Proceedings of the 19th Conference on Computer Science and Intelligence Systems (FedCSIS) pp. 355–362
DOI: 10.15439/2024F3577
ISSN 2300-5963 ACSIS, Vol. 39

Linked Labor Market Data: Towards a novel data housing strategy

Kristine Hein*
* Federal Institute for Vocational Education and Training (BIBB), Bonn, Germany

***Abstract*—The labor market is a domain rich in diverse data structures, both quantitative and qualitative, and numerous applications. This leads to challenges in the domain of data warehouse architecture and linked data. In this context, only a few approaches exist to generate linked data sets. For example, the multilingual classification system of European Skills, Competences, Qualifications, and Occupations (ESCO) and the German Labor Market Ontology (GLMO) serve as prominent examples showcasing the pivotal role of ontologies.**

This paper introduces an initial conceptualization and proof-of-concept for managing interoperable German labor market data, including qualitative and quantitative data, such as surveys and statistical data, as well as textual data, such as social media data or online job advertisements. Additionally, it presents a data management perspective on the research network infrastructure, with a particular focus on the challenges encountered when establishing a data warehouse architecture within the field of education management. In this context, vocational training research offers a unique opportunity to anticipate future developments in the education and training markets. To this end, however, a fast and qualitatively good analysis option must be created to meet the demands of our fast-paced modern world. This is why a novel automated data strategy is required to facilitate the accelerated automation of processes, including ETL and the utilisation of contemporary data stacks.

I. Introduction

LABOR markets are dynamic entities, shaped by political and technical innovations, as well as changes in society. These shifts result in new demands, requirements, and novel data. In this context, vocational education and training, as well as re-training, play a pivotal role in meeting these new demands [1], [2]. However, labor market research is also an area with a limited amount of available resources [3]. It encompasses a multitude of data structures, each serving a distinct purpose. These include the facilitation of connections between job seekers and the most suitable training or employment opportunities. However, the sole application of semantic web technologies in this domain is the multilingual classification of European Skills, Competences, Qualifications and Occupations (ESCO). This exemplifies the pivotal role of ontologies in this field, see [4], [5].

We will now proceed to discuss the manifold problems and the problem statements. Following this, we will discuss the research questions of this work.

A. Problem statement

The sheer volume of data is overwhelming. While data is collected by all, there is no standardised structure for its collection, format or preparation, let alone documentation. Gaining an overview of the individual data collections is challenging, and it is important to first cluster them and understand them at different levels. The field of research encompasses a variety of data types, which, in the context of survey data, can be broadly classified into three categories: survey data, process data, and structural data.

In the context of survey data, it is important to distinguish between qualitative and quantitative data. In the field of data science, the format of the data is initially distinguished from its content. This distinction is made, for instance, between text, image, video data, or formats such as XML and JSON. Furthermore, there is the realm of relational data (databases) in which standardised formats can be searched using a variety of search tools. These formats are designed with the objective of facilitating rapid searches and indexing.

Examples of data include survey data such as employment surveys, economic surveys such as the microcensus, job advertisement data, as well as already aggregated and harmonised data from economic structural research. Furthermore, data from social media, digitised archive data, structured data from education and training portals, market research data, and other sources are also utilised. Furthermore, qualitative data from internal BIBB surveys is also available. The data can be observed over time. The data is collected and updated at regular intervals, ranging from daily to once every six to ten years.

In the context of vocational training data, the occupation represents the common denominator of all data sets. However, some datasets can only be linked to each other via sectors, as surveys of companies (as a survey unit) and not of individuals were conducted.

Another challenge that arises in the context of surveys is the data protection of personal data. The person must have given their consent for the processing of GDPR-relevant data. As this consent is earmarked for a specific purpose, in some cases it is not possible to link the raw data at all. However, possibilities of abstraction and anonymisation can help here in order to still be able to work with the data collected. As a rule, the data volumes can be (artificially) increased by clever clustering so that conclusions cannot be drawn. Further work can take the direction of data boosting (see bootstrapping procedure [6]). In addition to these two defining characteristics, other common parameters of the datasets can be identified in certain instances, such as mapping to regions such as federal states, federal regions, cities/municipalities, districts, and so forth.

In addition to the amount of different data, there are also very different stakeholders at different levels of the data

warehouse. Standardisation of the input data. Uniqueness of the data records, storage in accordance with FAIR Data, manipulation- and access-protected. Ensure a minimalist principle for forwarding the data, and yet disclose it sufficiently so that the full scope of the data can be recognised, so that scientists have a comprehensive insight into and overview of the data.

