knowledge

MDPI

Article

Web Mining of Online Resources for German Labor Market Research and Education: Finding the Ground Truth?

Andreas Fischer [1,*,†] and Jens Dörpinghaus [2,3,*,†]

[1] Forschungsinstitut Betriebliche Bildung (F-BB), 90408 Nürnberg, Germany
[2] Federal Institute for Vocational Education and Training (BIBB), 53113 Bonn, Germany
[3] Department of Computer Science, University of Koblenz, 56070 Koblenz, Germany
* Correspondence: andreas.fischer@f-bb.de (A.F.); jens.doerpinghaus@bibb.de (J.D.)
† These authors contributed equally to this work.

Abstract: The labor market is highly dependent on vocational and academic education, training, retraining, and further education in order to master challenges such as advancing digitalization and sustainability. Further training is a key factor in ensuring a qualified workforce, the employability of all employees, and, thus, national competitiveness and innovation. In the contribution at hand, we explore an innovative way to derive knowledge about learning pathways by connecting the dots from different data sources of the German labor market. In particular, we focus on the web mining of online resources for German labor market research and education, such as online advertisements, information portals, and official government websites. A key question for working with different data sources is how to find the ground truth and common data structures that can be used to make the data interoperable. We discuss how to classify and summarize web data from different platforms and which methods can be used for extracting data, entities and relationships from online resources on the German labor market to build a network of educational pathways. Our proposed solution is based on the classification of occupations (KldB) and related document codes (DKZ), and combines natural language processing and knowledge graph technologies. Our research provides the foundation for further investigation into educational pathways and linked data for labor market research. While our work focuses on German data, it is also useful for other German-speaking countries and could easily be extended to other languages such as English.

Keywords: web mining; knowledge discovery and data mining; knowledge discovery; labor market research; research and development towards society

check for updates

Citation: Fischer, A.; Dörpinghaus, J. Web Mining of Online Resources for German Labor Market Research and Education: Finding the Ground Truth? *Knowledge* **2024**, *4*, 51–67. https://doi.org/10.3390/knowledge4010003

Academic Editor: Gautam Srivastava

Received: 19 December 2023
Revised: 18 January 2024
Accepted: 6 February 2024
Published: 19 February 2024

Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

The web mining of data about the German labor market offers new and innovative ways to derive knowledge about learning pathways by connecting the dots from different data sources. The labor market is a domain with a variety of data structures connected to a variety of related applications (e.g., recommending suitable jobs to job seekers, listing skills for occupational titles or selecting suitable candidates for a job). Labor market research is mostly based on traditional methods such as surveys or the analysis of official statistical data (e.g., [1]). In the paper at hand, we explore a different approach—sourcing, analyzing and linking open data on various aspects of the labor market through the web mining of online resources. As a proof of concept, we build and extend a network of education pathways.

This is an important issue, as there is no ground truth for this type of network. Not all programs are officially regulated by the state or federal government, and chambers do not always publish relevant data. In addition, not all education pathways are formalized but some education pathways emerge from the offers on the market (i.e., rather informal learning pathways, cf. [2]).

There are a number of web resources offered by the government and official institutes. For example, the Federal Employment Agency (BA) offers comprehensive information about professions as well as related forms of (further) vocational education and training on its BERUFENET information portal (see Figure 1). In addition, the BA lists current offers for vocational training (AUSBILDUNGSSUCHE), further education (WEITERBILDUNGSSUCHE), study programs (STUDIENSUCHE), and job-related language support (SPRACHFÖRDERUNG), as well as coaching and activation measures (COACHING-UNDAKTIVIERUNG) on an information portal called KURSNET. It also provides information on typical salaries (ENTGELTATLAS) and professional reorientation (NEW-PLAN), lists online job advertisements (JOBSUCHE), and maintains an applicant directory (BEWERBERBÖRSE).

The Federal Institute for Vocational Education and Training (Bundesinstitut für Berufsbildung—BIBB) provides an overview and information about regulations for state-recognized training occupations (Berufesuche), which contain information on a variety of topics, such as job descriptions ("Berufsbild"), examination requirements ("Prüfungsanforderungen"), vocational training plans ("Berufsbildungsplan"), courses ("Lehrgang"), vocational aptitude requirements ("Berufseignungsanforderungen"), and curricula ("Lehrplan").

Last but not least, there are a number of classification frameworks related to German qualifications and occupations—e.g., the classification of occupations (Klassifikation der Berufe, KldB), its European counterpart, the "European Skills/Competences, Qualifications and Occupations" (ESCO, [3]), and the International Standard Classification of Jobs (ISCO-08). Notably, official statistics on the labor market are often available only in aggregated form, e.g., based on the Classification of Economic Activities (Klassifikation der Wirtschaftszweige 2008, WZ-08). Connectable taxonomies such as ESCO, see [3], are a good example of the central role of ontologies and, in particular, knowledge graphs in this field. However, single taxonomies such as ESCO cannot provide all details of local labor market needs and do not provide direct links to other hierarchies of skills, vocational education and training (VET), and continuing vocational education and training (CVET) data. Fortunately there are official tables relating ESCO classifications to KldB codes.

In general, the labor market relies heavily on vocational education and training, retraining, and advanced vocational qualification to meet challenges such as the ongoing digitalization [4] or sustainability [5]. In particular, the German education system offers special pathways for skilled workers in tertiary education [6], and distinguishes between initial vocational education (training, "Ausbildung" or retraining, "Umschulung") and continuing vocational education, which includes continuing vocational training ("Weiterbildung") and upgrading training ("Fortbildung"). In this regard, upgrading training is usually formally regulated (e.g., at the federal level [BBiG/HwO] or by the federal states [7,8]; this type of accreditation can also be found in other countries and enables quality assurance, leading to official recognition and approval by the relevant legislative or professional authorities). There are 1004 (re-)trainings that are regulated by companies or by the Crafts and Trade Code, 542 of which are regulated by the German state. The number of informal courses is much higher.

Understanding the differences between formal education and informal learning pathways (but also between official statistics and the labor market represented by online job or training ads) is, therefore, crucial. In this respect, the mass of data available online could be used to bridge the gap between the relatively slow traditional research using survey data and official regulations dealing with rapid changes in the labor market.

In this paper, we explore an innovative approach to generate knowledge on both formal and informal education pathways—sourcing, analyzing and linking open data on different aspects of the labor market through the web mining of online resources. As a proof of concept, we build and extend a network of educational pathways based on the data available online.

Specifically, we examine how the different data sources can be related to each other and how knowledge about the German labor and vocational training market can be generated

from them. (1) Based on official information available through web mining, we will look at the relationships between occupations and how to find (a) entry requirements and (b) training opportunities for each occupation in order to derive knowledge about educational pathways. (2) In addition, we will explore how different types of data and classifications can be related to each other based on existing identifiers and taxonomies (e.g., based on BERUFENET IDs, KldB codes, ESCO classes and ISCO classes) in order to obtain further information about all occupations. In terms of linking data sources, we focus on two research questions: Our first research question (RQ1) is, what are common data structures that can be used to make crawled data interoperable? Our second research question (RQ2) is, what kind of methods can be used for data, entity, event, and relationship extraction from German online labor market resources?

The remainder of this paper is organized as follows: The next section provides an overview of related works. Section 3 presents our methodological approach to querying and linking the data, including an overview of data schemas, web resources and methods. Section 4 is devoted to the results, where we discuss BIBB and BA education pathways and their interoperability and provide some illustrative examples of the kind of knowledge we were able to derive from the data. The final section contains our conclusions and outlook.

All URLs in this paper were accessed in December 2023.

