

Deep Learning-based Computational Job Market Analysis: A Survey on Skill Extraction and Classification from Job Postings

Elena Senger^{1,3}, Mike Zhang², Rob van der Goot², Barbara Plank^{1,2}

¹MaiNLP, Center for Information and Language Processing, LMU Munich, Germany

²Department of Computer Science, IT University of Copenhagen, Denmark

³Fraunhofer Center for International Management and Knowledge Economy IMW, Germany

elena.senger@cis.lmu.de, {mikz, robv}@itu.dk, b.plank@lmu.de

Abstract

Recent years have brought significant advances to Natural Language Processing (NLP), which enabled fast progress in the field of *computational job market analysis*. Core tasks in this application domain are *skill extraction and classification* from job postings. Because of its quick growth and its interdisciplinary nature, there is no exhaustive assessment of this emerging field. This survey aims to fill this gap by providing a comprehensive overview of deep learning methodologies, datasets, and terminologies specific to NLP-driven skill extraction and classification. Our comprehensive cataloging of publicly available datasets addresses the lack of consolidated information on dataset creation and characteristics. Finally, the focus on terminology addresses the current lack of consistent definitions for important concepts, such as hard and soft skills, and terms relating to skill extraction and classification.

1 Introduction

Skill extraction and classification has recently been the subject of an increased amount of interest (Zhang et al., 2023; Clavié and Soulié, 2023), which shows in a high number of publications, driven by the advances in natural language processing (NLP) technology. For instance, through large language models (LLMs) the low resource tasks of skill extraction can be approached by using synthetic training data (Clavié and Soulié, 2023; Decorte et al., 2023). Surveys regarding skill extraction are emerging (Khaouja et al., 2021a; Pappoutsoglou et al., 2019), nevertheless, a comprehensive overview from an NLP perspective is still lacking—a gap we aim to fill in this survey. Our contributions are:

- Firstly, we aim to address the lack of standardized terminology in the field, bringing clarity to terms like hard and soft skills, as well as

phrases related to skill extraction and classification.

- Additionally, this survey is the first to examine various publicly accessible datasets and sheds light on their creation methodologies.
- In contrast to prior surveys, we adopt an NLP-centric focus, with a deep dive into the latest advancements of neural methods for skill extraction and classification.

While prior surveys exist, they focus typically on *Skill count* and *Topic modeling* methods for extracting skills. Skill count is performed manually or by matching n-grams with a skill base. Topic modeling is an unsupervised method utilizing word distributions to identify underlying topics in documents. Due to primary statistical basis and lack of defined skill spans or labels, topic modeling, as well as skill count, methods are not covered in this survey. For further details on skill count, see Khaouja et al. (2021a) and Ternikov (2022), and for topic modeling, please refer to Khaouja et al. (2021a), Ternikov (2022) and Ao et al. (2023).

Research Methodology For our search strategy we used several academic databases including the ACL Anthology, Google Scholar, arXiv, IEEE, ACM, Science Direct, and Springer Link. The primary search terms were “skill extraction” and “job”. To refine the search, we added terms like “deep learning”, “machine learning”, or “natural language processing” to our query for Google Scholar and Science Direct databases. This yielded the inclusion of 26 publications on neural skill extraction from job postings (JPs) that were published before November 2023.

2 Other Surveys

Previous surveys provide a foundation for our survey. Notable contributions include works from the social sciences, in particular, by Napierala and

Kvetan (2023) in the “Handbook of Computational Social Science for Policy” (Chapter 13). It focuses on changing skills in a dynamic world from a social science perspective. Moreover, Papoutsoglou et al. (2019) focus on studies regarding the software engineering labor market. Besides JPs, they research other sources like social networks or Q&A sites. Lastly, the survey by Khaouja et al. (2021a) on skill identification from JPs is the closest to this survey. It overviews papers using methodologies such as skill counts, topic modeling, skill embeddings, and other machine learning-based methods. With this survey, we steer away from manual and topic modeling approaches to delve deeply into recent extraction methodologies and deep learning-based innovations.

3 Skill-related Terminology

The terms *skill extraction, identification* (Li et al., 2023), *detection* (Beauchemin et al., 2022), *standardization* (Li et al., 2023) and *classification* are used differently, sometimes interchangeably, and describe the same or different tasks. We provide the following definition (See an example in Table 3 in the Appendix):

- **Skill Extraction (E):** as a generic (parent) category for retrieving skill-related information. Skill extraction $E : JP \rightarrow (S)$, where E maps a job posting (JP) to a set of skills S .
- **Skill Identification/Detection (I):** as the process of extracting skills without any predefined labels. It can be represented as $I : JP \rightarrow S$, where skills, especially skill spans, are extracted from JPs. It can also be formalized as a classification problem, $I : \text{Span} \rightarrow \{0, 1\}$, to determine whether a given span in a JP represents a skill (1) or not (0).
- **Skill Extraction with Coarse Labels (E_C):** as identifying broader categories of skill spans. It is formalized as $E_C : JP \rightarrow \{SC_1, SC_2, \dots, SC_n\}$, where each SC_i represents a skill span with a coarse label.
- **Skill Standardization (Std):** as the normalization process of skill terms, formalized as $Std : S \rightarrow S'$, mapping an initial set of skills S to a standardized set S' .
- **Direct Skill Classification (C_D):** as mapping skills to a predefined skill base for assigning fine-grained labels. This process can be formalized as $C_D : S \rightarrow L$, where C_D maps

a set of already extracted skills S to a set of fine-grained labels L .

- **Skill Classification with Extraction (C_E):** as mapping JPs to a predefined skill base for assigning fine-grained labels. This process can be formalized as $C_E : JP \rightarrow L$, where C_E maps a set of already extracted skills S entire JP or raw JP snippets to a set of fine-grained labels L .

Given these definitions, the skill extraction step can happen at different levels of **granularity** (of the input). Some works extract skills per JP (E_{JP} , the overall document), per sentence ($E_{sentence}$) or per n-gram (E_{n-gram}). A skill span (E_{span}) is a continuous n-gram sequence that capture a skill.

A **skill base** (B) is a knowledge base containing skill entities and terminology. A taxonomy is a hierarchically structured skill base, while ontologies provide a structure via relationships between concepts (Khaouja et al., 2021a). Several works use the term “skill dictionary” for a skill base, most often referring to an unstructured skill base or a list of skills (Gugnani and Misra, 2020; Yao et al., 2022). Two popular publicly-available skill bases, created by domain experts, and are frequently used and maintained are the European Skills, Competences, Qualifications and Occupations (ESCO; le Vrang et al., 2014) taxonomy and the US Occupational Information Network (O*NET; Council et al., 2010). We refer to Khaouja et al. (2021a) for more examples of skill bases.

4 What are Skills? On Skill Definitions

Understanding the concept of a *skill* is pivotal in the field of skill extraction. In this section, we investigate several definitions of skills by various publications and institutions, aiming to identify commonalities and distinctions across different sources, which is crucial for establishing a common ground in this emerging field.