Another challenge is the necessity of dealing with detailed data in special occupational areas or data gaps in other data sets. In such instances, estimates may be employed, provided that they are documented and labelled. In summary, the identification of 'fuzzy' data and estimates is to be achieved through the implementation of different data modelling techniques, which must be documented and disclosed in accordance with the relevant data pipelines. This is intended to ensure the traceability and reproducibility of the data by other scientists.

B. Research questions

In order to address the multitude of issues that have been identified, we will commence with three preliminary research inquiries.

1) What are the most effective methods for integrating labor market data into a data warehouse system?
2) What difficulties are encountered when conducting quantitative labor market research?
3) What specific challenges arise in the analysis of labor market research data, particularly in the context of historical, qualitative, and quantitative data?

The initial research question is relatively broad in scope, whereas the subsequent inquiries are more specific to the field of computational social sciences, with a particular focus on labor market research.

This paper is divided into six sections. The introductory section provides an overview of the subject matter and its relevance to the field. The second section offers a concise analysis of the current state of the art and related work. The third section delineates the methodological approach employed in this study. The fourth section presents the results and an evaluation of the approach. The final section presents the conclusions and offers a prospective outlook.

II. Background and Literature Review

Over the past decades, we find a growing interest in mining data from labor market data, educational databases, and information systems, see for example [7], [8], [9], [10], [3]. These studies have highlighted the importance of supporting decision-making and process management in labor market research. Data warehousing is a frequently employed methodology in the computational social sciences and big data pipelines [11], [12]. The generic challenges are typically the automated extraction of knowledge from data, which is usually interpreted passages from texts, and the mapping to existing data sets. However, there are still several challenges related to data and data integration. The research questions addressed in this field of study are diverse, encompassing topics such as occupational inequality [13], [14], migration and language skills [15], sustainability [16], discrimination [17], and students and later occupation [18].

The situation in German-speaking countries (Germany, Austria, and Switzerland) with regard to automated analysis of labor market data is not significantly different from that in English-speaking countries. As stated in [19], "Catalogs play a valuable role in providing a standardized language for the activities people perform in the labor market." While these catalogs are widely used to create and compute static values, manage labor market and educational needs, or recommend training and jobs, there is no single ground truth. According to Rodrigues (2021), one reason for this could be the fact that labor market concepts are modeled by multiple disciplines, each with a different perspective on the labor market. While there have been discussions about mapping between different standards, such as the European ESCO and the American O*NET [20], there are only limited mapping approaches between standards to date. For instance, there is no mapping between the German KldB and the Austrian AMS (we will discuss this later). This is the first gap. While there is a diverse field of different taxonomies, catalogs, and even word lists used in different institutions and for different research questions, existing tools tend to focus on only one of these perspectives, making more generic solutions difficult to implement.

For data integration, the necessity for more generic models has been discussed in the field of education, see Szabo et al. [21]. Ontologies and ontology-based methodologies have been extensively utilized. For instance, for the prediction and modeling of workshops and labor market needs, see [22], for the identification of job knowledge, see [23], but also for the analysis of particular jobs, see [24], or for the matching of educational content to generic texts, see [25]. Furthermore, these ontologies have been employed to predict the unemployment rate, as evidenced by the work of Li et al.[26]. However, these approaches have primarily concentrated on a specific labor market characteristic, such as skills, knowledge, educational content, or job classifications.

According to our best knowledge, no data warehouse or linked-data approach for labor market data has yet been proposed although some preliminary work was carried out [27], [28], [5], [29]. Thus, here we find the first gap for interdisciplinary research. In addition, we find only few works addressing the specific challenges in the analysis of labor market research data, particularly in the context of historical, qualitative, and quantitative data. Some work was carried out in the area of online social media data [30].

III. Method

In order to address the manifold challenges and dependencies inherent to this interdisciplinary research and data infrastructure area, it is necessary to employ a bundle of different methodological approaches. As previously discussed, despite the technical challenges that must be overcome, it is essential to address several domain-specific and research-specific questions. First and foremost, it is imperative to

house qualitative and quantitative data from a diverse range of sources, including NLP, classical social science surveys, labor market statistics, and social media data, among others. Second, we must ensure that the data is interoperable and queryable despite the lack of an overarching ontology or taxonomy for all data. Third, the data being subject to interoperability is not aggregated at the same level. For instance, we have occupations for vocational training and others linked to other domain-specific entities. Other data is linked to occupational groups, while others are linked to occupations. Fourth, quantitative data also comes with a rich assortment of metadata that describes processes and structures. These data are of great importance for the classical scientific approach in the social sciences and labor market research. Sixth, the data is often not only stored as raw data, but also in different aggregation levels, which are subject to domain-specific requirements and usually not interoperable between different data sets.

In addition to these domain-specific requirements, classical technical problems need to be solved. For example, long-term storage is needed, data protection and security needs to be addressed.