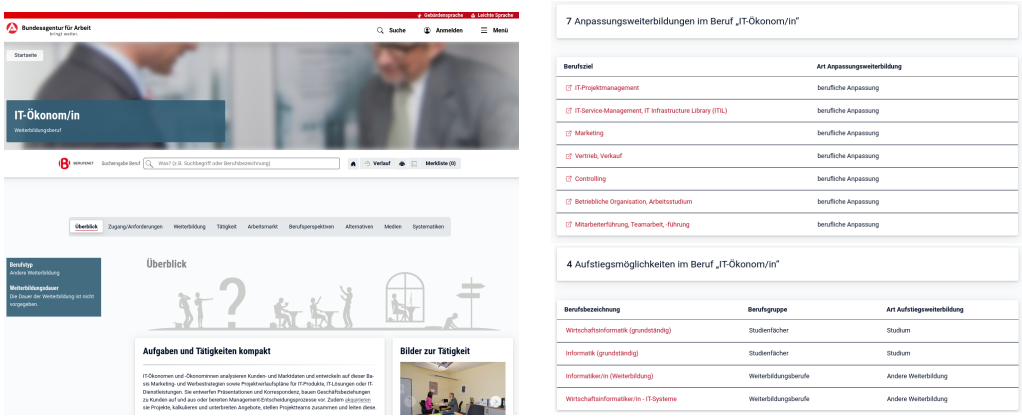

Figure 1. BERUFENET website for "IT-Economics (certified)", see https://web.arbeitsagentur.de/berufenet/beruf/15323 (accessed on 2 December 2023). (**Left**) The landing page with an overview. (**Right**) Continuing education programs that provide links to KURSNET and can be crawled using the API.

2. Related Work

Over the last decade, there has been an increasing interest in mining data from the web, e.g., educational databases, advertisements, and information systems [9–11]. Web mining refers to the application of data-mining techniques to discover and extract patterns and knowledge from web data (e.g., [5]). Supporting decision making and process management in education is key. The generic challenges are usually the automated extraction of knowledge from data (typically interpreted passages from texts) and the mapping to existing datasets. However, there are still several challenges related to the data and data integration [12]. Research questions that have been addressed with web mining techniques cover a wide range, for example, occupational inequality [13], questions of migration and language skills [14], discrimination [15], and students and later occupation [16].

With regard to the linkage of datasets on the labor market, Ortmann et al. used data from BERUFENET to quantify the similarity of jobs based on job competences and to relate this information via KldB code to a separate dataset on job changes from the national education panel [17]. For instance, they analyzed the proportion of job changes between different categories of the KldB by the amount of similarity of the jobs (distinguishing similar, related and complete career changes) and found job changes between 5-digit KldB

codes to be more likely for completely dissimilar job changes (49 percent) than for related (30 percent) or similar job changes (21 percent).

Another interesting area of research is the classification of online advertisements with regard to skills and taxonomies: Skill concepts have been widely used for the analysis of online job advertisements (OJAs) and provide a good starting point for matching open positions with corresponding employees [18,19]. OJAs are usually published in an online database such as Monster or Stepstone but also in databases of official organizations like the federal agency of employment (BA) in Germany. They contain various data about the hiring company, the position, and the requirements for the employee. OJAs are a well-studied topic, especially in the English language [15,20–22], and even historical advertisements have been studied [23]. So far, there is little research on German OJAs [24,25], although some work has been conducted, in particular, focusing on qualification development [26,27] and in the context of the greening of jobs [28,29]. The proposed technologies for skills extraction range from the automated mapping of search terms to the classification of skills [30] to complex applications of large language models (e.g., SkillGPT by [31]). Special attention has been paid to multi-label classification frameworks [32], building skill taxonomies [33], and, in general, the reflection on big data technologies [34], also for German OJAs [35]. While some authors treat competences, skills and knowledge as synonyms, we follow the KSAO model of competency proposed by Fischer and Neubert [36]: Knowledge, Skills, Abilities, and Other components (KSAO) are distinct components and prerequisites of competency—a context-specific disposition to perform well (cf. [37]). Therefore, competences and skills refer to related but distinct concepts.

Thus, while there are still some open questions with regard to common data structures and methods for data extraction (see Table 1), we can build on the experience with BERUFENET, OJAs and the existing taxonomies and structures for skills in German texts, for example, the KldB [38,39]. See Section 3.1 for more details. Specifically, we note several research gaps: First, to the best of our knowledge, no labor market research has been conducted on linking a wide range of online data sources. Second, the German labor market (in Germany, Austria and Switzerland) has several specific requirements, and only very limited work has focused on German texts regarding these requirements. Third, no work has been performed to link official training regulations to CVET advertisements, which would cover the majority of non-regulated training programs, see [40].

Since we can only rely on very limited previous work, we will first provide information on the data and continue with a detailed discussion of the methods used for our approach.

Table 1. Related work in the context of the proposed research questions for the German labor market.

Research Question	Literature
Common data structures	
Specific structures	[17,33,38,39]
General structures, ontologies	[41]
Methods for data extraction	
German OJAs	[24,25,42]
Skill extraction	[36,37]
Qualification development	[26,27]
Greening of job index (GOJI)	[28,29]
Other metadata like industrial sectors	[43]

3. Method

3.1. Data Schemata

Labor markets are complex fields with diverse data structures and multiple applications (for example, connecting job seekers to the right training or job [44]). As described above, the European ontology ESCO cannot provide all details of local labor market needs and does not provide links to other hierarchies of skills sufficiently. For example, in German-speaking countries, other taxonomies of occupations and skills are widely used. Thus,

when discussing data for occupational qualifications and certificates, we need to consider multiple data schemas and their relation to several relevant taxonomies.

In the context of occupations, the International Standard Classification of Occupations (ISCO) was developed by the International Labour Organization (ILO) (See https://www.ilo.org/public/english/bureau/stat/isco/isco08/) and was published in 1958, 1968, 1988, and, as its recent version, 2008. It was also used within the European Union (EU), and some German-speaking countries (Germany, Austria, and Switzerland) have linked their specific version to the ISCO 2008. ISCO maps to the ontology "European Skills, Competences, Qualifications and Occupations" (ESCO), which links skills and competences to occupations described in ISCO. Gonzalez et al. state that few works have described the analysis and use of ESCO (see [45]). Some work has been conducted on the semantic interoperability between skills and labor market documents, which was initially promised by ESCO [44]. Other researchers have tried to use data from ESCO and Wikidata for the text mining of the scientific literature (see [45]), or for curriculum analysis (see [46]). Recent research has provided a generic mining and mapping approach [47] and automated ontology alignment for ESCO and the English-language O*NET [48].

In Germany, the classification of occupations ("Klassifikation der Berufe", KldB) (See https://statistik.arbeitsagentur.de/DE/Navigation/Grundlagen/Klassifikationen/Klassifikation-der-Berufe/KldB2010-Fassung2020/KldB2010-Fassung2020-Nav.html) and related document codes (DKZ) are the reference for IAB (Institut für Arbeitsmarkt- und Berufsforschung) and the German Federal Employment Agency (Bundesagentur für Arbeit—BA). The most recent version is the 2020 revision of KldB 2010, which was completely redeveloped and renders the previous versions from 1988 and 1992 deprecated. It was developed to be compatible with ISCO-08. These data are used by the BA when matching candidates to jobs and are integrated into other IT applications. However, while part "B" of DKZ is dedicated to occupations, part "C" covers continuing professional development, "K" skills, and "A" higher education. All these parts are important to describe the access to education and training.

Formally, KldB codes (five-digit codes) are systematically related to DKZ identifiers: for instance, with regard to the classification of occupations there are eight-digit DKZ identifiers for each occupation (as well as for related education and training) which extend the corresponding KldB code by three additional digits. According to the BA, these DKZ 8-digits "form a much more dynamic 'sub-hierarchical level', which has a clear relationship to the KldB (each DKZ 8-digit code can be clearly assigned to a KldB 5-digit code), but is not part of the actual classification and can be adapted to changes in the real occupational landscape at high frequency" [49]. In the online edition of the KldB, eight-digit DKZ codes are currently returned when querying individual occupations (e.g., 43104-132 for "Data Scientist"), which also indicates the close connection between KldB and DKZ.

