Skills-Matching and Skills Intelligence through curated and data-driven ontologies

Usage scenarios, practical examples and the question: Who annotates education?

Robert Rentzsch[1], Mila Staneva[2]

Abstract: This paper reviews the fields of application of classification systems for job-relevant skills and competences. European and international public and private-sector developments for Skills Matching (in particular the matching of jobs and job seekers) and Skills Intelligence (labour market analyses and forecasts, in particular for political decision makers) are examined in an exemplary manner. We first provide an overview of the internationally most influential skills taxonomies and ontologies of the classical kind, which are created and curated by experts. These are contrasted with new, data-driven classifications. Subsequently, we describe the different areas of application of both classification types, followed by a wide variety of practical examples of initiatives, projects and products in the above-mentioned areas. The outlook section especially addresses the question of why and how, in addition to applicants and jobs, all types of educational programs should be annotated with skills taxonomies in the future.

Keywords: skills, competencies, qualifications, skills, jobs, matching, taxonomy, ontology, classification, ESCO, O*NET, recruiting, job market, analysis, education, training programs.

[1] Institute for Innovation and Technology (iit), Steinplatz 1, 10623 Berlin, rentzsch@iit-berlin.de, https://orcid.org/0000-0003-0932-8777

[2] Institute for Innovation and Technology (iit), Steinplatz 1, 10623 Berlin, staneva@iit-berlin.de

1 Overview

For a long time, systematically since the 1960s, states and organizations have been using systems to classify occupations and occupational groups for statistical purposes and in the context of standardization efforts. The internationally most influential classification of this kind is the "Standard Classification of Occupations" (ISCO)[3] of the International Labour Organization (ILO), to which the German Classification of Occupations (KldB)[4] is compatible. In the course of the increasing differentiation of the labour market and occupational roles, the first skills taxonomies were added in the 1990s, for example as part of the Occupational Information Network (O*NET)[5] database in the USA and in the form of the multilingual European Dictionary of Skills and Competencies (DISCO)[6]. Since 2010, the European Commission (EC) has been developing the classification system 'European Skills, Competences, Qualifications and Occupations' (ESCO)[7], now available in 27 languages. As with O*NET, this is not a single taxonomy. Rather, occupations or occupational fields as well as skills and knowledge concepts are hierarchically defined in two separate pillars, with cross-references describing, for example, which skills are required for a particular occupation; a third pillar, "qualifications", is taken up again in the conclusion section. ESCO and O*NET are thus ontologies that establish links between concepts from different taxonomies. As of June 2020, ESCO contains ~13,500 skill, knowledge and attitude/value concepts[8] as well as ~3,000 occupational concepts. An overview of further national classifications is given in [SF17].

The skills and occupations in ESCO and O*NET are each curated by teams of sector experts based on their domain knowledge. Many smaller, sector-specific taxonomies are created in a similar way. Examples are the EU projects DigComp[9] and e-Competence Framework[10] in the IT sector, EntreComp[11] for entrepreneurial skills and the German National Competence-Based Learning Objectives Catalogue for Medicine (NKLM)[12] or the international Medbiquitous[13] for medical professions. Alternatively, a variety of data-driven ontologies have been developed in recent years by the private sector, both by specialised software providers for the areas of personnel recruitment and -management (HR) and by tech giants such as Google and LinkedIn. In this context, very large datasets of job advertisements are usually analysed and processed using Natural Language Processing (NLP) and Machine Learning (ML) algorithms. Based on the learned

[3] https://www.ilo.org/public/english/bureau/stat/isco
[4] https://statistik.arbeitsagentur.de/Navigation/Statistik/Grundlagen/Klassifikationen/Klassifikation-der-Berufe/Klassifikation-der-Berufe-Nav.html
[5] https://www.onetcenter.org
[6] http://disco-tools.eu
[7] https://ec.europa.eu/esco/portal/home
[8] https://ec.europa.eu/esco/portal/escopedia/Skills_pillar
[9] https://ec.europa.eu/jrc/en/digcomp
[10] https://www.ecompetences.eu/
[11] https://ec.europa.eu/social/main.jsp?catId=1317
[12] http://www.nklm.de
[13] https://www.medbiq.org/competency_object

dependency relationships between skills and professional roles as well as similarity relationships within skills and job titles, ontologies are created that can be updated at any time and adapted to the dynamics of the labour market; often additional existing taxonomies and/or manual curation steps are included. While these proprietary approaches are usually not explained in detail, they are sometimes found in scientific publications, e.g. [DS18, KMC19, MJB20]. Finally, governmental initiatives are also increasingly pursuing the data-driven approach.