It is necessary to retrieve data from a variety of sources, including applications that retrieve data from different application programming interfaces (APIs), data analysts who work with data portals, and researchers with different privileges. It is evident that no single solution will be universally applicable; however, we will provide a comprehensive discussion of future challenges.

To tackle these these remaining interdisciplinary challenges, we will provide some methodological ideas and discuss their impact within this complex setting.

A. Structure

The data warehouse (DWH) is comprised of two primary components, see Figure 1 for an illustration. The data archive, in this instance, is the data lake, serving as the initial point of reference. The data derived from this archive is subsequently integrated into the data warehouse in a structured format.

Today, modern data stacks are mainly used with standardised connectors to operationalise ETL processes. Transformation and orchestration are usually performed by standard tools. However, due to the heterogeneity of vocational training data, custom connectors must be developed that prepare the respective data records, adapt formats, cluster and cleanse data and ensure automated data integration in such a way that they can be further used in the DWH.

The overall integration, preparation and cleansing process occurs in level 0, which is the basementof the DWH. With regard to the content of the data, it is necessary to differentiate it into three categories: raw data, L0 indicators (aggregated data that has already been processed by external systems) and a classification system. Furthermore, the documentation of the metadata is conducted in parallel. This is the point at which the expertise of domain experts is applied to the documentation.

The data is linked at the second level. As previously outlined, the GLMO classification system employs a multi-faceted approach, encompassing a range of factors, including occupational classification (Kldb, ISCO), differentiation between gainful employment and training occupations, competence assessment (ESCO, AMS, BIBB Comp), economic indicators, and regional analysis. all taking into account the temporal course, possibly necessary anonymisation clustering (e.g. formation of occupational groups, main groups see [31]. In terms of content, the data and documentation level is of particular importance here, particularly in consideration of the findings of data analysts.

At the third level, the linked data is prepared and, if necessary, enriched, formatted for data reports or the dashboard or a data portal. In principle, this is the business intelligence (BI) level before the data is exported.

Prior to its dissemination, the data must be subjected to a final verification process and, if necessary, anonymised. At this stage, it is of particular importance to conduct plausibility and format tests. Furthermore, additional test pipelines can be developed to monitor the data analysis process.

At this level, the L2 data must be prepared by BI experts in such a way that the findings about the data discovered in Level 1 are appropriately represented, prepared and documented for the respective stakeholders. These include scientists, the CEO, and other standard users who should have access to the data, for example, in order to develop a data portal for young people for career guidance.

B. Data schemata and linked data

This section will present a selection of data schemata, which follow the from star to galaxy schema. See Figure 2 in this section for an example. Quantitative and qualitative data will be used as examples, including:

- "Datensystem Auszubildende" (DAZUBI) is a system that collates data from the vocational training statistics of the statistical offices of the Federal Government. The annual total survey encompasses data on vocational training in accordance with the Vocational Training Act (BBiG) and the Crafts Code (HwO), including trainee, contract, and examination data.
- The QuBe project is a repository of data pertaining to future qualifications and occupations. Based on economic structure models, data is forecast up to 2050. The data set contains both past data and forecast data in 726 dimensions per job, branch, region.
- Two text corpora comprising approximately nine million online job advertisements (OJAs) are available for analysis. They consist of an average of 80dim. monthly data set.
- A substantial corpus of advertisements for continuing education.
- The labor market archive.
- A diverse array of online social media data with extracted sentiments for each job
- The quali panel, which examines the structures and developments in a longitudinal perspective of company