While these two classifications are widely used, other approaches have been introduced by Blossfeld [50], Erikson–Goldthorpe–Portocarero [51], the "Internationaler Sozioökonomischer Index des beruflichen Status (ISEI)" [52], and the "Standard International Occupational Prestige Scale (SIOPS)" [53]. Recently, the German Labor Market Ontology (GLMO) was introduced [41], providing linked data between KldB, ISCO and ESCO.

To summarize, based on detailed classification identifiers such as DKZ digits, occupations can be categorized using different national and international taxonomies.

3.2. Web Resources

One important source for up-to-date information on occupations and different forms of (further) vocational education and training are the information portals of the BA, especially BERUFENET. The Application Programming Interfaces (APIs) of BERUFENET and other services of the BA, among many others, have recently been documented by a civil society initiative called bund.dev (see https://bund.dev). For instance, information on the API of BERUFENET is available at https://github.com/bundesAPI/berufenet-api, see also Figure 2.

A complete list of the occupations available on BERUFENET can be obtained by a simple GET-request per page ($n = 179$), starting with page 0 (Listing 1).

Listing 1. GET-request for a page with occupations from BERUFENET.

```
berufe=$(curl -m 60 \
-H "X-API-Key: d672172b-f3ef-4746-b659-227c39d95acf" \
"https://rest.arbeitsagentur.de/infosysbub/"\
"bnet/pc/v1/berufe?suchwoerter=*&page=0")
```

Given the ID of an occupation, detailed information can be obtained by another GET-request per occupation, for instance, for BERUFENET-ID 15322 (Listing 2). In this way, it is possible to call up detailed information on all occupations (and related forms of education and training) online. Other services of the BA function in a similar way—the interested reader may refer (or even contribute) to the documentation provided online by bund.dev and the first author of this study (see Figure 2):

Listing 2. GET-request for a details on a occupation from BERUFENET.

```
berufeinfo=$(curl -m 60 \
-H "X-API-Key: d672172b-f3ef-4746-b659-227c39d95acf" \
"https://rest.arbeitsagentur.de/infosysbub/"\
"bnet/pc/v1/berufe/15322")
```

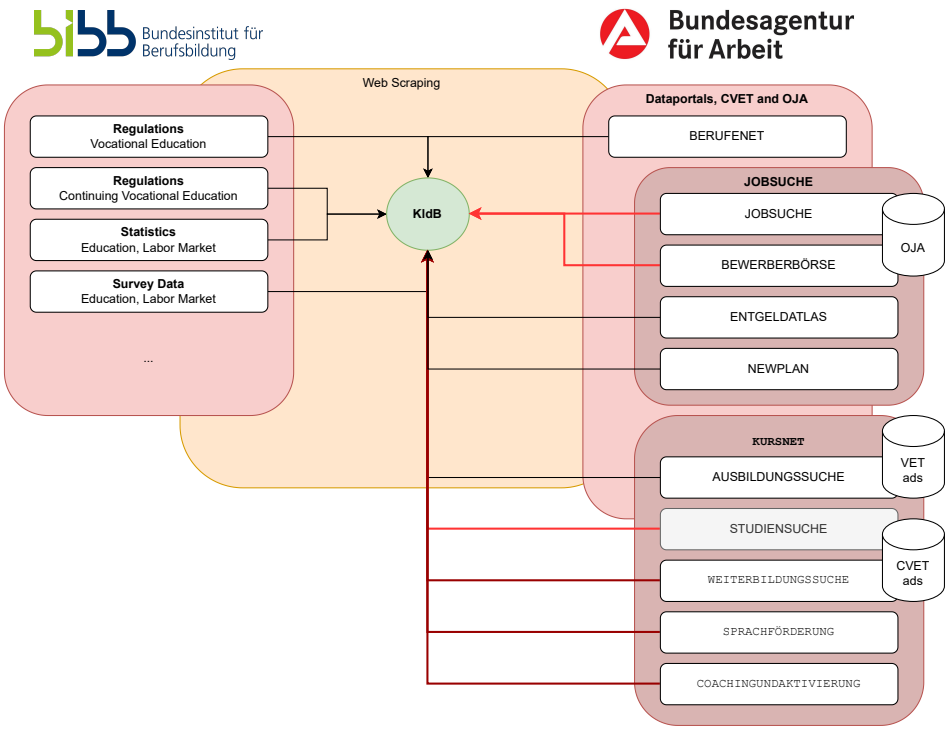

Figure 2. A visualization of the resources considered (see also Table 2) and how they can be linked to the classification of occupations (KldB): they provide a direct mapping (black arrows), can be mapped directly by string matching (red arrows), or the data are only partially available and require a more complex matching because the naming does not follow the standardized form (dark red arrows).

Table 2. Overview of available online data. Some data are available via an API, others only as a file download. Here, the source [DBA] refers to the download portal of the BA at https://download-portal.arbeitsagentur.de/files/ and [BS] to BIBB-Berufesuche at https://www.bibb.de/dienst/berufesuche/de/index_berufesuche.php. Obviously, very few historical data are available. See also Figure 2.

Data Set	Source	Initial Format	Historical Data	Data Records
BERUFENET (API)	BA	JSON		3569
AUSBILDUNGSSUCHE (API)	BA	JSON		>24 K
STUDIENSUCHE (API)	BA	JSON		>15 K
WEITERBILDUNGSSUCHE (API)	BA	JSON		>5 M
SPRACHFÖRDERUNG (API)	BA	JSON		>35 K
COACHINGUNDAKTIVIERUNG (API)	BA	JSON		> 100 K
JOBSUCHE (API)	BA	JSON		>900 K
BEWERBERBÖRSE (API)	BA	JSON		>1.7 M
ENTGELTATLAS (API)	BA	JSON		<3569
NEWPLAN (API)	BA	JSON		<3569
Higher Education Degrees [DBA]	BA	CSV		797
Occupations (KldB) [DBA]	BA	CSV, XML	X	33,802
Continuing professional development [DBA]	BA	CSV		542
Continuing professional development [BS]	BIBB	CSV, PDF	X	218
Vocational Education [BS]	BIBB	CSV, PDF	X	357
Skills [DBA]	BA	XML		9078
Tools [54]	BIBB	CSV		10,978

- BERUFENET (https://github.com/bundesAPI/berufenet-api);
- KURSNET (https://andreasfischer1985.github.io/arbeitsagentur-apis/):
  - AUSBILDUNGSSUCHE (https://github.com/bundesAPI/ausbildungssuche-api);
  - WEITERBILDUNGSSUCHE (https://github.com/bundesAPI/weiterbildungssuche-api);
  - STUDIENSUCHE (https://github.com/bundesAPI/studiensuche-api);
  - SPRACHFÖRDERUNG (https://github.com/bundesAPI/berufssprachkurssuche-api);
  - COACHINGUNDAKTIVIERUNG (https://github.com/bundesAPI/coachingangebote-api).
- JOBSUCHE (https://github.com/bundesAPI/jobsuche-api);
- BEWERBERBÖRSE (https://github.com/bundesAPI/bewerberboerse-api);
- ENTGELTATLAS (https://github.com/bundesAPI/entgeltatlas-api);
- NEWPLAN (https://github.com/bundesAPI/newskills-api).

In the Federal Republic of Germany, the Vocational Training Act (BBiG) of 1969, which was reformed in 2005 and in 2020, was passed to create a political framework for shaping the work of vocational regulation [55]. If occupations are newly created or updated, several parties are involved: (a) the companies and chambers (employers), (b) the trade unions (employees), (c) the federal states, and (d) the federal government. Finally, the federal government provides the legal framework for vocational education and training through laws and regulations. The Federal Institute for Vocational Education and Training (BIBB), founded in 1970 on the basis of the Vocational Training Act (BBiG), provides the content of the training regulations online:

- Ausbildungsordnungen (https://www.bibb.de/dienst/berufesuche/de/index_berufesuche.php/) with classification (KldB 2010), statistical data, and a genealogy of occupations;
- Fortbildungsordnungen (https://www.bibb.de/dienst/berufesuche/de/index_berufesuche.php/; see Figure 3 for an example) with classification (KldB 2010) and statistical data;
- Berufesuche (https://www.bibb.de/de/40.php).