The concept of *skill* can be seen as one broad concept (Green et al., 2022; Wild et al., 2021; Fang et al., 2023) or split into subclasses, with multiple possibilities for the split. In the latest version of the ESCO taxonomy the “skill pillar” is divided into four categories: “Transversal skills”, “Skills”, “Knowledge” and “Language skills and knowledge”.¹ O*NET is structured in six domains

¹https://esco.ec.europa.eu/en/classification/skill_main

(Council et al., 2010), the domain most fitting for skill extraction from JP is “Worker Requirements”. This domain entails four subcategories: basic skills, cross-functional skills, knowledge, and education.² But publications considered in this survey that define skills, mainly distinguish between hard and soft skills (Tamburri et al., 2020; Beauchemin et al., 2022; Sayfullina et al., 2018), which is therefore also the separation used in this survey.

Hard Skills Tamburri et al. (2020) delineate hard skills as professional competencies, activities, or knowledge pertinent to organizational functions, processes, and roles, essential for the successful completion of specific tasks. This definition emphasizes the practicality and functionality of hard skills within a professional setting. Aligning with this, the study by Beauchemin et al. (2022) views hard skills as task-oriented technical competencies, drawing upon Lyu and Liu (2021) to define them as formal technical abilities for performing certain tasks. Furthermore, Gugnani and Misra (2020) expand on this perspective by incorporating technological terminologies for skill identification and therefore integrating knowledge as a fundamental component of hard skills.

By incorporating knowledge as a component of hard skills, the definitions of hard skills and knowledge categories of O*NET and ESCO can be combined. O*NET’s definition of hard skills states that they are developed abilities that enable learning or knowledge acquisition, coupled with their definition of knowledge as “Organized sets of principles and facts applying in general domains”.³ This comprehensive definition underscores not only technical proficiency but also the ability to adapt and apply knowledge. Similarly, ESCO, referencing the European Qualifications Framework, defines skills as “the ability to apply knowledge and use know-how to complete tasks and solve problems”, while defining knowledge as “the outcome of the assimilation of information through learning”.⁴

In conclusion, we define hard skills as a wide variety of professional abilities, ranging from measurable technical skills to the more general capacity for learning and effectively applying knowledge. They are quantifiable and teachable competencies,

predominantly technical, yet intrinsically linked to the ability to adapt and apply them in diverse professional scenarios.

Soft Skills Sayfullina et al. (2018), referencing the Collins dictionary (HarperCollins Publishers, 2023), views soft skills as innate, non-technical qualities highly sought after in employment, diverging from reliance on acquired knowledge. In a more social context, Tamburri et al. (2020) characterizes soft skills as encompassing personal, emotional, social, or intellectual aspects, further known as behavioral skills or competencies. Echoing this sentiment, Beauchemin et al. (2022), drawing from Lyu and Liu (2021), identifies soft skills as a variety of personal attributes and behaviors crucial for effective workplace interaction, collaboration, and adaptability.

Adding to these perspectives, ESCO characterizes soft skills as *transversal skills*, highlighting their wide applicability across various occupations and sectors and their fundamental role in individual growth.⁵ Similarly, O*NET classifies these skills under Cross-Functional Skills, defining them as developed capacities that enhance the performance of activities common across different jobs, encompassing areas like Social Skills and Complex Problem Solving Skills.⁶ Both sources underscore the universal relevance of soft skills.

These previous definitions lead to our converged definition that soft skills cover a vast array of personal, social, and intellectual competencies, all of which are indispensable for successful interpersonal engagement and personal development in professional settings.

5 Operationalization of Skill Definitions

In this section, we explore various methodologies for operationalizing skill definitions in skill extraction and classification research.

Using a Skill Base By using a given skill base, a pre-defined definition of the concept of skills is provided by the authors of the skill base. Numerous studies employ established skill bases such as the ESCO taxonomy (Zhang et al., 2023, 2022b; Clavié and Soulié, 2023; Decorte et al., 2023, 2022) or O*NET (Gugnani and Misra, 2020). However, it is often ambiguous whether these studies use

² <https://www.onetcenter.org/content.html>

³ See footnote 2.

⁴ <https://esco.ec.europa.eu/en/about-esco/escopedia/escopedia/knowledge> and <https://esco.ec.europa.eu/en/about-esco/escopedia/escopedia/skill>

⁵ <https://esco.ec.europa.eu/en/about-esco/escopedia/escopedia/transversal-knowledge-skills-and-competences>

⁶ See footnote 2.

all or only specific subcategories (Li et al., 2023; Decorte et al., 2022; Gugnani and Misra, 2020). Some papers mention explicitly the use of all subclasses (Zhang et al., 2022b,a; Gnehm et al., 2022a) other times it can be inferred from the number of skill spans used (Clavié and Soulié, 2023; Decorte et al., 2023). However, one should note that the interpretations of ESCO definitions differ based on the ESCO version and authors’ perspective. Zhang et al. (2022a,b) used ESCO version 1.0 with a different soft skill category than discussed in Section 4 and implemented two labels: “knowledge” aligns with ESCO’s “Knowledge” category, and “Skills” as a fusion of the hard and soft skills. In contrast, Colombo et al. (2019) using the same ESCO version, but treat soft skills separate from hard skills. Most of the publications used all subcategories as skills without differentiating (Clavié and Soulié, 2023; Gnehm et al., 2022a; Decorte et al., 2023).

Beyond these, there are other skill bases, such as the Russian professional standard in Botov et al. (2019) or the Chinese Occupation Classification Grand Dictionary used in Cao and Zhang (2021); Cao et al. (2021). Additionally, non-official skill bases exist, like the list of 1K soft skills in (Sayfullina et al., 2018) or LinkedIn’s in-house taxonomy for skill extraction (Shi et al., 2020). In general, for transparency and reproducibility, it is helpful to state which subset of fine-grained labels L of the skill base (B) and which skill base version is used.

Leveraging Automated Tools Some studies leverage automated tools like AutoPhrase (Shang et al., 2018) or Microsoft Azure Analytics Service for NER for initial skill term detection, followed by manual verification and refinement (Yao et al., 2022; Kortum et al., 2022). Also Vermeer et al. (2022) extract parts of their training data using an automated tool, while others are taken from a skill base.⁷ Lastly, Gugnani and Misra (2020) employ an IBM tool for skill identification, which forms a part of a larger skill identification framework.⁸ While some previous work did not apply manual verification (Gugnani and Misra, 2020; Vermeer et al., 2022), we recommend it to reduce automation bias from the tool impacting the data.