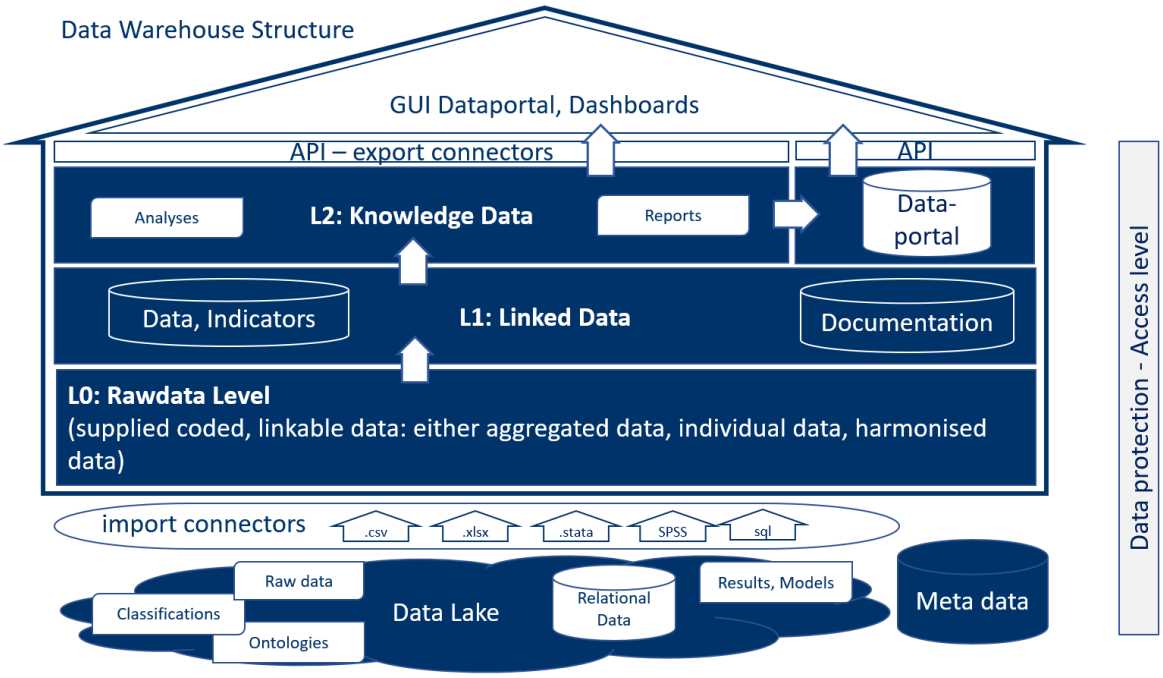

Fig. 1. Data warehouse structure: The data archive, in this instance, is the data lake, serving as the initial point of reference. The data derived from this archive is subsequently integrated into the data warehouse in a structured format.

activities and measures for training and securing skilled workers.
- The pension insurance data in the form of longitudinal and cross-sectional data contain data products for several years. The cross-sectional data products describe characteristics for reference dates (e.g. pension (e.g. pension portfolio) or reporting years (e.g. pension entries, pension histories).
- The collective training allowances for which collective agreements exist (only occupations with higher numbers of trainees)
- 42 quantitative and qualitative datasets on nursing training from 2021-2024 (only one job)
- employment surveys and 560 related indicators, like e.g. job satisfaction

It is necessary to obtain data instances that differ from one another. For instance, one instance may be used for ground truth data, while another may be employed for testing purposes. Some data will be subjected to processing. The size of the data is a significant challenge. Table I presents some information about the corresponding dimension of selected data sets.

Some data records are only collected once, whereas the majority of data records are collected and updated on an ongoing basis. In particular, for the data pertaining to job advertisements, which can be retrieved on a daily basis, an automation pipeline was designed to retrieve the new data, check for duplicates and update changes. Furthermore, an anomaly detection system has been integrated into the data retrieval process, in order to identify any technical or content-related anomalies that may occur during the processing stage.

Due to the high dimensionality of the data and the limitations of computing resources, dimension reduction must be achieved through the preselection of the data. However, this depends on the specific use case and the objective of the data analysis. The early aggregation or elevation of data to a higher level of abstraction may result in the undesirable blurring of distinctions. Prior to the commencement of the data selection process, it is necessary to define the dimensions that are deemed to be of interest, with this selection being dependent on the specific application in question. This may entail a reduction in the number of parameters or columns, but it may also involve imposing temporal constraints or preselecting specific occupational categories, occupational sectors, or occupational groups. A preliminary selection must therefore be made by domain experts or by means of feature extraction using data science methods. At present, the selection process is still carried out manually, using prior knowledge.

In order to illustrate the methodology, we will utilize two example data sets. The first data set comprises approximately 3.5 million tweets from Twitter/X on labor market data. The second comprises the metadata of approximately 5 million YouTube videos. In Figure 1, we describe the data schema for tweets. They follow the star schema, centered around tweets. These tweets are linked data, as the `job_id` is linked to the classification of occupations (KldB) according to GLMO. Additionally, named entities are linked to CSO and GLMO.

TABLE I
SOME EXAMPLES OF DATA SETS, TYPE AND UPDATE.

Dataset	Type	Update
Dazubi	quantitive	anual
Twitter/X, YouTube, Kununu	quantitive	none (single survey)
Employment surveys (e.g. ETB)	quantitativ	every 6 years
QualiPanel	qualitive	anual
QuBe	quantitativ	anual
Job advertisements	quantitativ	daily
Indicators derived from ETB and microcensus	quantitave	anual
Nursing training	quantitativ/qualitativ	anual

In Figure 2, we show the corresponding YouTube metadata schema. While this schema houses different data, it is evident that it is similarly connected to KldB and various other ontologies, such as GLMO and CSO. This produced linked data.