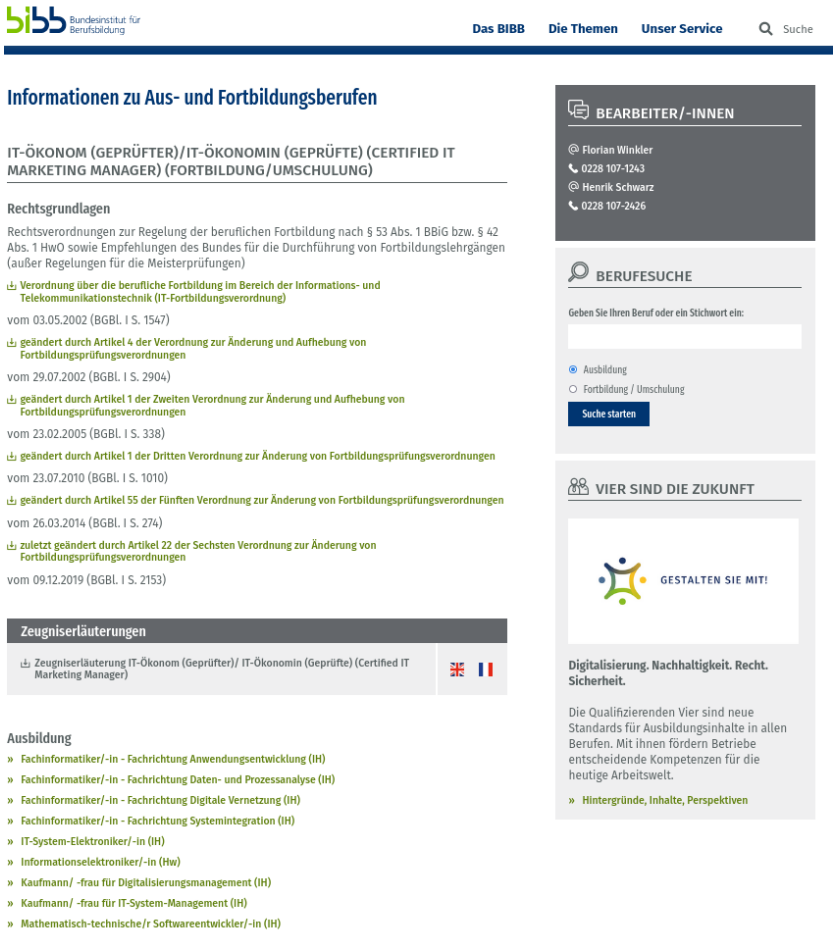

Figure 3. BIBB "Berufesuche" website for "IT-Economics (certified)", see https://www.bibb.de/dienst/berufesuche/de/index_berufesuche.php/profile/advanced_training/56tz67z8. It offers official regulations and some other information. Compare with the BA BERUFESUCHE website in Figure 1.

The official regulations contain many different types of documents. The main components of regulations are (a) an occupation title ("Bezeichnung des Ausbildungsberufes"), (b) the length of the program ("Ausbildungsdauer"), (c) the occupational skills, knowledge and abilities ("beruflichen Fertigkeiten, Kenntnisse und Fähigkeiten"), (d) the structure ("sachliche und zeitliche Gliederung"), and (e) the requirements ("Prüfungsanforderungen").

It seems noteworthy that these official regulations are not static as the genealogy of vocational education demonstrates (see Figure 3), and as a result, people trained in a deprecated education are available on the labor market. However, while the legal basis and advanced (vocational) training regulations are available for each point in time, they are not always available in a machine-readable form (see Figure 3, middle), and, what is more, the labor market and its demands are usually evolving much faster than regulations. Thus, we also want to make labor market data on occupations and CVET advertisements interoperable in order to find a ground truth concerning educational pathways.

The BIBB data are, therefore, mostly complementary to the BA data. However, at the intersection of the two sets of data is vocational education and training, particularly vocational and continuing training programs. The BA data also provide information on academic programs and could also provide further data on informal CVET programs.

3.3. Methods

We have compiled a complete list of official (continuing) professional development and training regulations from the BIBB database in order to derive a knowledge graph

of education pathways according to the BIBB, see Section 4.1. Similarly, we compiled a complete list of occupations as well as detailed information for each occupation on BERUFENET ($n = 3569$) via the documented API. From these data, we extracted occupation titles, BERUFENET IDs (numbers with three or more digits), KldB codes (five digits), DKZ codes (eight digits, preceded by the letter "B") as well as entry requirements and training opportunities for each occupation. From the information on entry requirements (entries under "Zugangsberufe/Zugangstätigkeiten") and training opportunities (entries under "Weiterbildung (beruflicher Aufstieg)"), we derived a knowledge graph of the education pathways according to the BA (consisting of the relation "is qualification for" between the nodes linked), in order to identify additional information (see Section 4.2).

To identify opportunities to link the data with other data sources, we took a closer look at selected entries in the databases of the BA. We extracted (a) information for the following selected occupations in AUSBILDUNGSSUCHE, WEITERBILDUNGSSUCHE, STUDIENSUCHE as well as in JOBSUCHE, ENTGELTATLAS, NEWPLAN, and BEWERBERBÖRSE:

- "Fachinformatiker/in—Daten- und Prozessanalyse"/KldB/DKZ-code "B 43112-919";
- "Data Scientist"/KldB/DKZ-code "B 43104-132".

And we extracted (b) information for exemplary entries in SPRACHFOERDERUNG and COACHINGUNDAKTIVIERUNG. On this basis, we inspected the data available and their interoperability (see Section 4.3).

With regard to web resources of the BIBB, we scraped the BIBB "Berufesuche" to obtain information on vocational education and training or continuing vocational education and training (in particular, Ausbildung, Fachpraktiker, Fortbildung/Umschulung, and Pflegeberufe). A list of all data entries could be obtained by appending all possible initial letters (and the letter sequence "xyz") to the URL one after the other. For instance, data entries for occupations starting with x, y, or z could be retrieved by the query https://www.bibb.de/dienst/berufesuche/de/index_berufesuche.php/alphabetical/apprenticeship/xyz. Each individual data entry has a particular ID (e.g., apprenticeship/8234101) and there are data for each ID on a second page (e.g., apprenticeship/8234101?page=2). From these data, we extracted basic information, e.g., KldB codes (five digits) and occupation titles (e.g., "Fachangestellter für Medien- und Informationsdienste/Fachangestellte für Medien- und Informationsdienste—Fachrichtung Bibliothek (Ausbildung)").

Data from different sources were linked via string matching of KldB/DKZ codes, occupation titles or similar data (see Figure 3). We extracted the KldB for each occupation on BERUFENET and related it to the corresponding classes in ESCO and ISCO-08 via look-up tables provided by the BA (e.g., the "Umsteigeschluessel-KldB2010-ISCO-08.xls", which can be found online under https://statistik.arbeitsagentur.de/DE/Statischer-Content/Grundlagen/Klassifikationen/Klassifikation-der-Berufe/KldB2010-Fassung2020/Arbeitsmittel/Umschluesselungstabellen.html).

4. Results

Figure 2 gives an overview of the data sources we considered and how they can be linked to each other via the classification of occupations (KldB).

4.1. BIBB Education Pathways

In this section, we describe the results of the web scraping official (continuing) professional development and training regulations of the BIBB. As described above, the BIBB offers information about regulations for 1004 (re-)trainings (CVET) that are regulated by enterprises or by the Crafts and Trade Code, 542 of which are regulated by the German state, and all regulations for vocational education in Germany. The latter include a KldB code (five digits) stating the occupation resulting from the education. CVET data additionally contain required qualifications for the education. Thus, the mapping from regulations and KldB codes can be obtained using string-matching. But this is, however, not the case for all CVET programs. For example, "AOK-Betriebswirt" (AOK business economist) is

special training only offered by AOK. It remains unclear if a generic mapping to a business economist meets the program.