Definition through Labeling Domain experts play a crucial role for labeling data and therefore impact how the definition of skills is put into work

⁷<https://www.textkernel.com/de/>

⁸<https://www.ibm.com/products/natural-language-understanding>

(Shi et al., 2020; Tamburri et al., 2020; Beauchemin et al., 2022). Tamburri et al. (2020) additionally provide a codebook with skill definitions to address ambiguities. Shi et al. (2020) used next skills identified by hiring experts and skills common among successful applicants as training data. The study by Bhola et al. (2020) treat the companies filing the JPs as domain experts by using their labels (see also Section 6). Besides domain experts, crowd workers and the people writing the guidelines for the workers oftentimes determine which terms are skills. Some studies do not mention who labels the data (Wild et al., 2021; Cao and Zhang, 2021; Botov et al., 2019). We suggest being clear about the labeling process and guidelines, making them public for transparency and re-use/standardization, and using domain experts if possible for accurate labeling.

6 Data

In this section, we provide a comprehensive description of publicly available datasets, with an overview in Table 1.

SAYFULLINA by Sayfullina et al. (2018) is a dataset derived from a publicly available Kaggle dataset, containing JPs from within the UK and representing a variety of sectors.⁹ The authors retrieved soft skill spans by exact matching with a list of 1,072 soft skills. Each identified span is accompanied by up to 10 surrounding words. Crowdsourcing was used to determine whether the highlighted skill belongs to a job applicant. To ensure reliability, the workers were tested on a small set of JPs and each snippet was evaluated by at least three workers. This process led to a dataset with high class imbalance due to more positive examples. To counter this, additional skill spans were added, including those usually not describing candidates (marked as negative) and those consistently labeled positive.

GREEN by Green et al. (2022) uses the same Kaggle dataset as SAYFULLINA. The labeling was done via crowdsourcing, they did not use experts but only workers who passed a test were included, and encouraged to follow the guidelines. Apart from the “Skill” label capturing hard and soft skills, the labels “Occupation”, “Domain”, “Experience”, “Qualification”, and “None” are used in a

⁹https://www.kaggle.com/datasets/airiddha/trainrev1?select=Train_rev1.csv



Publication	Approach	Granularity	Skill type	Use case	Size	
(Sayfullina et al., 2018)	Crowdsourced	span-level	soft	I	7411 spans	✗
(Green et al., 2022)	Crowdsourced	span-level	hard + soft	E_C	10,606 spans	✓
(Beauchemin et al., 2022)	Expert	span-level	soft	E_C	47 JPs - 932 spans	✗
(Zhang et al., 2022a)	Expert	span-level	hard + soft	E_C	265 JP - 9,633 spans	✓
(Zhang et al., 2022b)	Expert	span-level	hard + soft	E_C+C_D	60 JP - 920 spans	✓
(Decorte et al., 2022)	Manual	span-level	hard + soft	$I+C_D$	1,618 spans	✓
(Gnehm et al., 2022b)	Expert	span-level	hard + soft	E_C+C_D	10,995 spans	✗
(Bhola et al., 2020)	Skill Inventory	document-level	unknown	C_E	20,298 JP	✗

Table 1: Overview of publicly-available labeled datasets.  indicates if the authors used guidelines (not necessarily publicly available).

BIO scheme. The authors reduced errors by label aggregation with a preference towards labels from higher-performing workers. Additionally, they reclassified specific “Experience” spans, as “Skill” spans, and manually split multi-term spans into separate spans.

FIJO by Beauchemin et al. (2022) was created in partnership with Canadian insurance companies, and consists of cleaned and de-identified French JPs published between 2009 and 2020. The dataset focus on soft skills and includes 867 JPs with 47 annotated JPs, selected and annotated by a domain expert. The annotated spans are unevenly distributed across four classes: “Thoughts”, “Results”, “Relational”, and “Personal”.

SKILLSPAN by Zhang et al. (2022a) consists of the anonymized raw data and annotations of skill and knowledge spans from three JP datasets, one of which cannot be made publicly available due to its license. The available datasets are:

- **HOUSE**: A static in-house dataset with different types of JPs from 2012-2020 and
- **TECH**: The StackOverflow JP platform, consisting mostly of technical jobs collected between June 2020 and September 2021.

The development of the publicly available annotation guidelines involved an iterative process, starting with a few JPs and progressing through several rounds of annotation and refinement by three domain experts.

KOMPETENCER by Zhang et al. (2022b) consists of Danish JPs with annotated skill and knowledge spans, see Table 4 in the Appendix. The same skill definitions, guidelines, and metrics as in SKILLSPAN are used for annotation. This dataset can be used for skill extraction with coarse labels, but the authors have also added fine-grained annotations to evaluate a classification with the ESCO

taxonomy. For fine-grained annotations, they query the ESCO API with the annotated spans and use Levenshtein distance to determine the relevance of each obtained label. Then, the quality of these distantly supervised labels is assessed through human evaluation. They also repeated this process for the English SKILLSPAN dataset but only manually checked a sample for calculating statistics.

DECORTE by Decorte et al. (2022) is a variant of the SKILLSPAN dataset with annotated ESCO labels. They used the identified skill without the skill and knowledge labels, but they can be recreated by matching the dataset with SKILLSPAN, see Table 4 in the Appendix. Unlike in KOMPETENCER they manually matched the skills with fitting ESCO labels (if they exist) to create a gold standard.

GNEHM-ICT by Gnehm et al. (2022b) is a Swiss-German dataset where they annotated for Information and Communications Technology (ICT)-related entity recognition. These could be ICT tasks, technology stack, responsibilities, and so forth. The used dataset is a combination of two other Swiss datasets namely the Swiss Job Market Monitor and an online job ad dataset (Gnehm and Clematide, 2020; Buchmann et al., 2022). There are around 25,000 sentences in the dataset.

BHOLA by Bhola et al. (2020) was obtained from a government website¹⁰ in Singapore. The preprocessing steps for this English language dataset include converting text to lowercase and removing stop words and rarely used words. The companies filing the JPs added skill labels, which are mapped to the whole JP document. This makes the dataset suitable for performing multi-label classification by predicting a set of required skills for a given JP.

¹⁰<https://www.mycareersfuture.gov.sg/>.

7 Methods

In this section, we survey methods for skill extraction and classification. As in Section 3 the goal of the extraction is to identify skill spans with (EC) or without coarse labels (I). The classification section covers direct classification methods (CD) and classification methods with extraction (CE), both aim to retrieve fine-grained skill labels.

7.1 Skill Extraction

This chapter delineates the evolution of skill extraction methodologies, grouped into three categories: skill identification as span labeling, skill identification through binary classification, and skill extraction with coarse span labels. Starting with LSTM neural networks in 2018 the methods in all three sub-chapters used after the introduction of BERT (Devlin et al., 2019) in 2019 heavily BERT and BERT-based models. Recent advancements continue to diversify the landscape, integrating a broader array of language models (LMs).