The primary objective when linking the data is to identify the levels at which the data can be linked. These levels are particularly relevant in the context of the classification systems used in the GLMO of occupations at different levels. If the data records do not have an assignment to the Kldb, an ontology-based textual mapping can be employed. Alternatively, an occupation mapping via ISCO is a potential option. In some cases, datasets lack an occupation assignment but include sector information, necessitating a mapping via economic sectors. Additionally, some datasets undergo regional mapping via location parameters. In the majority of instances, individual data records are represented as individual Star Schemata. The classification system is employed to create a galaxy schema 2 through mapping.

C. Stakeholders and roles

The data warehouse stakeholders are subject to different authorisation requirements depending on whether they are accessing the individual database instances or the individual levels.

It is the responsibility of the IT architects to ensure the backup and recovery of data on the technical side. However, they are not privy to the content of the data. In contrast, data architects are responsible for the overall structuring of the data architecture, which necessitates access to the content of the data.

In addition, data engineers have access to Level 0 data, as they are responsible for developing import pipelines, harmonisation procedures and transformation matrices. In particular, data experts with a background in economics and social impact are required for the harmonisation and alignment of marginal totals between different data sets. These domain experts then design the corresponding transformation models for the respective individual case. In subsequent work, it should be determined whether the modelling can be generalised and automated.

Those engaged in the practice of data analysis and business intelligence (BI) typically commence their training at the introductory level, which encompasses linked data and preparation for reporting, including the data portal.

IV. RESULTS

The results of our methodological approach deliver several results. First, we can demonstrate that even in this challenging interdisciplinary environment we can house linked data and make data available for querying for different stakeholders. The data can be integrated automatically via ELT pipelines with this concept and delivered for the corresponding use case or report request, as we could show for the integration of job advertisements on a daily basis. However, subsequent analysis revealed that research question 1 was overly ambitious, given the numerous challenges that were encountered. It is not possible to provide a one-size-fits-all solution. Rather, the solution must be selected on a case-by-case basis.

Second, we could identify some aspects which are subject for further research. Selecting data remained challenging, in particular because data cannot be preprocessed in enddless much situations. However, ad-hoc query are not possible without further research. The necessity for analysis increases exponentially with each additional dimension of the data, indicating that manual analyses should be conducted on a select number of phenomena in the future. It is recommended that standard analysis models be developed by data scientists and that their findings be fed back into the data pool. In this context, validated results are then treated as additional multidimensional indicators analogous to weighting matrices (expert knowledge). The graphical structure of the Galaxy schema precludes the possibility of working with the storage of all abstraction levels. Consequently, it is necessary to work with abstract views and ad-hoc generations on the respective data aspect, as the overall system has to deal with limited data storage.

Third, it is necessary to consider the possibility of automatic clustering, for example at a higher occupational level due to the lack of data (sparseness of data), separately. At this time, rule-based automatic solutions are being developed.

Furthermore the consideration of qualitative data has not yet been included in the analysis and must be analysed in further work.

V. CONCLUSION AND OUTLOOK

The objective of this study was to identify the challenges and approaches to solutions that would enable the heterogeneous data landscape of vocational education and training (VET) research to be linked and reusable.