In Figure 4, we present some results for IT professions. For instance, according to the data from web resources of the BIBB, a "Telecommunications Electronics Technician" can use the CVET program "Computer Scientist (Certified)", which leads to "Professions in Software Development—professionally oriented activities", which in turn gives opportunities for the four additional CVET programs: (1) "IT Project Manager (Certified)", leading to "Managers, IT Network Engineering, Coordination, Administration, Organization"; (2) "IT consultant (certified)", leading to "professions in IT application consulting"; (3) "IT economist (certified)" (see also Figure 3), leading to "professions in IT sales"; and (4) "IT developer (certified)", which finally leads to "occupations in IT coordination".

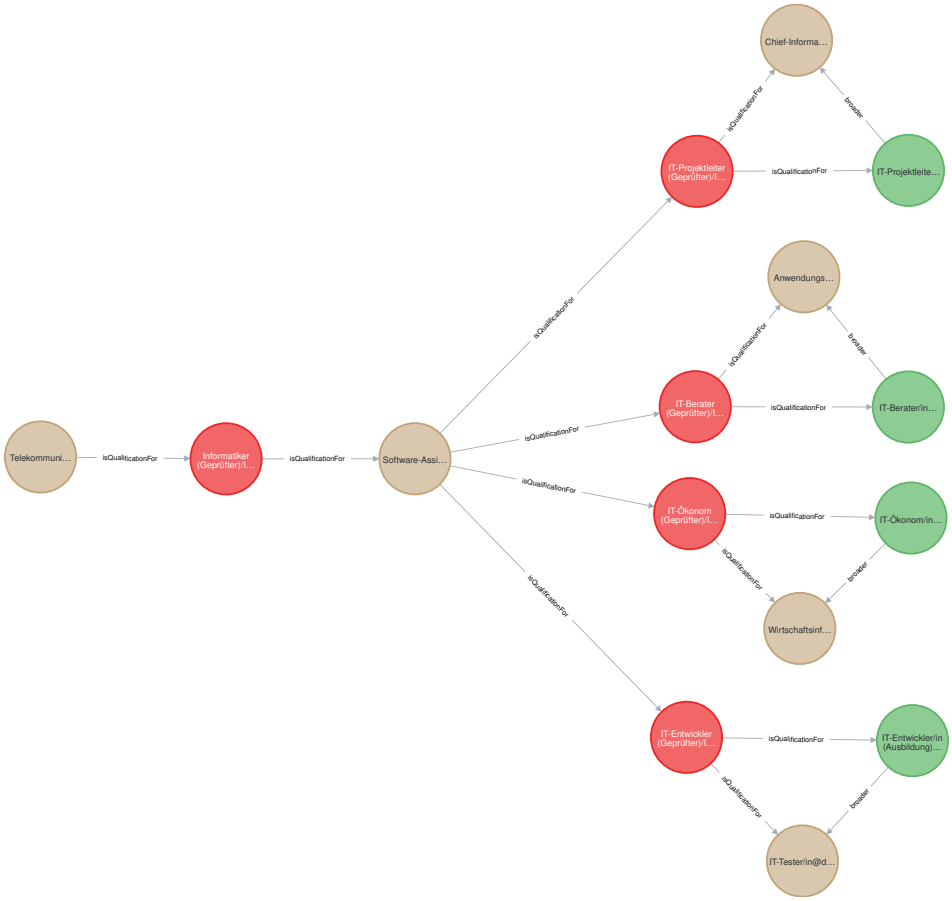

Figure 4. The pathway of a "Telecommunications Electronics Technician" via the CVET program "Computer Scientist (Certified)", which leads to "Professions in Software Development—professionally oriented activities", which in turn gives opportunities for the four additional CVET programs. Yellow and green nodes refer to occupations, and red nodes to CVET programs.

In part, the data analyzed in this section suggest some very complex education pathways (see Figure 5). For instance, on the top left of Figure 5, we find the occupations of beekeeping ("Berufe in der Imkerei"), which qualify for further education in nature and landscape conservation ("Berufe in der Natur- und Landschaftspflege") with several specializations like cemetery gardener and professions in nursery gardens. However, the data reported so far only reflect the official regulations and, therefore, may not reflect the realities of the labor or (further) education and training markets.

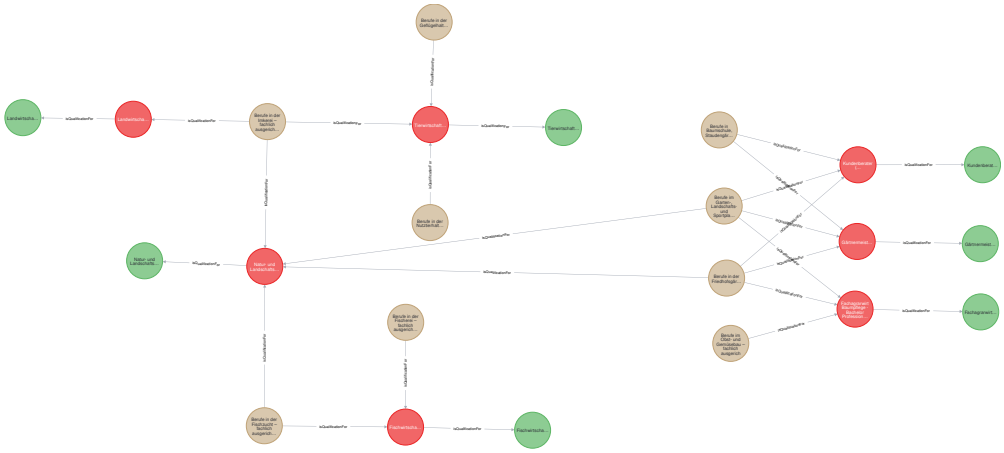

Figure 5. This graphic shows the complexity of career pathways by CVET programs. Yellow and green nodes refer to occupations, and red nodes to CVET programs.

4.2. BA Education Pathways

In this section, we show how data from the knowledge graph described in Section 4.1 can be extended by data from the BA. In this case, we derived education pathways from information on (a) entry requirements and (b) advancement courses listed for each occupation in BERUFENET. Figure 6 shows the results for an exemplary occupation, namely "IT-Economist (certified)" (BERUFENET-ID 15322, KldB/DKZ-code B 43233-105). As a rule, you need to have passed the exam as an IT economist in order to work as an IT economist, so BERUFENET lists "IT-Economist (certified)" (BERUFENET-ID 15323, KldB/DKZ-code B 43233-903) as an entry requirement for working in this occupation.

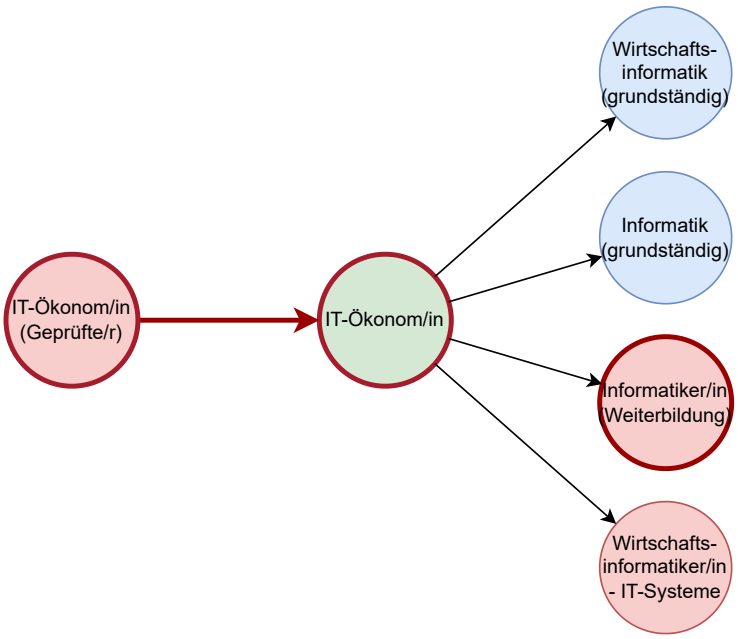

Figure 6. An extended version of the BIBB pathways shown in Figure 4, focusing on the leaf "IT Economist (certified)". All outgoing links are scraped from BA BERUFENET. All nodes with a dark red line are also included in the BIBB data, in particular, also the CVET program "Computer scientist (certified)". However, we find new links and, in particular, study programs (blue nodes). Green nodes refer to occupations, and red nodes to CVET programs.