7.1.1 Skill Identification as Span Labeling

In this category approach skill identification as a span labeling task. The primary objective is to accurately identify skill spans, encompassing both the identification of the relevant skill phrases and their precise boundaries. Jia et al. (2018) are the first to use sequence tagging for identifying skills from JPs in 2018. The authors use a pre-trained LSTM neural network (Lample et al., 2016) for identifying skill terms on the word-level. Tamburri et al. (2020) also employed binary classification, but at the sentence-level, using a Dutch JP dataset. Their best-performing model, BERT Multilingual Cased, was fine-tuned on expert-annotated JP sentences, suggesting potential improvement with more data and optimization. Further publications retrieve embeddings using a pre-trained BERT model (Wild et al., 2021; Cao and Zhang, 2021; Cao et al., 2021). Notably, Cao et al. (2021) and Cao and Zhang (2021) combine BERT’s pre-trained vectors with a Bi-LSTM and a CRF layer for finer entity classification. This approach aligns with previous research demonstrating the efficacy of a CRF layer in NER tasks (Souza et al., 2020). In Zhang et al. (2023), they further built upon the domain-adaptive pre-training paradigm (Gururangan et al., 2020). They make use of the ESCO taxonomy (le Vrang et al., 2014) and integrate this in a multilingual XLM-R model (Conneau et al., 2020),

using this taxonomy-driven pre-training method, they introduce a new state-of-the-art for all skill identification benchmarks. For analysis, they show that performance increases especially for skills that are shorter in length, due to ESCO skills also being shorter.

In contrast to these single-model approaches, Gugnani and Misra (2020) adopted a multi-faceted methodology to predict the relevance of identified skill spans. Their methodology encompassed four modules: using part-of-speech (PoS) tagging, parsing sentences with skill bases (O*NET, Hope, and Wikipedia), leveraging a ready-made sequence tagging solution, and employing a pre-trained word2vec model for final score determination through cosine similarity.¹¹

7.1.2 Skill Identification as binary Classification Task

In this category, skill identification is framed as a binary classification task. The focus is on determining whether a given sequence either constitutes or contains a (specific) skill. The task in Sayfullina et al. (2018) differs from the other publications. They extract skill spans by exact match and aim to decide whether skill spans refer to a candidate or something else, like a company. They experiment with various classifiers and input representations, such as Soft Skill Masking, Embedding, and Tagging, finding the LSTM classifier with skill tagging most effective on their dataset. Tamburri et al. (2020) employed binary classification at the sentence-level to determine if it contains a skill. Their best-performing model, BERT Multilingual Cased, was fine-tuned on expert-annotated JP sentences using a Dutch JP dataset. Yao et al. (2022) classify individual words as skill-related or not. They split JPs into individual words, analyzing each through character-level and word-level encoders, integrating linguistic features like POS tags and capitalization. Their initial training employs AutoPhrase (Shang et al., 2018) for automatic skill term identification, followed by manual verification and expert-labeled samples. The model is further refined using Positive-Unlabeled learning, where the classifier’s predictions on unlabeled data help expand the skill base for continuous adaptation.

¹¹<https://www.ibm.com/products/natural-language-understanding>.

Paper	Model	Skill Type	Granularity	Use Case
(Fang et al., 2023)	Custom pre-trained LM	soft + hard	word-level	C_E
(Goyal et al., 2023)	FastText skip-gram, GNN	unknown	word-level	C_E
(Clavié and Soulié, 2023)	GPT-4	soft + hard	span-level	C_E
(Li et al., 2023)	XMLC - LLM	soft + hard	document-level	C_E
(Decorte et al., 2023)	GPT-3.5	soft + hard	sentence-level	C_E
(Zhang et al., 2023)	Multilingual XLM-R	soft + hard	span-level	E_C
(Decorte et al., 2022)	RoBERTa	soft + hard	sentence-level	C_E
(Zhang et al., 2022c)	RoBERTa, JobBERT	soft + hard	span-level	C_D
(Gnehm et al., 2022a)	JobBERT-de, SBERT	soft + hard	span-level	$E_C + C_D$
(Zhang et al., 2022b)	BERTbase, DaBERT	soft + hard	span-level	C_E
(Beauchemin et al., 2022)	Bi-LSTM, CamemBERT	soft	span-level	E_C
(Yao et al., 2022)	BERT, word2vec	unknown	word-level	I
(Anand et al., 2022)	LaBSE model	soft + hard	title	C_E
(Vermeer et al., 2022)	RobBERT	soft + hard	document-level	C_E
(Wild et al., 2021)	BERT, spaCy	soft + hard	span-level	I
(Khaouja et al., 2021b)	Sent2vec, SBERT	soft + hard	sentence-level	C_E
(Cao et al., 2021)	BERT-BiLSTM-CRF	soft + hard	span-level	I
(Cao and Zhang, 2021)	BERT-BiLSTM-CRF	soft + hard	span-level	I
(Li et al., 2020)	Deep Averaging Network, FastText	unknown	span-level	C_E
(Tamburri et al., 2020)	BERT Multilingual Cased	soft + hard	sentence-level	I
(Bhola et al., 2020)	BERTbase	unknown	document-level	C_E
(Gugnani and Misra, 2020)	Word2vec	soft + hard	span-level	I
(Botov et al., 2019)	Word2vec	unknown	span-level	C_E
(Jia et al., 2018)	LSTM	unknown	word-level	I
(Sayfullina et al., 2018)	CNN, LSTM, HAN	soft	span-level	I
(Javed et al., 2017)	Word2vec	soft + hard	span-level	C_E

Table 2: Publications regarding neural skill extraction and classification. The skill type was not always explicitly mentioned in some cases it’s derived from examples given in the paper.

7.1.3 Skill Extraction with Coarse Labels

This section explores advancements in skill extraction with coarse labels, where each publication extract spans from two to four different categories. The studies of Gnehm et al. (2022a) and Zhang et al. (2022a) both utilize sequence tagging-based models. Gnehm et al. (2022a) focusing on iterative training and annotation with jobBERT-de, a German LM tailored for JPs. Zhang et al. (2022a) compare BERT-based (Devlin et al., 2019) and SpanBERT-based (Joshi et al., 2020) models, highlighting the importance of domain adaptation. On the other hand, Beauchemin et al. (2022) and Fang et al. (2023) delve into the intricacies of training and optimizing LMs for skill extraction. Beauchemin et al. (2022) examine the sensitivity of Bi-LSTM and CamemBERT (Martin et al., 2020) models to training data volume, with CamemBERT unfrozen yielding the highest mean token-wise accuracy. Fang et al. (2023) introduce RecruitPro, a specialized model for skill extraction from recruitment texts, employing innovative techniques for dealing with data noise and label imbalances. Collectively, these papers emphasize the need for tailored approaches and continuous innovation in model development.