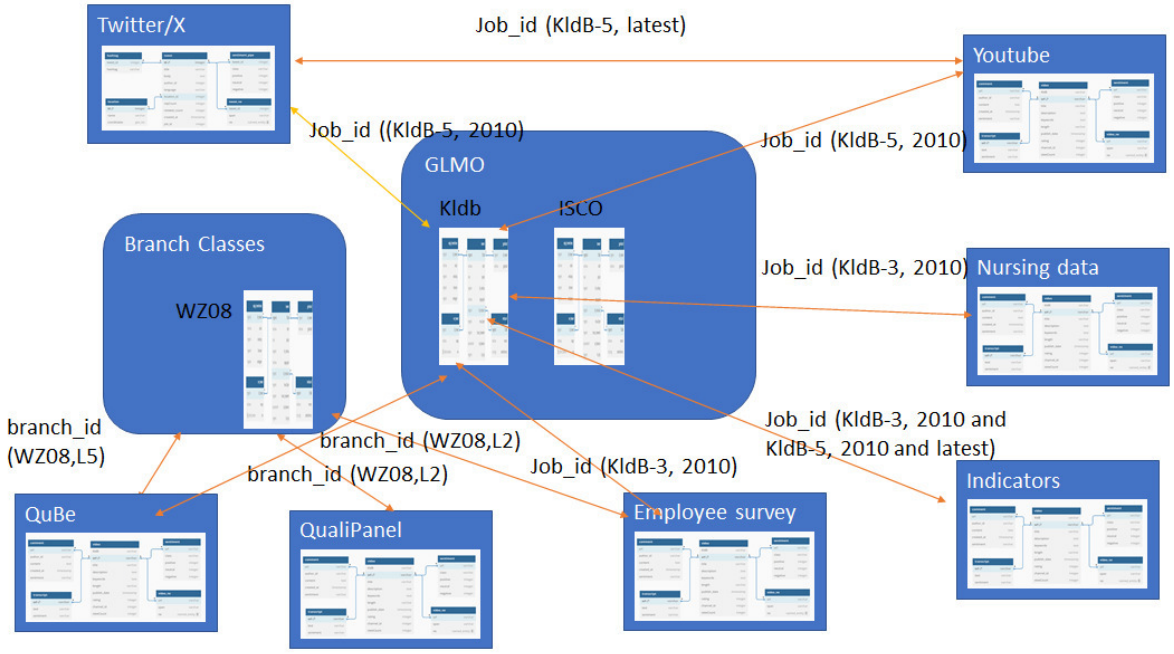

Fig. 2. Galaxy Data Schema Structure, that includes all single star and classification schemata

Although the research questions cannot be fully answered at this stage, we have identified a number of potential avenues for further investigation. These include the challenges associated with setting up a data warehouse structure for different types of data and the requirements of different stakeholders.

General challenges are how to intercept the temporal progression of data and classification systems, measurement system and dimensions change. The data can be categorised according to the level of detail provided. The following data types are to be considered: survey data, aggregated data, and linked data. A brief description of the linking process, including any pertinent notes on potential issues and expert opinions and analyses of the data.

The scientific reuse and further utilisation of all types of data requires the development of suitable procedures that ensure transparent and structured data documentation and the definition of minimum standards for the individual levels of a project dataset. We are currently developing a corresponding data management template. It would be beneficial to implement transparency and documentation, as well as data self-service, in order to facilitate the accessibility of data to the research community, without the necessity for experts in specific domains.

In light of the ever-evolving landscape of data technologies, it is prudent to adopt a modular and state-of-the-art structure wherever feasible. Migrations are a costly undertaking, requiring significant resources (time and money) and more. These include the possibility of system downtime, dissatisfaction, inconsistencies between data records, and a loss of confidence in data quality.

The harmonisation of data is still only possible with the input of experts, who are in short supply. In subsequent work, it should be determined whether the modelling can be generalised and automated.

References

[1] R. Dobischat, B. Käpplinger, G. Molzberger, and D. Münk, "Digitalisierung und die folgen: Hype oder revolution?" *Bildung 2.1 für Arbeit 4.0?*, pp. 9–24, 2019.

[2] R. Helmrich, M. Tiemann, K. Troltsch, F. Lukowski, C. Neuber-Pohl, A. C. Lewalder, and B. Gunturk-Kuhl, *Digitalisierung der Arbeitslandschaften: keine Polarisierung der Arbeitswelt, aber beschleunigter Strukturwandel und Arbeitsplatzwechsel*. Wissenschaftliche Diskussionspapiere, 2016, no. 180.

[3] J. Dörpinghaus, D. Samray, and R. Helmrich, "Challenges of automated identification of access to education and training in germany," *Information*, vol. 14, no. 10, p. 524, 2023.

[4] J. De Smedt, M. le Vrang, and A. Papantoniou, "Esco: Towards a semantic web for the european labor market." in *Ldow@ www*, 2015.

[5] J. Dörpinghaus, J. Binnewitt, S. Winnige, K. Hein, and K. Krüger, "Towards a german labor market ontology: Challenges and applications," *Applied Ontology*, no. 18(4), pp. 1–23, 2023.

[6] P. Koch, W. Konen, and K. Hein, "Gesture recognition on few training data using slow feature analysis and parametric bootstrap," in *International Joint Conference on Neural Networks*, Barcelona, Spain, Jul. 2010, p. 8 pages.

[7] A. Dutt, M. A. Ismail, and T. Herawan, "A systematic review on educational data mining," *Ieee Access*, vol. 5, pp. 15 991–16 005, 2017.

[8] S. K. Mohamad and Z. Tasir, "Educational data mining: A review," *Procedia-Social and Behavioral Sciences*, vol. 97, pp. 320–324, 2013.

[9] C. Romero and S. Ventura, "Educational data mining: A survey from 1995 to 2005," *Expert systems with applications*, vol. 33, no. 1, pp. 135–146, 2007.

[10] J. Dörpinghaus and M. Tiemann, "Vocational education and training data in twitter: Making german twitter data interoperable," *Proceedings of the Association for Information Science and Technology*, vol. 60, no. 1, pp. 946–948, 2023.