In the following, the various areas of application of skills taxonomies will be examined in more detail, before section 3 describes concrete practical examples of initiatives, projects and products in the areas of skills- or job-matching and labour market analysis. The outlook particularly deals with the question of why and how learning and training opportunities should be annotated with skills taxonomies in the future, in addition to applicants and jobs.

2 Fields of application

Classifications such as those described above are used in a variety for business, politics and administration purposes, but also by individuals, even if this use is usually indirect: The taxonomies or ontologies, together with NLP and ML algorithms, form the basis of a wide range of search, matching and analysis tools. Their areas of application include:

- Machine-readable annotation: Skills taxonomies are used in (typically commercial) assistance systems for creating resumes (CVs) and job advertisements, which suggest appropriate taxonomy concepts based on the free text provided. These can then be integrated into the respective documents in a machine-readable form, which allows further processing (see below). The approach can also be extended to more complex ontologies.

- Semantic searching: Ontologies form the core of semantic search technologies. In contrast to keyword-based searches, semantic searches take into account the complex relationships between concepts (such as skills and occupations) captured in an ontology and thus approach the semantic meaning of the search term. Not only linguistically synonymous concepts, but also such that are merely semantically related can be taken into account.

- Semantic matching: As an extension of semantic searching, entire documents can be matched, e.g. CVs with job profiles in HR. For this purpose, all documents are first annotated as described above. Then, an algorithm determines the semantic proximity of the concepts found, respectively. Based on this, a ranking is created, listing e.g. the best candidates for a job or the most suitable jobs for a candidate. Talent management tools for (re)assigning employees to tasks and positions in a company also make use of this approach, as well as systems that identify individual skills gaps or skills mismatches and suggest suitable further training.

- Multilingualism: multilingual classifications such as DISCO or ESCO enable the uniform annotation of CVs and job advertisements in different languages. This facilitates the exchange of information between employers and jobseekers across borders and makes transnational surveys and forecasting (see below) possible.

- Labour market analysis: This deals with statistics and predictions on the supply of and demand for skills and qualifications in the labour market. Skills intelligence tools developed by the public and private sectors, which analyse large numbers of job advertisements using the methods described above [Ri16], are being used to identify structural skills gaps and skills mismatches and to enable political countermeasures. In the medium term, protracted trend analyses by means of surveys, currently the global standard in labour market analysis, could be supplemented or replaced by daily updated data. This provides a significantly higher level of detail, too: In addition to the results that, for example, programmers are being sought, it can be distinguished that Python is in high demand in region A whereas employers mainly look for Ruby coders in region B. Such aspects are of interest not least for education providers that want to align their offers closely with the needs of the labour market.

3 Practical examples

The following are concrete examples of tools, websites and initiatives that use and/or create skills taxonomies and ontologies.[14]

Creation of applicant and job profiles

- Europass: The new Europass portal, to be launched in 2020, will include an online CV editor with an ESCO-based recommendation/annotation system for the machine-readable description of users' skills and prior knowledge. It is planned that the resulting profile can be used for direct job searches via EURES (see below) and easily be transferred to other platforms too, e.g. LinkedIn.[15]

- European Broadcasting Union (EBU): The HR department of the EBU has sustainably integrated and expanded ESCO in its working methods with the aim to produce more detailed job advertisements in the media sector.[16] This and other pilot projects also contribute to the further development of ESCO.

