. 80174–80183, 2022.
- [GC20] Gnehm, A.-S.; Clematide, S.: Text zoning and classification for job advertisements in German, French and English. In: *Proceedings of the Fourth Workshop on Natural Language Processing and Computational Social Science*. Pp. 83–93, 2020.
- [GF10] Grubic, T.; Fan, I.-S.: Supply chain ontology: Review, analysis and synthesis. *Computers in Industry* 61/8, pp. 776–786, 2010.
- [GGS20] González, L.; García-Barriocanal, E.; Sicilia, M.-A.: Entity Linking as a Population Mechanism for Skill Ontologies: Evaluating the Use of ESCO and Wikidata. In: *Research Conference on Metadata and Semantics Research*. Springer, pp. 116–122, 2020.
- [GT23] Grollmann, P.; Tiemann, M.: Skilled and Unskilled Work. From Theoretical Concepts to Social Practices. In (Herzog, L.; Zimmermann, B., eds.): *Shifting Categories of Work. Unsettling the Ways we Think about Jobs, Labor, and Activities*. Routledge, pp. 223–238, 2023.
- [HCH14] Hailu, N. D.; Cohen, K. B.; Hunter, L. E.: Ontology translation: A case study on translating the Gene Ontology from English to German. In: *Natural Language Processing and Information Systems: 19th International Conference on Applications of Natural Language to Information Systems, NLDB 2014, Montpellier, France, June 18-20, 2014. Proceedings* 19. Springer, pp. 33–38, 2014.
- [He16] Helmrich, R. et al.: Digitalisierung der Arbeitslandschaften: keine Polarisierung der Arbeitswelt, aber beschleunigter Strukturwandel und Arbeitsplatzwechsel. *Wissenschaftliche Diskussionspapiere*, 2016.

- [HS16] Hermes, J.; Schandock, M.: Stellenanzeigenanalyse in der Qualifikationsentwicklungsforschung. Die Nutzung maschineller Lernverfahren zur Klassifikation von Textabschnitten. Bundesinstitut für Berufsbildung, Bonn/, 2016.
- [KH04] Kauppinen, T.; Hyvönen, E.: Bridging the semantic gap between ontology versions. In: Web Intelligence Symposium, Finnish AI Conference. Vol. 20, pp. 2–3, 2004.
- [KKG21] Khaouja, I.; Kassou, I.; Ghogho, M.: A survey on skill identification from online job ads. *IEEE Access* 9/, pp. 118134–118153, 2021.
- [KKM20] Kovalev, S.; Kolodenkova, A.; Muntyan, E.: Educational data mining: current problems and solutions. In: 2020 V International Conference on Information Technologies in Engineering Education (Inforino). IEEE, pp. 1–5, 2020.
- [KM22] Krebs, B.; Maier, T.: Die QuBe-Kompetenzklassifikation als verdichtende Perspektive auf berufliche Anforderungen, tech. rep., Wissenschaftliche Diskussionspapiere, 2022.
- [Kr18] Kreuzer, C.: Visualisierung der Opportunity Recognition-Kompetenz von Industriekaufleuten. *Zeitschrift für Berufs- und Wirtschaftspädagogik* 114/2, pp. 247–271, 2018.
- [Ma17] Manyika, J. et al.: Jobs lost, jobs gained: What the future of work will mean for jobs, skills, and wages./, 2017.
- [MT13] Mohamad, S. K.; Tasir, Z.: Educational data mining: A review. *Procedia-Social and Behavioral Sciences* 97/, pp. 320–324, 2013.
- [NPU20] Ningrum, P. K.; Pansombut, T.; Ueranantasun, A.: Text mining of online job advertisements to identify direct discrimination during job hunting process: A case study in Indonesia. *Plos one* 15/6, e0233746, 2020.
- [Ro20] Ros, R. et al.: Mining Wages in Nineteenth-Century Job Advertisements. The Application of Language Resources and Language Technology to study Economic and Social Inequality. In: Proceedings of the Workshop about Language Resources for the SSH Cloud. Pp. 27–32, 2020.
- [RV07] Romero, C.; Ventura, S.: Educational data mining: A survey from 1995 to 2005. *Expert systems with applications* 33/1, pp. 135–146, 2007.
- [SHC11] Spohr, D.; Hollink, L.; Cimiano, P.: A machine learning approach to multilingual and cross-lingual ontology matching. In: The Semantic Web–ISWC 2011: 10th International Semantic Web Conference, Bonn, Germany, October 23–27, 2011, Proceedings, Part I 10. Springer, pp. 665–680, 2011.
- [Vr14] le Vrang, M. et al.: Esco: Boosting job matching in europe with semantic interoperability. *Computer* 47/10, pp. 57–64, 2014.
- [Zi12] Ziegler, P.: Zur Verwendung von Berufsinformation im Hinblick auf Matching in Deutschland und Österreich, tech. rep., AMS info, 2012.