Additionally, BERUFENET lists four opportunities for advancement, see Figure 6, each with a short occupation title and BERUFENET ID.

In general, on BERUFENET, each occupation is related to (a) measures of further educational training (including study programs), and to (b) measures of adaptation training. For instance, besides the four measures of further educational training in Figure 6, the occupation "IT-Economist (certified)" is additionally related to seven measures of adaptation training on WEITERBILDUNGSSUCHE:

- "IT Project Management";
- "IT Service Management; IT Infrastructure Library (ITIL)";
- "Marketing";
- "Sales";
- "Controlling";
- "Business organization, work study";
- "Employee leadership, teamwork, leadership".

Each measure is stored with an ID that represents an education goal (which can be used as a value for parameter "bildungsziel" or "oberknoten" in the WEITERBILDUNGSSUCHE to query current vocational training offers). For instance, the first entry in this list has "ID" 122937, which represents offers of adaptation training with regard to IT project management in the WEITERBILDUNGSSUCHE:

- https://web.arbeitsagentur.de/weiterbildungssuche/suche?bildungsziel=122937.

Figure 6 shows an example of the benefit of extending the BIBB education pathways based on education pathways derived from BA data. Four observations are noteworthy: First, more dependent occupations and trainings are added to the existing data (e.g., for "Computer scientist (certified)"). Second, it adds a few more continuing training programs that are not regulated at the federal level. Third, it adds study programs. Fourth, it shows a complex network of education rather than a network of mainly pathways.

In addition, there is a great deal of further information that could be used to further enrich the above-mentioned knowledge graph. For instance, each occupation on BERUFENET is related to a set of competences and related chunks of knowledge and skills—e.g., for the "IT-Economist (certified)", it lists the following set of core competences:

- "Acquisition";
- "Business administration";
- "Controlling";
- "Information technology, computer technology";
- "IT coordination";
- "IT organization";
- "Calculation";
- "Cost and performance accounting";
- "Customer service, care";
- "Marketing",
- "Human resource".

An allocation to various taxonomies can also be established via KldB/DKZ codes (8-digits). For instance, with regard to European and international frameworks of classification, the "IT-Economis (certified)" can be considered a narrow match to ESCO-occupation "ICT business development manager" (ESCO-code 2434.2) and automatically classified as ISCO Unit group 2434 "Information And Communications Technology Sales Professionals" based on its KldB/DKZ-code B 43233-105.

4.3. Interoperability

In building and extending the knowledge graph based on multiple data sources, we gained several insights:

- The most efficient way of relating information from BERUFENET to other data sources or to classification systems seems to be the KldB/DKZ codes (eight digits), which are stored in BERUFENET and many other data sources (e.g., AUSBILDUNGSSUCHE; see Figure 1); data sources of the BA that do not contain a KldB/DKZ code can often

be related to a KldB/DKZ code by matching short occupation titles (although short occupation titles, unlike BERUFENET IDs or the eight-digit variants of KldB/DKZ codes do not differ for training and for the occupational activity).

- Results of the JOBSUCHE have an attribute "beruf" that contains occupational titles that could be matched with BERUFENET' short occupation titles; the JOBSUCHE API does not seem to provide KldB-/DKZ-codes.
- Results of the AUSBILDUNGSSUCHE have an attribute "abschlussbezeichnung" that contains training job designations that could (after removing HTML tags) be matched with BERUFENET's short occupation titles.
- Results of the STUDIENSUCHE have an attribute "Studienfaecher", which contains one or more course designations that could be matched with BERUFENET's short occupation titles.

In addition, entries in the BEWERBERBÖRSE do provide an attribute "berufe" that can be matched with the short occupation titles of the BERUFENET API. It is also possible to query corresponding results using the "was" parameter and setting it to short occupation titles. The APIs for ENTGELTATLAS and NEWPLAN show the KldB/DKZ code in the results and allow for requesting results based on KldB/DKZ codes via parameters. The APIs of SPRACHFÖRDERUNG and COACHINGUNDAKTIVIERUNG have an attribute "systematiken" but do not contain theBERUFENET's short occupation titles or KldB entries.

5. Conclusions and Outlook

Labor markets heavily rely on vocational education and training, re-training and advanced vocational qualification. In this paper, we inspected different sources of data and data schemata to explore the interconnectivity of data on the job market in Germany and to derive knowledge on learning pathways from information on the relation between different jobs and occupations. In order to structure the discussion of our main findings, we would like to take up the two research questions that we posed in the introduction:

Our first research question (RQ1) was how to derive knowledge about educational pathways from data on entry requirements and training opportunities. We have found that knowledge about a complex variety of possible educational pathways can be derived from BA and BIBB data, and that linking different data sources can reveal pathways that complement an examination of individual data sources well—e.g., we were able to extend our knowledge tree on education pathways, which was derived from the BIBB information about professional development and training regulations, based on the knowledge graphs we derived from BERUFENET of the BA. As each occupation on BERUFENET can be related to a KldB/DKZ code, knowledge trees can easily be extended by adding further kinds of data points available on BERUFENET (e.g., competences, skills and knowledge) or from different data sources (as we demonstrated in Section 4.2).

Our second research question (RQ2) was how different kinds of data and classifications could be related to each other for data, entity, event, and relationship extraction from German online labor market resources. In general, eight-digit KldB/DKZ codes seemed to be the most reliable way of relating data on occupations between different data sources. Short occupation titles also worked well, at least for data from a single data provider (i.e., within data from the BA), but it seems noteworthy that (a) sources from different providers differed in spelling details such as gender, and (b) the difference between training and occupational activity was found in the eight-digit variants of KldB/DKZ codes but not in the occupational titles of BERUFENET (e.g., "IT-Economist"). It should also be noted that the eight-digit KldB/DKZ code is probably less stable than the five-digit KldB and may be subject to change (cf. [49]). In this regard, a similarity-based classification based on sentence embeddings of occupational titles by large language models (e.g., [2,31]) may be a promising alternative to simple string matching for many use cases.

In addition, it was possible to relate German occupations to European and international classification frameworks, but not at the level of individual occupations, which resulted in an inherent fuzziness: the ISCO-08 classification is not designed for the occupational level,

and even the ESCO classification provided only approximate results for many occupations. This implies that international education research is still very much tied to an aggregate level, although sentence embeddings by large language models (cf., [2,31]) could allow for a classification at the level of individual occupations in future studies.

In summary, it seems possible to find the ground truth by linking different data sources on the labor market and on (further) vocational education and training, but the data include many domain-specific aspects, and the relationships to existing occupations are not always clear. For example, many offers of further education in the WEITERBILDUNGSSUCHE of the BA are not linked to occupations in BERUFENET.

Looking to the future, many more data sources could be included to create knowledge graphs about careers and vocational (further) education. For the contribution at hand, we focused primarily on two official data sources. To obtain a more comprehensive overview, it would be interesting to include additional data sources, such as those from chambers of crafts or trade, as well as additional sources of CVET advertisements. In this respect, it should be noted that the web resources studied in our work contain a lot of structured data. This cannot be taken for granted when analyzing a wider range of job portals such as Monster.com, StepStone, or Academics, or other data portals such as kununu or Glassdoor.

Future work will, therefore, need to spend a considerable amount of effort on linking unstructured textual information in order to obtain a complete picture of the labor market. Thus, optimizing and extending ontologies such as GLMO could be a promising direction of research in this area. In addition to ontological aspects and some of the above-mentioned innovative ways of linking data sources, there are other research directions to consider: For example, how do these data relate to other sources such as data from chambers of trade or craft? Can statistical data be used to extend the network of educational pathways? Is it possible to model the movement of labor to other occupations?