- EU Skills Profile Tool for Third Country Nationals[17]: Developed by the EC, this tool uses the ESCO classification to record the skills and professional experience of third-country nationals in a standardised way. It aims to support authorities, employment

[14] For brevity, the corresponding NLP and ML methods are not discussed in detail here.
[15] https://ec.europa.eu/futurium/en/europass/e-portfolio-and-web-based-tools
[16] https://ec.europa.eu/esco/portal/escopedia/European_Brodcasting_Union__40_EBU_41_
[17] https://ec.europa.eu/migrantskills

agencies, social services and non-governmental organisations in counselling, job placement and validation of previous experience.

- SkillLab: The Dutch start-up[18] has developed an assessment app that identifies relevant skills of refugees (similar to the EC-developed tool described above) and converts this information into standardised application documents. The app suggests ESCO skills based on reported professional and other experience. SkillLab cooperates with the ILO, among others.

Optimization of job search engines

- Actonomy: This Belgian provider of HR software has been collecting data for more than 15 years and has linked its HR ontology, consisting of more than 500,000 concepts, with ESCO, to optimize its "xMP Semantic Mind" recruiting application.[19]

- Google: A job search feature, "Google for Jobs", has recently been added to the popular search engine. It builds upon the in-house Cloud Talent Solution[20], which is marketed separately in the HR area. This is based on two interlinked ontologies, O*NET and a proprietary one that is again "learned" through data mining and includes more than 50,000 skill concepts. The result is an ontology with 30 job categories, over 1,000 job families and approximately 250,000 job roles, all linked to specific skills[21].

- JANZZ.technology: The core product of the Swiss company is the JANZZon![22] ontology for job-related data, which is available in several languages and, in contrast to the purely data-driven approaches mentioned above, also integrates more than 80 curated classifications such as O*Net, ESCO, DISCO, UK Skills Taxonomy (see below) and KldB. Together with a semantic-matching product it forms the basis for the job portal JANZZ.jobs.

- LinkedIn, Monster: The LinkedIn career network builds its own ontology (or "knowledge graph") of job titles, skills and qualifications. This is based on user-generated content and external data[23]. In an ESCO pilot project, a mapping between the skills in the in-house taxonomy and those in ESCO was created[24]. In another pilot project, Monster uses ESCO as a basis for the semantic job search on its Spanish portal.[25]

- Textkernel: The Dutch company uses a proprietary ontology for its semantic (job) search solutions, which is based on data mining and quality assured by domain

[18] https://skilllab.io/en-us
[19] https://www.actonomy.com/whats-new/blog/combining-actonomys-ontology-with-esco
[20] https://cloud.google.com/solutions/talent-solution
[21] https://cloud.google.com/blog/products/gcp/cloud-jobs-api-machine-learning-goes-to-work-on-job-search-and-discovery
[22] https://janzz.technology/janzz-on
[23] https://engineering.linkedin.com/blog/2016/10/building-the-linkedin-knowledge-graph
[24] https://ec.europa.eu/esco/portal/escopedia/Joint_pilot_project_with_LinkedIn
[25] https://ec.europa.eu/esco/portal/news/61ef2465-7b91-4c1f-a5a0-afc390f42b90

experts[26]. This has been linked to common taxonomies and ontologies such as ISCO, ESCO, O*NET and ROME to ensure interoperability with other systems.

Job Matching algorithms

- DOCEBO[27]: This Italian software provider with several international branches offers an ESCO- and ML-based enterprise learning management system. Among other things, the latter allows for comparing the skills of the employees with the needs of the company in order to identify skills gaps. It can then make tailor-made training proposals.

- Economic Modeling Specialists Inc. (EMSI): The US firm is a direct competitor of Burning Glass Technologies (see below) in various areas. However, a rare feature is its publically available online matching tool for individuals[28]. This matches users with suitable professions, based on a selection of skills via a suggestion system and/or an uploaded CV. The difference to traditional tools of this kind from the field of career guidance is the underlying in-house taxonomy (and the collected data behind it) together with the associated matching algorithms[29]. Another tool, which was made freely available due to the Corona pandemic and its impact on the labour market, makes suggestions for the optimization of CVs.[30]

- HR software: There are a number of ontology-based semantic-matching products, e.g. Match! by Textkernel, JANZZsme! by JANZZ.technology, xMP Semantic Mind by Actonomy, the AI Matching Engine by Sovren[31] or Google's Cloud Talent Solution (see above). These tools enable the automated matching of applications and candidates with job descriptions, and vice versa. As a rule, both structured data (e.g. from in-house databases) and unstructured data (e.g. from social networks) can be processed.