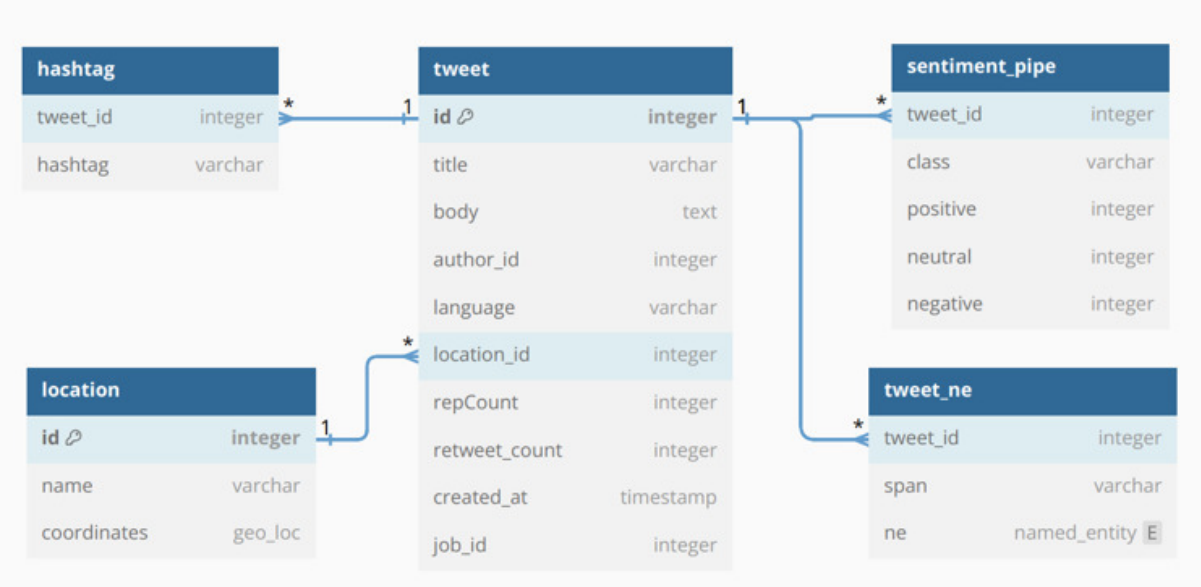

Fig. 3. Simplified star schema ERD for Twitter/X data.

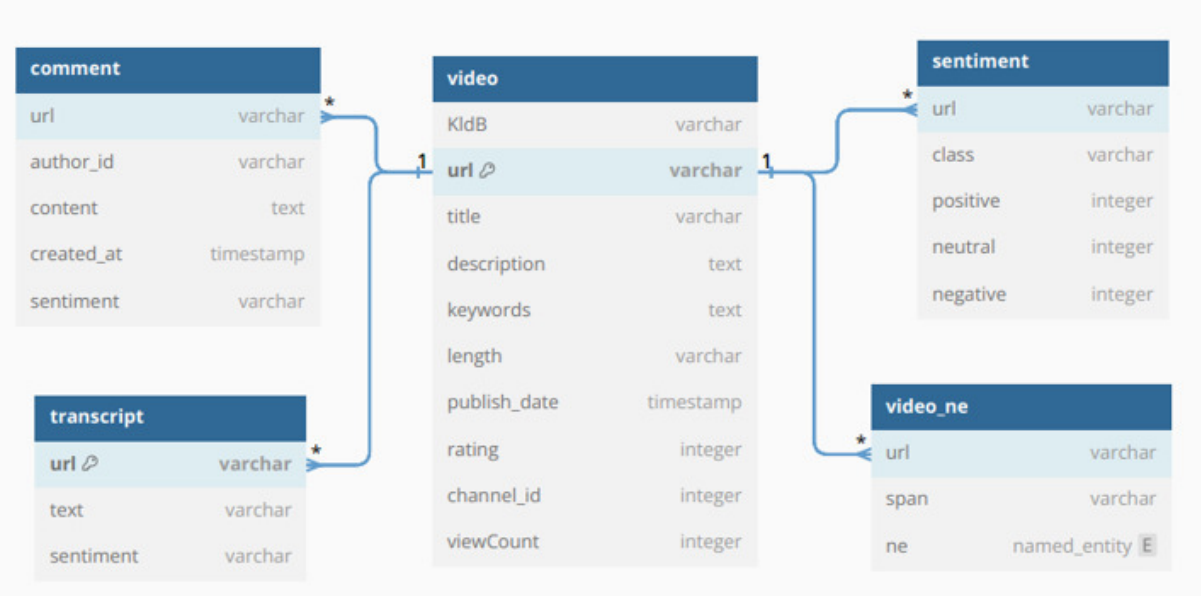

Fig. 4. Simplified star schema ERD for Youtube data.

[11] R. M. Chang, R. J. Kauffman, and Y. Kwon, "Understanding the paradigm shift to computational social science in the presence of big data," *Decision Support Systems*, vol. 63, pp. 67–80, 2014, 1. Business Applications of Web of Things 2. Social Media Use in Decision Making. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0167923613002212

[12] I. Moalla, A. Nabli, L. Bouzguenda, and M. Hammami, "Data warehouse design approaches from social media: review and comparison," *Social Network Analysis and Mining*, vol. 7, no. 1, p. 5, 2017.