Many data sources on the labor market can be accessed online, and APIs are well documented thanks in part to civil society initiatives such as bund.dev. Linking between different sources is possible via common classification systems such as KldB/DKZ or via occupation titles. From an educational science perspective, it would be desirable for the data to be linked by researchers or to be directly provided as Linked Open Data (LOD) by the BA or the BIBB. This would lower the barriers for many scientists and make research easier as well as less error-prone. It is worth noting that first steps in this direction have already been taken by the BA, which provides explicit links from BERUFENET occupations to a subset of related education advertisements in WEITERBILDUNGSSUCHE. Such steps should be extended and applied to other APIs as well.

Author Contributions: Conceptualization, A.F. and J.D.; methodology, A.F. and J.D.; software, A.F. and J.D.; validation, A.F. and J.D.; formal analysis, A.F. and J.D.; investigation, A.F. and J.D.; resources, A.F. and J.D.; data curation, A.F. and J.D.; writing—original draft preparation, A.F. and J.D.; writing—review and editing, A.F. and J.D.; visualization, A.F. and J.D.; supervision, A.F. and J.D.; project administration, A.F. and J.D.; funding acquisition, A.F. and J.D. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding. This article was funded by the Open Access Publication Fund of the Federal Institute for Vocational Education and Training (BIBB), Bonn.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The German Labor Market Ontology (GLMO) is available at http://w3id.org/glmo.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

BA	German Federal Employment Agency (Bundesagentur für Arbeit)
BIBB	Federal Institute for Vocational Education and Training (Bundesinstitut für Berufsbildung)
CVET	Continuing Vocational Education and Training
DKZ	Documentation number, "Dokumentationskennziffer"
ESCO	European Skills/Competences, Qualifications and Occupations
GLMO	German Labor Market Ontology
ISCO	International Standard Classification of Occupations
KldB	German Classification of Occupations, "Klassifikation der Berufe"
KSAO	Knowledge, Skills, Abilities, and Other components
OJA	Online Job Advertisements
VET	Vocational Education and Training

References

1. Fischer, A.; Hecker, K.; Wittig, W. *Arbeitsmarktbedarfsanalyse zu beruflichen Kompetenzen und Teilqualifikationen: Eine repräsentative Unternehmensbefragung*; Forschungsinstitut Betriebliche Bildung (F-BB): Nürnberg, Germany, 2020.
2. Fischer, A.; Jöchner, A.; Pabst, C.; Lorenz, S.; Schley, T. *KI-Basierte Personalisierung Berufsbezogener Weiterbildung: Ein Praxisleitfaden für Bildungsanbieter*; wbv-Verlag: Nürnberg, Germany, 2023.
3. *ESCO Handbook: European Skills, Competences, Qualifications and Occupations*, 2nd ed.; Publication Office of the European Union: Luxembourg, 2019. [CrossRef]
4. Helmrich, R.; Tiemann, M.; Troltsch, K.; Lukowski, F.; Neuber-Pohl, C.; Lewalder, A.C.; Gunturk-Kuhl, B. *Digitalisierung der Arbeitslandschaften: Keine Polarisierung der Arbeitswelt, aber beschleunigter Strukturwandel und Arbeitsplatzwechsel*; Wissenschaftliche Diskussionspapiere; Federal Institute for Vocational Education and Training (BIBB): Bonn, Germany, 2016; Number 180.
5. Fischer, A.; Hilse, P.; Schütt-Sayed, S. Rahmenlehrpläne–Spiegel der Bedeutung nachhaltiger Entwicklung. In *Zum Konzept der Nachhaltigkeit in Arbeit, Beruf und Bildung—Stand in Forschung und Praxis*; Barbara Budrich: Nürnberg, Germany, 2023; pp. 281–302.
6. Graf, L.; Lohse, A.P. Advanced skill formation between vocationalization and academization: The governance of professional schools and dual study programmes in Germany. In *Governance Revisited: Challenges and Opportunities for Vocational Education and Training*; Gonon, P., Bürgi, R., Eds.; Peter Lang Group AG: Lausanne, Switzerland, 2021.
7. Dikau, J. Rechtliche und organisatorische Bedingungen der beruflichen Weiterbildung. In *Handbuch der Berufsbildung*; Springer: Wiesbaden, Germany, 1995; pp. 427–440.
8. Bauer, R.; Bauer, R. Die Debatte über die Zukunft der dualen Berufsausbildung. In *Verberuflichung von Weiterbildung und die Zukunft der Dualen Berufsausbildung: Eine Berufssoziologische Analyse am Beispiel des Kraftfahrzeuggewerbes*; Springer: Wiesbaden, Germany, 2000; pp. 21–84.
9. Dutt, A.; Ismail, M.A.; Herawan, T. A systematic review on educational data mining. *IEEE Access* **2017**, *5*, 15991–16005. [CrossRef]
10. Mohamad, S.K.; Tasir, Z. Educational data mining: A review. *Procedia-Soc. Behav. Sci.* **2013**, *97*, 320–324. [CrossRef]
11. Romero, C.; Ventura, S. Educational data mining: A survey from 1995 to 2005. *Expert Syst. Appl.* **2007**, *33*, 135–146. [CrossRef]
12. Kovalev, S.; Kolodenkova, A.; Muntyan, E. Educational data mining: Current problems and solutions. In Proceedings of the 2020 V International Conference on Information Technologies in Engineering Education (Inforino), Moscow, Russia, 14–17 April 2020; pp. 1–5.
13. Marlis, B.; Buchs, H.; Ann-Sophie, G. Occupational Inequality in Wage Returns to Employer Demand for Types of Information and Communications Technology (ICT) Skills: 1991–2017. *Kölner Z. Soziologie Sozialpsychologie* **2020**, *72*, 455–482.
14. Settelmeyer, A.; Bremser, F.; Lewalder, A.C. Migrationsbedingte Mehrsprachigkeit—Ein „Plus" beim Übergang von der Schule in den Beruf. In *Interkulturelle und Sprachliche Bildung im Mehrsprachigen Übergang Schule-Beruf*; Waxman: Münster, Germany, 2017; pp. 135–150.
15. 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* **2020**, *15*, e0233746. [CrossRef] [PubMed]
16. Smirnov, I. Estimating educational outcomes from students' short texts on social media. *EPJ Data Sci.* **2020**, *9*, 27. [CrossRef]
17. Ortmann, T.T.; Bönke, D.H.; Hammer, L. *Bessere Perspektiven bei Jobwechseln. Zur Ähnlichkeit Beruflicher Übergänge*; Gieselmann: Gütersloh, Germany, 2023.
18. Degenhardt, S. Kompetenzen für eine digitalisierte Arbeitswelt–Anforderungen an Aus-und Weiterbildung. In *Digitaler Wandel in der Sozialwirtschaft*; Nomos Verlagsgesellschaft mbH & Co.: Baden-Baden, Germany, 2018; pp. 259–272.
19. Kreuzer, C. Visualisierung der Opportunity Recognition-Kompetenz von Industriekaufleuten. *Z. Berufs Wirtsch.* **2018**, *114*, 247–271. [CrossRef]
20. Beręsewicz, M.; Pater, R. *Inferring Job Vacancies from Online Job Advertisements*; Publications Office of the European Union: Luxembourg, 2021.