- Milk & Sugar: Based on ESCO, this German company has developed the matching tool AI-Match for its JobStairs job portal. This tool compares a short profile of job seekers with job offers and yields individualised recommendations.[32]

- OpenSKIMR[33]: The ERASMUS+-funded project has developed algorithms to support individual career planning. A matching algorithm assigns people to occupations and specific jobs based on their skills profile, while another one captures the gap between existing skills and those required for the desired job. A route-

[26] https://www.textkernel.com/ontology-mining
[27] https://www.docebo.com/
[28] https://match.emsiskills.com
[29] https://skills.emsidata.com/faqs#taxonomy
[30] https://skills.emsidata.com/resume
[31] https://www.sovren.com/products/ai-matching
[32] https://www.milchundzucker.de/matching-technologie-der-ai-match-von-jobstairs-ganz-schoen-smart
[33] https://openskimr.eu

planning algorithm then suggests further training measures. ESCO forms the basis for these algorithms.[34]

Multilingualism

- European Employment Services (EURES)[35]: This cooperation network of EU employment services aims to promote the mobility of workers despite language barriers. To this end, plans exist to integrate ESCO into the EURES job search portal. According to the corresponding implementing decision, this is necessary "to facilitate the exchange of job vacancies and job seeker profiles and to ensure a high quality matching across languages and national contexts" [EU18a]. To connect to EURES, the EU member states must either use the ESCO occupations classification by 2021 or map their national systems to it [EU18a, EU18b]. Accordingly, EU-funded mapping projects are currently underway in at least 12 countries, including Germany.

Labour market analysis

- EU Skills Panorama[36]: The Skills Intelligence Portal is being developed by the European Centre for the Development of Vocational Training (Cedefop) on behalf of the EC and is aimed primarily at policy makers and analysts, career and education advisors as well as labour market researchers. It offers central access to data and information skills demand across the labour markets of EU states and regions. Together with other classifications[37] and datasets[38], the tool uses ESCO to annotate about 67 million job advertisements from 2018/19 with skills. The underlying project "Skills Online Vacancy Analysis Tool for Europe" (Skills-OVATE)[39] uses a mixed approach of API usage, web crawling and -scraping to collect job advertisements from more than 300 public and commercial portals [CED19].
- OECD Skills For Jobs Database: The Organisation for Economic Cooperation and Development (OECD) uses this database as the basis for a dashboard[40] on skills supply and demand for actors in politics, education and the world of work. It currently covers more than 40 countries, in some cases down to the regional and sectoral level. O*NET, among others, is used for data generation.
- OpenSkills project: Internationally, there are numerous smaller projects similar to those mentioned above, for example for certain regions or sectors [MM19]. Additionally, the T3 Innovation Network[41], founded in 2019 by the US Chamber of Commerce Foundation together with other foundations and major employers such as

[34] https://ec.europa.eu/esco/portal/escopedia/Openskimr
[35] https://ec.europa.eu/eures/public/homepage
[36] https://skillspanorama.cedefop.europa.eu
[37] https://skillspanorama.cedefop.europa.eu/en/about-us/about-the-data
[38] https://skillspanorama.cedefop.europa.eu/en/datasets
[39] https://www.cedefop.europa.eu/en/data-visualisations/skills-online-vacancies
[40] https://www.oecdskillsforjobsdatabase.org
[41] https://www.uschamberfoundation.org/t3-innovation

Walmart, will most likely develop a tool for the US similarly comprehensive to that of Cedefop in the coming years, based on the OpenSkills project[42].