7.2 Skill Classification

While skill standardization can be achieved through classification, other methods such as clustering (Bernabé-Moreno et al., 2019; Lukauskas et al., 2023), matching n-grams based on string similarity (Boselli et al., 2018), or identifying semantically similar skills (Bernabé-Moreno et al., 2019; Colombo et al., 2019; Gröger and Schneider, 2019) also lead to standardized skill spans. These methods simplify the variety and quantity of skill spans without assigning standardized labels. Transitioning from these methods, we now focus on skill classification, a crucial step for assigning standardized labels to effectively organize and understand skills. Most publications skip a traditional extraction and match the JPs directly to the skill base (C_E), which can be seen as skill extraction against a skill base. Exceptions are Gnehm et al. (2022a), which perform extraction of skill spans with coarse labels before the fine-grained classification step, and Zhang et al. (2022b) who rely on prior work for extraction and focus solely on the matching of skill spans to ESCO (C_D). We divide the publications by methodology into those that match based on semantic similarity and those using extreme multi-label classification to solve the matching task.

7.2.1 Similarity-based Approaches

The publications with similarity-based approaches split the JPs into sentences or n-grams before matching them. All of the following publications use skill embedding methods, which can be seen as an advancement of the skill count methods (Section 1). The advances in text embeddings over time are reflected in the scope of the approaches. While [Javed et al. \(2017\)](#) and [Botov et al. \(2019\)](#) improve the matching using word2vec embeddings ([Mikolov et al., 2013](#)), later [Li et al. \(2020\)](#) use FastText ([Bojanowski et al., 2017](#)) leveraging sub-word information to handle out-of-vocabulary words and capture more detailed semantic and syntactic information. [Khaouja et al. \(2021b\)](#) compare using sent2vec trained on Wikipedia sentences, and SBERT ([Reimers and Gurevych, 2019](#)) trained on millions of paraphrase sentences for embeddings. Moreover, [Zhang et al. \(2022c\)](#) uses LMs like RoBERTa and JobBERT to match n-grams from JP sentences with the ESCO taxonomy. They also experiment with context and frequency-aware embeddings. [Gnehm et al. \(2022a\)](#) performed direct skill extraction using context-aware embeddings and the SBERT model similar to [Zhang et al. \(2022c\)](#), additionally they contextualize skill areas within spans and ontology terms using their hierarchical structure. The study explores techniques to enhance BERT model similarity, including in-domain pretraining, transformer-based sequential denoising auto-encoder (TSDAE; [Wang et al., 2021](#)) for domain-specific terminology, and Siamese BERT Networks for training sentence embeddings ([Reimers and Gurevych, 2019](#)). They further leverage MNR loss in Siamese networks ([Henderson et al., 2017](#)), using ontology data to create positive text pairings for better label matching. SkillGPT ([Li et al., 2023](#)) is the first tool to use an LLM for the matching task, they convert ESCO entries into structured documents, which are vectorized by the LM. Then, they summarize the input text, and use an embedding of the summary to retrieve the closest ESCO entries.

7.2.2 Extreme Multi-label Classification Approaches

[Bhola et al. \(2020\)](#) were the first to formulate skill extraction against a skill base as an extreme multi-label classification (XMLC). They classify multiple skill labels per document using the labels of the BHOLA dataset (around 2500 labels) as a skill base. Their BERT–XMLC framework, involves a Text

Encoder that uses the pre-trained BERTbase model to convert JP texts into dense vector representations, a Bottleneck Layer that reduces overfitting by compressing these representations ([Liu et al., 2017](#)) and subsequently a fully connected layer for multi-label classification of the skills. Enhancements include focusing on semantic skill label representation and skill co-occurrence, using bootstrapping to augment training data, and improve skill correlation capture. Their model outperformed XMLC baselines. [Vermeer et al. \(2022\)](#) adapted this approach for using RobBERT and additional linear layers, validating on BHOLA and a non-public Dutch dataset. Similarly, [Anand et al. \(2022\)](#) extended the model to predict skill importance using LaBSE-encoded ([Feng et al., 2022](#)) job titles, ranking skills from an in-house database based on a 0-1 scale of importance.

Subsequent publications have concentrated on XMLC for skill extraction and classification using the ESCO taxonomy with around 13000 labels. For a pure skill classification for already identified skill spans [Zhang et al. \(2022b\)](#) use distant supervision by querying the ESCO API for the fine-grained skill labels. For model training, they employ zero-shot cross-lingual transfer learning techniques using various BERT models and fine-tune them on Danish JPs. The effectiveness of the models is tested on an adapted version of SKILLSPAN and KOMPETENCER. The same year [Decorte et al. \(2022\)](#) addressed the XMLC task on the sentence-level, again using distant supervision with the ESCO taxonomy. They enhance binary skill classifier training with three negative sampling strategies, involving siblings in ESCO hierarchy, Levenshtein distance, and cosine similarity of RoBERTa-encoded skill names. Their model employs a frozen pre-trained RoBERTa with mean pooling for sentence representation, followed by separate binary classifiers for each skill, evaluated on DECORTE.

As for the similarity-based approaches, LLMs are prominent in recent XMLC approaches. Unlike [Li et al. \(2023\)](#), [Decorte et al. \(2023\)](#) use the LLM solely during training to reduce latency and enhance reproducibility. They create a synthetic training dataset using the LLM, then optimize a bi-encoder through contrastive training, to effectively represent both skill names and corresponding sentences in close proximity within the same space. This method outperforms the distance supervision baseline by [Decorte et al. \(2022\)](#) (see Table 5). Sim-

ilarly, Clavié and Soulié (2023) treat the skill extraction and classification task as individual binary classification problems, using GPT-3.5 like Decorte et al. (2023) but generating more spans per skill for synthetic training. They propose two extraction methods: one using linear classifiers for each skill, employing hard negative sampling (Robinson et al., 2021) for improved skill differentiation, and another based on similarity, utilizing E5-LARGE-V2 embeddings (Wang et al., 2022) for cosine similarity calculations between JP extracts and ESCO labels or synthetic sentences. Potential skills are then reranked using an LLM. In evaluations using the DECORTE dataset, their methods achieved high performance with GPT-4, though results with GPT-3.5 were lower than Decorte et al. (2023), see Table 5 in the Appendix.

Goyal et al. (2023) present JobXMLC, a unique framework for the XMLC task, distinct from the prevailing methods. JobXMLC integrates a job-skill graph to represent job-skill interconnections, utilizes a GNN for multi-hop embeddings from the graph’s structure, and incorporates an extreme classification system with skill attention based on skill frequency in the dataset. The framework’s effectiveness is validated on the BHOLA and a proprietary StackOverflow dataset, see Table 5 in the Appendix.

8 Conclusions and Future Directions

Recent publications indicate two emerging trends in skill extraction. Firstly, extracting skills against skill bases like ESCO is gaining popularity, facilitating cross-industry and regional comparisons. Secondly, LLMs are increasingly applied in skill extraction and classification, proving particularly advantageous due to the scarcity of training data in this domain.

Future research in skill extraction and classification could focus on emerging skills and the extraction of implicit skills. Methods like those by Javed et al. (2017) and Khaouja et al. (2021b) update skill bases with emerging technologies and frequently used keywords, but evaluating these remains difficult without a standard benchmark. The challenge of extracting implicit skills, not directly stated in job postings, is also gaining attention. Techniques include prompting LLMs to generate training data with implied skills (Clavié and Soulié, 2023) and using complete sentences to encompass both explicit and implicit skills (Decorte et al., 2022, 2023).