21. Khaouja, I.; Kassou, I.; Ghogho, M. A survey on skill identification from online job ads. *IEEE Access* **2021**, *9*, 118134–118153. [CrossRef]
22. Carnevale, A.P.; Jayasundera, T.; Repnikov, D. *Understanding Online Job Ads Data*; Technical Report; Center on Education and the Workforce, Georgetown University: Washington, DC, USA, 2014.
23. Ros, R.; Van Erp, M.; Rijpma, A.; Zijdeman, R. 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*; Association for Computational Linguistics: Stroudsburg, PA, USA, 2020; pp. 27–32.
24. 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*; Association for Computational Linguistics: Stroudsburg, PA, USA, 2020; pp. 83–93.
25. Buchmann, M.; Buchs, H.; Busch, F.; Clematide, S.; Gnehm, A.S.; Müller, J. Swiss Job Market Monitor: A Rich Source of Demand-Side Micro Data of the Labour Market. *Eur. Sociol. Rev.* **2022**, *38*, 1001–1014. [CrossRef]
26. Hermes, J.; Schandock, M. Stellenanzeigenanalyse in der Qualifikationsentwicklungsforschung. In *Die Nutzung Maschineller Lernverfahren zur Klassifikation von Textabschnitten*; Forschungsinstitut Betriebliche Bildung (F-BB): Nürnberg, Germany, 2016.
27. Ziegler, M.; Horstmann, K.; Wehner, C. *Machbarkeitsstudie: Teilqualifikationen in Online-Jobanzeigen (OJA)*; Humboldt-Universität zu Berlin: Berlin, Germany, 2022.
28. Janser, M. *The Greening of Jobs in Germany: First Evidence from a Text Mining Based Index and Employment Register Data*; Technical report, IAB-Discussion Paper; Institut für Arbeitsmarkt- und Berufsforschung (IAB): Nürnberg, Germany, 2018.
29. Binnewitt, J.; Schnepf, T. Join us to turn the wor(l)d greener!—Investigating online apprenticeship advertisements' reference to environmental sustainability. In *Zum Konzept der Nachhaltigkeit in Arbeit, Beruf und Bildung—Stand in Forschung und Praxis*; Federal Institute for Vocational Education and Training (BIBB): Bonn, Germany, 2022.
30. Ziegler, P. *Zur Verwendung von Berufsinformation im Hinblick auf Matching in Deutschland und Österreich*; Technical report, AMS Info; Leibniz Information Centre for Economics: Hamburg, Germany, 2012.
31. Li, N.; Kang, B.; De Bie, T. SkillGPT: A RESTful API service for skill extraction and standardization using a Large Language Model. *arXiv* **2023**, arXiv:2304.11060.
32. Bhola, A.; Halder, K.; Prasad, A.; Kan, M.Y. Retrieving skills from job descriptions: A language model based extreme multi-label classification framework. In Proceedings of the 28th International Conference on Computational Linguistics, Online, 9–13 December 2020; pp. 5832–5842.
33. Khaouja, I.; Mezzour, G.; Carley, K.M.; Kassou, I. Building a soft skill taxonomy from job openings. *Soc. Netw. Anal. Min.* **2019**, *9*, 1–19. [CrossRef]
34. International Labour Office. *The Feasibility of Using Big Data in Anticipating and Matching Skills Needs*; International Labour Office: Geneva, Switzerland, 2020.
35. Stops, M.; Bächmann, A.C.; Glassner, R.; Janser, M.; Matthes, B.; Metzger, L.J.; Müller, C.; Seitz, J. *Machbarkeitsstudie Kompetenz-Kompass: Teilprojekt 2: Beobachtung von Kompetenzanforderungen in Stellenangeboten*; Bundesministerium für Arbeit und Soziales: Berlin, Germany, 2020.
36. Fischer, A.; Neubert, J.C. The multiple faces of complex problems: A model of problem solving competency and its implications for training and assessment. *J. Dyn. Decis. Mak.* **2015**, *1*, 6.
37. Fischer, A.; Hecker, K.; Pfeiffer, I. Berufliche Kompetenzen von Geflüchteten erkennen? Exemplarische Befunde zur Kompetenzmessung im Bereich der Metallbearbeitung und Metallverarbeitung. *Z. Weiterbildungsforschung* **2019**, *42*, 115–131. [CrossRef]
38. Bundesagentur für Arbeit. *Band 1: Systematischer und Alphabetischer Teil mit Erläuterungen*; Bundesagentur für Arbeit: Nuremberg, Germany, 2010.
39. Paulus, W.; Matthes, B. The German classification of occupations 2010–structure, coding and conversion table. *FDZ-Methodenreport* **2013**, *8*, 2013.
40. Dörpinghaus, J.; Binnewitt, J.; Hein, K. Lessons from Continuing Vocational Training Courses for Computer Science Education. In Proceedings of the ITiCSE 2023: Innovation and Technology in Computer Science Education, Turku, Finland, 7–12 July 2023; p. 636.
41. Dörpinghaus, J.; Binnewitt, J.; Winnige, S.; Hein, K.; Krüger, K. Towards a German labor market ontology: Challenges and applications. *Appl. Ontol.* **2023**, *18*, 343–365. [CrossRef]
42. Dörpinghaus, J.; Samray, D.; Helmrich, R. Challenges of Automated Identification of Access to Education and Training in Germany. *Information* **2023**, *14*, 524. [CrossRef]
43. Fechner, R.; Dörpinghaus, J.; Firll, A. Classifying Industrial Sectors from German Textual Data with a Domain Adapted Transformer. In Proceedings of the 2023 18th Conference on Computer Science and Intelligence Systems (FedCSIS), Warsaw, Poland, 17–20 September 2023; pp. 463–470.
44. le Vrang, M.; Papantoniou, A.; Pauwels, E.; Fannes, P.; Vandensteen, D.; De Smedt, J. Esco: Boosting job matching in europe with semantic interoperability. *Computer* **2014**, *47*, 57–64. [CrossRef]
45. 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 Proceedings of 14th International Conference, MTSR 2020, Madrid, Spain, 2–4 December 2020; pp. 116–122.

46. Kitto, K.; Sarathy, N.; Gromov, A.; Liu, M.; Musial, K.; Buckingham Shum, S. Towards skills-based curriculum analytics: Can we automate the recognition of prior learning? In Proceedings of the LAK '20: 10th International Conference on Learning Analytics and Knowledge, Frankfurt, Germany, 23–27 March 2020; pp. 171–180.
47. Fareri, S.; Melluso, N.; Chiarello, F.; Fantoni, G. SkillNER: Mining and mapping soft skills from any text. *Expert Syst. Appl.* **2021**, *184*, 115544. [CrossRef]
48. Neutel, S.; de Boer, M.H. Towards Automatic Ontology Alignment using BERT. In Proceedings of the AAAI 2021 Spring Symposium on Combining Machine Learning and Knowledge Engineering (AAAI-MAKE 2021), Palo Alto, CA, USA, 22–24 March 2021.
49. Fischer, A. Toot 111039750735796601 on Chaos.Social. Technical Report, 2023. Available online: https://chaos.social/@AFischer1985/111039750735796601 (accessed on 1 January 2024).
50. Schimpl-Neimanns, B. *Mikrodaten-Tools: Umsetzung der Berufsklassifikation von Blossfeld auf die Mikrozensen 1973–1998*; GESIS—Leibniz-Institut für Sozialwissenschaften: Mannheim, Germany, 2003.
51. Brauns, H.; Steinmann, S.; Haun, D. Die Konstruktion des Klassenschemas nach Erikson, Goldthorpe und Portocarero (EGP) am Beispiel nationaler Datenquellen aus Deutschland, Großbritannien und Frankreich. *Zuma Nachrichten* **2000**, *24*, 8–63.
52. Ganzeboom, H. Questions and Answers about ISEI-08. *Stand* **2010**, *13*, 2016.
53. Ganzeboom, H.B.; Treiman, D.J. Internationally comparable measures of occupational status for the 1988 International Standard Classification of Occupations. *Soc. Sci. Res.* **1996**, *25*, 201–239. [CrossRef]
54. Güntürk-Kuhl, B. *Die Taxonomie der Arbeitsmittel des BIBB*; Federal Institute for Vocational Education and Training (BIBB): Bonn, Germany, 2017.
55. Kuppe, A.M.; Lorig, B.; Schwarz, H.; Stöhr, A. *Ausbildungsordnungen und wie sie Entstehen*; Federal Institute for Vocational Education and Training (BIBB): Bonn, Germany, 2015.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.