[13] B. Marlis, H. Buchs, and G. Ann-Sophie, "Occupational inequality in wage returns to employer demand for types of information and communications technology (ict) skills: 1991–2017," *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, vol. 72, pp. 455–482, 2020.

[14] J. Dörpinghaus, J. Binnewitt, and K. Hein, "Lessons from continuing vocational training courses for computer science education," in *Proceedings of the 2023 Conference on Innovation and Technology in Computer Science Education V. 2*, 2023, pp. 636–636.

[15] A. Settelmeyer, F. Bremser, and A. C. Lewalder, "Migrationsbedingte mehrsprachigkeit—ein "plus" beim übergang von der schule in den beruf," *Interkulturelle und sprachliche Bildung im mehrsprachigen Übergang Schule-Beruf*, pp. 135–150, 2017.

[16] F. Derksen and J. Dörpinghaus, "Digitalization and sustainability in german continuing education," in *INFORMATIK 2023 - Designing Futures: Zukünfte gestalten*. Bonn: Gesellschaft für Informatik e.V., 2023, pp. 1945–1953.

[17] P. K. Ningrum, T. Pansombut, and A. Ueranantasun, "Text mining of online job advertisements to identify direct discrimination during job hunting process: A case study in indonesia," *Plos one*, vol. 15, no. 6, p. e0233746, 2020.

[18] I. Smirnov, "Estimating educational outcomes from students' short texts on social media," *EPJ Data Science*, vol. 9, no. 1, pp. 1–11, 2020.

[19] C. Ospino, "Occupations: Labor market classifications, taxonomies, and ontologies in the 21st century," *Inter-American Development Bank*, 2018.

[20] S. Guru Rao, "Ontology matching using domain-specific knowledge and semantic similarity," Master's thesis, University of Twente, 2022.
[21] I. Szabó, "The implementation of the educational ontology," in *Proceedings of the 7th European Conference on Knowledge Management, Corvinus University of Budapest, Hungary, ACL, UK*, 2006, pp. 541–547.
[22] E. Boldyreva and V. Kholoshnia, "Ontological approach to modeling the current labor market needs for automated workshop control in higher education." in *MICSECS*, 2019.
[23] M. Khobreh, F. Ansari, M. Fathi, R. Vas, S. T. Mol, H. A. Berkers, and K. Varga, "An ontology-based approach for the semantic representation of job knowledge," *IEEE Transactions on Emerging Topics in Computing*, vol. 4, no. 3, pp. 462–473, 2015.
[24] M. Papoutsoglou, A. Ampatzoglou, N. Mittas, and L. Angelis, "Extracting knowledge from on-line sources for software engineering labor market: A mapping study," *IEEE Access*, vol. 7, pp. 157 595–157 613, 2019.
[25] A. Poletaikin, S. Sinitsa, L. Danilova, Y. Shevtsova, and N. Dvurechenskaya, "Ontology approach for the intelligent analysis of labor market and educational content matching," in *2021 International Symposium on Knowledge, Ontology, and Theory (KNOTH)*. IEEE, 2021, pp. 50–55.
[26] Z. Li, W. Xu, L. Zhang, and R. Y. Lau, "An ontology-based web mining method for unemployment rate prediction," *Decision Support Systems*, vol. 66, pp. 114–122, 2014.
[27] T.-P. Liang and Y.-H. Liu, "Research landscape of business intelligence and big data analytics: A bibliometrics study," *Expert Systems with Applications*, vol. 111, pp. 2–10, 2018.
[28] T. Avdeenko and M. Bakaev, "Modeling information space for decision-making in the interaction of higher education system with regional labor market," in *2014 12th International Conference on Actual Problems of Electronics Instrument Engineering (APEIE)*. IEEE, 2014, pp. 617–623.
[29] A. Fischer and J. Dörpinghaus, "Web mining of online resources for german labor market research and education: Finding the ground truth?" *Knowledge*, vol. 4, no. 1, pp. 51–67, 2024.
[30] B. Batrinca and P. C. Treleaven, "Social media analytics: a survey of techniques, tools and platforms," *Ai & Society*, vol. 30, pp. 89–116, 2015.
[31] M. Tiemann, H.-J. Schade, R. Helmrich, A. Hall, U. Braun, and P. Bott, "Berufsfeld-definitionen des bibb auf basis der klassifikation der berufe 1992," *Schriftenreihe des Bundesinstituts für Berufsbildung Bonn*, vol. 105, no. 1, p. 57, 2008.