However, these methods need thorough evaluation, presenting an open field for future exploration.

Limitations

A limitation that should be considered is that only publications in the English language (although data was from multiple languages) were surveyed in this paper. Second, to allow for a deeper focus publications regarding topic modeling were excluded even if they used deep-learning-based methods.

Acknowledgements

We thank the reviewers for their insightful feedback. ES acknowledges financial support with funds provided by the German Federal Ministry for Economic Affairs and Climate Action due to an enactment of the German Bundestag under grant 46SKD127X (GENESIS). MZ is supported by the Independent Research Fund Denmark (DFF) grant 9131-00019B and BP is supported by ERC Consolidator Grant DIALECT 101043235.

References

- Sarthak Anand, Jens-Joris Decorte, and Niels Lowie. 2022. *Is it required? ranking the skills required for a job-title*.
- Ziqiao Ao, Gergely Horváth, Chunyuan Sheng, Yifan Song, and Yutong Sun. 2023. *Skill requirements in job advertisements: A comparison of skill-categorization methods based on wage regressions*. *Information Processing & Management*, 60(2):103185.
- David Beauchemin, Julien Laumonier, Yvan Le Ster, and Marouane Yassine. 2022. *"fijo": a french insurance soft skill detection dataset*.
- Juan Bernabé-Moreno, Álvaro Tejada-Lorente, Julio Herce-Zelaya, Carlos Porcel, and Enrique Herrera-Viedma. 2019. *An automatic skills standardization method based on subject expert knowledge extraction and semantic matching*. *Procedia Computer Science*, 162:857–864. 7th International Conference on Information Technology and Quantitative Management (ITQM 2019): Information technology and quantitative management based on Artificial Intelligence.
- Akshay Bhola, Kishaloy Halder, Animesh Prasad, and Min-Yen Kan. 2020. *Retrieving skills from job descriptions: A language model based extreme multi-label classification framework*. In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 5832–5842, Barcelona, Spain (Online). International Committee on Computational Linguistics.

- Piotr Bojanowski, Edouard Grave, Armand Joulin, and Tomas Mikolov. 2017. [Enriching word vectors with subword information](#). *Transactions of the Association for Computational Linguistics*, 5:135–146.
- Roberto Boselli, Mirko Cesarini, Fabio Mercorio, and Mario Mezzanica. 2018. [Classifying online job advertisements through machine learning](#). *Future Generation Computer Systems*, 86:319–328.
- Dmitriy Botov, Julius Klenin, Andrey Melnikov, Yuri Dmitrin, Ivan Nikolaev, and Mikhail Vinel. 2019. [Mining labor market requirements using distributional semantic models and deep learning](#). In *Business Information Systems - 22nd International Conference, BIS 2019, Seville, Spain, June 26-28, 2019, Proceedings, Part II*, volume 354 of *Lecture Notes in Business Information Processing*, pages 177–190. Springer.
- Marlis Buchmann, Helen Buchs, Felix Busch, Simon Clematide, Ann-Sophie Gnehm, and Jan Müller. 2022. Swiss job market monitor: A rich source of demand-side micro data of the labour market. *European Sociological Review*.
- Lina Cao and Jian Zhang. 2021. [Skill requirements analysis for data analysts based on named entities recognition](#). In *2021 2nd International Conference on Big Data and Informatization Education (ICB-DIE)*, pages 64–68.
- Lina Cao, Jian Zhang, Xinquan Ge, and Jindong Chen. 2021. [Occupational profiling driven by online job advertisements: Taking the data analysis and processing engineering technicians as an example](#). *PLoS ONE*, 16.
- Benjamin Clavié and Guillaume Soulié. 2023. [Large language models as batteries-included zero-shot esco skills matchers](#).
- Emilio Colombo, Fabio Mercorio, and Mario Mezzanica. 2019. [Ai meets labor market: Exploring the link between automation and skills](#). *Information Economics and Policy*, 47:27–37. The Economics of Artificial Intelligence and Machine Learning.
- Alexis Conneau, Kartikay Khandelwal, Naman Goyal, Vishrav Chaudhary, Guillaume Wenzek, Francisco Guzmán, Edouard Grave, Myle Ott, Luke Zettlemoyer, and Veselin Stoyanov. 2020. [Unsupervised cross-lingual representation learning at scale](#). In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 8440–8451, Online. Association for Computational Linguistics.
- National Research Council, Nancy Thomas Tippins, Margaret L Hilton, et al. 2010. *A database for a changing economy: Review of the Occupational Information Network (O*NET)*. National Academies Press.
- Jens-Joris Decorte, Jeroen Van Haute, Johannes Deleu, Chris Develder, and Thomas Demeester. 2022. [Design of negative sampling strategies for distantly supervised skill extraction](#). In *Proceedings of the 2nd Workshop on Recommender Systems for Human Resources (RecSys-in-HR 2022)*, volume 3218, page 7. CEUR.
- Jens-Joris Decorte, Severine Verlinden, Jeroen Van Haute, Johannes Deleu, Chris Develder, and Thomas Demeester. 2023. [Extreme multi-label skill extraction training using large language models](#).
- Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. [BERT: Pre-training of deep bidirectional transformers for language understanding](#). In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.
- Chuyu Fang, Chuan Qin, Qi Zhang, Kaichun Yao, Jingshuai Zhang, Hengshu Zhu, Fuzhen Zhuang, and Hui Xiong. 2023. [Recruitpro: A pretrained language model with skill-aware prompt learning for intelligent recruitment](#). In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, KDD '23*, page 3991–4002, New York, NY, USA. Association for Computing Machinery.
- Fangxiaoyu Feng, Yinfei Yang, Daniel Cer, Naveen Ariavazhagan, and Wei Wang. 2022. [Language-agnostic BERT sentence embedding](#). In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 878–891, Dublin, Ireland. Association for Computational Linguistics.
- Ann-sophie Gnehm, Eva Bühlmann, Helen Buchs, and Simon Clematide. 2022a. [Fine-grained extraction and classification of skill requirements in German-speaking job ads](#). In *Proceedings of the Fifth Workshop on Natural Language Processing and Computational Social Science (NLP+CSS)*, pages 14–24, Abu Dhabi, UAE. Association for Computational Linguistics.
- Ann-Sophie Gnehm, Eva Bühlmann, and Simon Clematide. 2022b. [Evaluation of transfer learning and domain adaptation for analyzing German-speaking job advertisements](#). In *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, pages 3892–3901, Marseille, France. European Language Resources Association.
- Ann-Sophie Gnehm and Simon Clematide. 2020. [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*, pages 83–93, Online. Association for Computational Linguistics.
- Nidhi Goyal, Jushaan Kalra, Charu Sharma, Raghava Mutharaju, Niharika Sachdeva, and Ponnurangam