- Burning Glass Technologies (BGT): This US software provider has been active in the field of skills matching and skills intelligence for more than 15 years. According to the company, its database consists of approximately 3.5 million active job advertisements, for which more than 50,000 sources on the Internet are automatically queried on a daily basis. From this, BGT creates and continuously adapts a dynamic job- and skills-ontology [BG19]. The wide range of products mainly aims at decision-makers in education and politics[43]. BGT also regularly participates in joint studies with stakeholders from research and politics. For example, the company supplied the data for the UK Skills Taxonomy (see below) and is also currently exchanging data with Cedefop.
- Economic Modeling Specialists International (EMSI): This BGT-competitor has been on the market for just as long and, according to its own statements, evaluates several 100 million job advertisements together with over 100 million online CVs and candidate profiles. Amongst other things, this data is used to create a taxonomy with over 30,000 skills that is updated fortnightly. In terms of product and target group, EMSI is focusing even more strongly on labour market analyses, especially regional and sector-specific ones. The firm also operates a UK branch, like BGT, and is further active in Canada and Australia. An example of a trend analysis application is EMSI's Job Postings Dashboard[44], where users can filter by skills, among other things.
- UK Skills Taxonomy: The British NESTA Foundation, in collaboration with BGT and others, has developed a data-driven, interactive skills taxonomy for the UK labour market[45]. This was supplemented by estimates of skills demand in the national and regional labour market (based on the frequency of skills being mentioned in job postings) and the market value of individual skills (based on salary information provided therein). This should make it possible to identify skills in high demand. In the future, the developers also plan to relate educational offers to this demand.

4 Outlook

Due to the rapidly increasing possibilities in the field of "Big Data", there is currently a great momentum in the development of classifications for skills and occupations, as well as in their main application fields of skills matching and skills intelligence. This particularly concerns the area of data-driven ontologies, where, in addition to the examples mentioned here, many more players (above all the operators of the major job search

[42] http://dataatwork.org
[43] https://www.burning-glass.com/solutions/government-and-workforce
[44] https://www.economicmodeling.com/job-posting-dashboard
[45] https://www.nesta.org.uk/data-visualisation-and-interactive/making-sense-skills

portals) are developing their own "skills clouds" using the above-mentioned methods[46]. Particularly the combination of the advantages of expert-created classifications such as ESCO and O*NET with those of new, data-driven ontologies holds great potential: The former can be used to reliably record macroscopic and longer-term labour market trends and to consistently describe occupations and transferable skills such as soft skills or languages, the latter to map labour market demand (almost) in real-time and to track micro-trends in in-demand skills.

Notably, the topic of annotating educational offers with taught skills – compared to annotating job offers with requested skills – has so far received little attention in both public and private-sector developments. Many recent publications in the academic field have addressed the aim of uniformly describing learning outcomes and making curricula comparable within individual fields of study or occupations ("curriculum alignment"), e.g. [KBS19] and [WT19]. However, even there the link to comprehensive developments such as ESCO or O*NET, and thus the possibility of direct usage for the bridging "Education ↔ Skills ↔ Labour market", has so far rarely been established. This would be particularly important for skills intelligence analyses to unfold their full potential: In addition to the demand for skills, annotating education would also make it possible to measure skills supply, in terms of corresponding learning and training opportunities. This would further allow to estimate the supply of workers with these skills, e.g. via alumni statistics or publically available data on individual educational pathways from career profiles. These measures could then serve as guidelines for education initiatives and programs.

As of June 2020, the ESCO "Qualifications" pillar contains educational offers from only eight countries, not including Germany. Therefore, the degree of cooperation of and uptake by the responsible national institutions is thus far limited. For the offers shown, the data is often patchy. The skills annotation option[47] that is available in the ESCO metadata schema for qualifications has not been used at all. This is hardly surprising, given the lack of national requirements of this kind for education providers. Encouragingly, based on a study [LAW19] commissioned by the ESCO developers that has recognised this problem, a pilot project aiming to test parsing and assistance systems (similar to the one described above) has been launched[48]. The additional challenge there: In contrast to job offers and CVs, documents describing educational offers of all kinds so far hardly adhere to any uniform structure.