- Kumaraguru. 2023. **JobXMLC: EXtreme multi-label classification of job skills with graph neural networks**. In *Findings of the Association for Computational Linguistics: EACL 2023*, pages 2181–2191, Dubrovnik, Croatia. Association for Computational Linguistics.
- Thomas Green, Diana Maynard, and Chenghua Lin. 2022. **Development of a benchmark corpus to support entity recognition in job descriptions**. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, pages 1201–1208, Marseille, France. European Language Resources Association.
- Joscha Gröger and Georg Schneider. 2019. **Automated analysis of job requirements for computer scientists in online job advertisements**. In *Proceedings of the 15th International Conference on Web Information Systems and Technologies, WEBIST 2019*, page 226–233, Setubal, PRT. SCITEPRESS - Science and Technology Publications, Lda.
- Akshay Gugnani and Hemant Misra. 2020. **Implicit skills extraction using document embedding and its use in job recommendation**. In *The Thirty-Fourth AAAI Conference on Artificial Intelligence, AAAI 2020, The Thirty-Second Innovative Applications of Artificial Intelligence Conference, IAAI 2020, The Tenth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2020, New York, NY, USA, February 7-12, 2020*, pages 13286–13293. AAAI Press.
- Suchin Gururangan, Ana Marasović, Swabha Swayamdipta, Kyle Lo, Iz Beltagy, Doug Downey, and Noah A. Smith. 2020. **Don't stop pretraining: Adapt language models to domains and tasks**. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 8342–8360, Online. Association for Computational Linguistics.
- HarperCollins Publishers. 2023. **Collins COBUILD Advanced Learner's Dictionary: Soft Skills**. <https://www.collinsdictionary.com/dictionary/english/soft-skills>.
- Matthew Henderson, Rami Al-Rfou, Brian Strope, Yunhsuan Sung, Laszlo Lukacs, Ruiqi Guo, Sanjiv Kumar, Balint Miklos, and Ray Kurzweil. 2017. **Efficient natural language response suggestion for smart reply**.
- Faizan Javed, Phuong Hoang, Thomas Mahoney, and Matt McNair. 2017. **Large-scale occupational skills normalization for online recruitment**. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 31, pages 4627–4634.
- Shanshan Jia, Xiaoran Liu, Ping Zhao, Chang Liu, Lianying Sun, and Tao Peng. 2018. **Representation of job-skill in artificial intelligence with knowledge graph analysis**. In *2018 IEEE symposium on product compliance engineering-asia (ISPCE-CN)*, pages 1–6. IEEE.
- Mandar Joshi, Danqi Chen, Yinhan Liu, Daniel S. Weld, Luke Zettlemoyer, and Omer Levy. 2020. **SpanBERT: Improving pre-training by representing and predicting spans**. *Transactions of the Association for Computational Linguistics*, 8:64–77.
- Imane Khaouja, Ismail Kassou, and Mounir Ghogho. 2021a. **A survey on skill identification from online job ads**. *IEEE Access*, 9:118134–118153.
- Imane Khaouja, Ghita Mezzour, and Ismail Kassou. 2021b. **Unsupervised skill identification from job ads**. In *2021 IEEE 22nd International Conference on Information Reuse and Integration for Data Science (IRI)*, pages 147–151.
- Henrik Kortum, Jonas Rebstadt, and Oliver Thomas. 2022. **Dissection of AI job advertisements: A text mining-based analysis of employee skills in the disciplines computer vision and natural language processing**. In *55th Hawaii International Conference on System Sciences, HICSS 2022, Virtual Event / Maui, Hawaii, USA, January 4-7, 2022*, pages 1–10. ScholarSpace.
- Guillaume Lample, Miguel Ballesteros, Sandeep Subramanian, Kazuya Kawakami, and Chris Dyer. 2016. **Neural architectures for named entity recognition**. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 260–270, San Diego, California. Association for Computational Linguistics.
- Martin le Vrang, Agis Papantoniou, Erika Pauwels, Pieter Fannes, Dominique Vandestein, and Johan De Smedt. 2014. **Esco: Boosting job matching in europe with semantic interoperability**. *Computer*, 47(10):57–64.
- Nan Li, Bo Kang, and Tijl De Bie. 2023. **Skillgpt: a restful api service for skill extraction and standardization using a large language model**.
- Shan Li, Baoxu Shi, Jaewon Yang, Ji Yan, Shuai Wang, Fei Chen, and Qi He. 2020. **Deep job understanding at linkedin**. In *Proceedings of the 43rd International ACM SIGIR conference on research and development in Information Retrieval, SIGIR 2020, Virtual Event, China, July 25-30, 2020*, pages 2145–2148. ACM.
- Jingzhou Liu, Wei-Cheng Chang, Yuexin Wu, and Yiming Yang. 2017. **Deep learning for extreme multi-label text classification**. In *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval, Shinjuku, Tokyo, Japan, August 7-11, 2017*, pages 115–124. ACM.
- Mantas Lukauskas, Viktorija Šarkauskaitė, Vaida Pilinkienė, Alina Stundziene, Andrius Grybauskas, and Jurgita Bruneckienė. 2023. **Enhancing skills demand understanding through job ad segmentation using nlp and clustering techniques**. *Applied Sciences*, 13.

- Wenjing Lyu and Jin Liu. 2021. [Soft skills, hard skills: What matters most? evidence from job postings](#). *Applied Energy*, 300:117307.
- Louis Martin, Benjamin Muller, Pedro Javier Ortiz Suárez, Yoann Dupont, Laurent Romary, Éric de la Clergerie, Djamé Seddah, and Benoît Sagot. 2020. [CamemBERT: a tasty French language model](#). In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 7203–7219, Online. Association for Computational Linguistics.
- Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. [Efficient estimation of word representations in vector space](#).
- Joanna Napierala and Vladimír Kvetan. 2023. [Changing Job Skills in a Changing World](#), pages 243–259. Springer International Publishing, Cham.
- Maria Papoutsoglou, Apostolos Ampatzoglou, Nikolaos Mittas, and Lefteris Angelis. 2019. [Extracting knowledge from on-line sources for software engineering labor market: A mapping study](#). *IEEE Access*, 7:157595–157613.
- Nils Reimers and Iryna Gurevych. 2019. [SentenceBERT: Sentence embeddings using Siamese BERT-networks](#). In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 3982–3992, Hong Kong, China. Association for Computational Linguistics.
- Joshua David Robinson, Ching-Yao Chuang, Suvrit Sra, and Stefanie Jegelka. 2021. [Contrastive learning with hard negative samples](#). In *9th International Conference on Learning Representations, ICLR 2021, Virtual Event, Austria, May 3-7, 2021*. OpenReview.net.
- Luiza Sayfullina, Eric Malmi, and Juho Kannala. 2018. [Learning representations for soft skill matching](#). In *Analysis of Images, Social Networks and Texts*, pages 141–152, Cham. Springer International Publishing.
- Jingbo Shang, Jialu Liu, Meng Jiang, Xiang Ren, Clare R. Voss, and Jiawei Han. 2018. [Automated phrase mining from massive text corpora](#). *IEEE Transactions on Knowledge and Data Engineering*, 30(10):1825–1837.
- Baoxu Shi, Jaewon Yang, Feng Guo, and Qi He. 2020. [Salience and market-aware skill extraction for job targeting](#). In *KDD '20: The 26th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, Virtual Event, CA, USA, August 23-27, 2020*, pages 2871–2879. ACM.
- Fábio Souza, Rodrigo Nogueira, and Roberto Lotufo. 2020. [Portuguese named entity recognition using bert-crf](#).
- Damian A. Tamburri, Willem-Jan Van Den Heuvel, and Martin Garriga. 2020. [Dataops for societal intelligence: a data pipeline for labor market skills extraction and matching](#). In *2020 IEEE 21st International Conference on Information Reuse and Integration for Data Science (IRI)*, pages 391–394.
- Andrei Ternikov. 2022. [Soft and hard skills identification: insights from it job advertisements in the cis region](#).
- Ninande Vermeer, Vera Provatorova, David Graus, Thilina Rajapakse, and Sepideh Mesbah. 2022. [Using robert and extreme multi-label classification to extract implicit and explicit skills from dutch job descriptions](#). In *compjobs '22: Computational Jobs Marketplace, Feb 25, 2022*. ACM.
- Kexin Wang, Nils Reimers, and Iryna Gurevych. 2021. [TSDAE: Using transformer-based sequential denoising auto-encoder for unsupervised sentence embedding learning](#). In *Findings of the Association for Computational Linguistics: EMNLP 2021*, pages 671–688, Punta Cana, Dominican Republic. Association for Computational Linguistics.
- Liang Wang, Nan Yang, Xiaolong Huang, Binxing Jiao, Linjun Yang, Daxin Jiang, Rangan Majumder, and Furu Wei. 2022. [Text embeddings by weakly-supervised contrastive pre-training](#).
- Simon Wild, Soyhan Parlar, Thomas Hanne, and Rolf Dornberger. 2021. [Naïve bayes and named entity recognition for requirements mining in job postings](#). In *2021 3rd International Conference on Natural Language Processing (ICNLP)*, pages 155–161.
- Kaichun Yao, Jingshuai Zhang, Chuan Qin, Peng Wang, Hengshu Zhu, and Hui Xiong. 2022. [Knowledge enhanced person-job fit for talent recruitment](#). In *2022 IEEE 38th International Conference on Data Engineering (ICDE)*, pages 3467–3480.
- Mike Zhang, Kristian Jensen, Sif Sonniks, and Barbara Plank. 2022a. [SkillSpan: Hard and soft skill extraction from English job postings](#). In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 4962–4984, Seattle, United States. Association for Computational Linguistics.
- Mike Zhang, Kristian Nørgaard Jensen, and Barbara Plank. 2022b. [Kompetencer: Fine-grained skill classification in Danish job postings via distant supervision and transfer learning](#). In *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, pages 436–447, Marseille, France. European Language Resources Association.
- Mike Zhang, Kristian Nørgaard Jensen, Rob van der Goot, and Barbara Plank. 2022c. [Skill extraction from job postings using weak supervision](#). In *RecSys in HR'22: The 2nd Workshop on Recommender Systems for Human Resources, in conjunction with the 16th ACM Conference on Recommender Systems*,

September 18–23, 2022, Seattle, USA. CEUR Workshop Proceedings.

Mike Zhang, Rob van der Goot, and Barbara Plank. 2023. [ESCOXLM-R: Multilingual taxonomy-driven pre-training for the job market domain](#). In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 11871–11890, Toronto, Canada. Association for Computational Linguistics.

A Appendix

A.1 Terminology Example

In Table 3, we present an example sentence for better terminology understanding.

	Familiar with building tests in python					
I :	O	O	B	I	O	B
E_C :	O	O	B_{skill}	I_{skill}	O	$B_{knowl.}$
C_D/C_E :	“Python (computer programming)”, “ plan ” “software testing”					

Table 3: An example with annotations for the different tasks described in Section 3. For skill classification (C), we used the ESCO taxonomy in this example, and for skill extraction with coarse labels (E_C) we follow the guidelines of SkillSpan (Zhang et al., 2022a)

A.2 Number of Skill and Knowledge Spans

In Table 4, we show the number of labeled spans for skills and knowledge in the SKILLSPAN (Zhang et al., 2022a), DECORTE (Decorte et al., 2022), and KOMPETENCER (Zhang et al., 2022b) dataset.

A.3 Scores of Selected Models

In Table 5, we display the scores of recent LMM-based approaches on the DECORTE (Decorte et al., 2022) dataset for comparison. Furthermore, we show results of Zhang et al. (2023); Goyal et al. (2023) and (Bhola et al., 2020) on the BHOLA (Bhola et al., 2020) dataset.

Source	# Skill Spans	# Knowledge Spans
SKILLSPAN - HOUSE	2,146	1,418
DECORTE - HOUSE	509*	210*
SKILLSPAN - TECH	2,241	3,828
DECORTE - TECH	419	480*
KOMPETENCER	665	255

Table 4: Number of labeled spans. The star * indicates, that two values found in the Decorte HOUSE test dataset (tagged as knowledge) were actually from the Skillspan TECH dataset; eight values found in the Decorte TECH test dataset (four skill spans, four knowledge spans) were actually from the Skillspan HOUSE dataset.

Model	Source	HOUSE*			TECH*			BHOLA		
		MRR	RP@5	RP@10	MRR	RP@5	RP@10	MRR	R@5	R@10
<i>Classifier^{neg}</i>	(Decorte et al., 2022)	0.299	30.82	38.69	0.326	31.71	39.09	N/A	N/A	N/A
<i>GPTsentences^{aug}</i>	(Decorte et al., 2023)	0.428	45.74	N/A	0.529	54.62	N/A	N/A	N/A	N/A
<i>GPT3.5Re - ranking</i>	(Clavié and Soulié, 2023)	0.427	43.57	51.44	0.488	52.50	59.75	N/A	N/A	N/A
<i>GPT4Re - ranking</i>	(Clavié and Soulié, 2023)	0.495	53.34	61.02	0.537	61.50	68.94	N/A	N/A	N/A
<i>BERT[~]XMLC + CAB</i>	(Bhola et al., 2020)	N/A	N/A	N/A	N/A	N/A	N/A	0.9049	21.67	40.49
<i>JobXMLC</i>	(Goyal et al., 2023)	N/A	N/A	N/A	N/A	N/A	N/A	0.90	18.29	32.33
<i>ESCOXML - R</i>	(Zhang et al., 2023)	N/A	N/A	N/A	N/A	N/A	N/A	0.907	N/A	N/A

Table 5: Scores of selected models on the benchmarking datasets DECORTE and BHOLA.