It is possible that, for example, the booming online- and offline-market for continuing education (with financial incentives for private providers and universities alike) and, more generally, the growing efforts towards digital credentialing will promote and accelerate these developments. Particularly in the case of newly created, innovative educational offers, describing the skills taught and proven in examinations in a standardised manner

[46] https://joshbersin.com/2020/01/the-war-of-the-skills-clouds-skillscloud
[47] https://ec.europa.eu/esco/portal/escopedia/Qualifications_pillar
[48] https://ec.europa.eu/esco/portal/escopedia/Linking_pilot

from the outset, for example with or on the basis of subject-specific extensions of ESCO, could provide advantages for both education seekers and providers. These include better comparability and greater international visibility. Fundamentally, any additional machine-readable annotation of offers otherwise described in free text, regardless of how coarse and inadequate it might initially be, would benefit the types of projects and products addressed in this article, and thus all users and stakeholders. The immense efforts towards ontology development in both the public and private sectors, as described in this article, must, sooner better than later, also trigger an echo in the educational system.

Funding

The project behind this study was funded by the German Federal Ministry of Education and Research (Bundesministerium für Bildung und Forschung) via grant number M532600. Sole responsibility for the contents of this publication lies with the authors.

Bibliography

[BG19] Burning Glass Technology: Mapping the Genome of Jobs. The Burning Glass skills taxonomy. White Paper, 2019.

[CED19] Cedefop, European Centre for the Development of Vocational Training: Online job vacancies and skills analysis: a Cedefop pan-European approach. Luxembourg: Publications Office, 2019.

[DS18] Djumalieva, J.; Sleeman, C.: An Open and Data-driven Taxonomy of Skills Extracted from Online Job Adverts. ESCoE Discussion Paper 2018-13, 2018.

[EU18a] European Commission: Commission Implementing Decision (EU) 2018/1020 of 18 July 2018 on the adoption and updating of the list of skills, competences and occupations of the European classification for the purpose of automated matching through the EURES common IT platform. Official Journal of the European Union L 183/17, 2018.

[EU18b] European Commission: Commission Implementing Decision (EU) 2018/1021 of 18 July 2018 on the adoption of technical standards and formats necessary for the operation of the automated matching through the common IT platform using the European classification and the interoperability between national systems and the European classification. Official Journal of the European Union L 183/20, 2018.

[KMC19] Khaouja, I.; Mezzour, G.; Carley, K.M.; Kassou, I.: Building a soft skill taxonomy from job openings. Social Network Analysis and Mining 1/2019, 2019.

[KBS19] Konert, J.; Buchem, I.; Stoye, J.: Digital competence directory with semantic web technology. Linking digital proofs of competence with machine-readable competence definitions: A practical solution used in the Open Virtual Mobility project. In (Schulz, S., Publ.): Proceedings of DELFI Workshops. Berlin. S.61-69, 2019.

[LAW19] Luomi-Messerer, K.; Andersen, T.; Wilson, F.; Blakemore, M. RfS 43 - Support to the Development of the ESCO qualifications pillar. Final Report v7. European Commission, Brussels, 2019.

[MM19] Mezzanzanica, M.; Mercorio, F.: Big Data for Labour Market Intelligence: An Introductory Guide. European Training Foundation, 2019.

[MJB20] Mrsic, L.; Jerkovic, H.; Balkovic, M.: Interactive Skill Based Labor Market Mechanics and Dynamics Analysis System Using Machine Learning and Big Data. In (Sitek P.; Pietranik M.; Krótkiewicz M.; Srinilta C., Publ.): Intelligent Information and Database Systems. ACIIDS 2020. Communications in Computer and Information Science, vol 1178. Springer, Singapore, S. 505-516, 2020.

[Ri16] Rihova, H.: Using Labour Market Information. Guide to Anticipating and Matching Skills and Jobs, Vol. 1. In (European Training Foundation; European Centre for the Development of Vocational Training; International Labour Office, Publ.): Using Labour Market Information. Luxembourg: Publications Office of the European Union, 2016.

[SF17] Siekmann, G.; Fowler, C.: Identifying Work Skills: International Approaches. Discussion Paper of the National Centre for Vocational Education Research (NCVER), 2017.

[WT19] Wunderlich, J.; Tilebein, M.: Transforming Intended Learning Outcomes expressed in natural language into elements of an ontology: Towards the formalisation of Intended Learning Outcomes for use in Curriculum Maps. In (Schulz, S., Publ.): Proceedings of DELFI Workshops. Berlin. S.101-109, 2019.