IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received July 13, 2021, accepted July 28, 2021, date of publication August 19, 2021, date of current version August 31, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3106120

A Survey on Skill Identification From Online Job Ads

IMANE KHAOUJA [1,2], ISMAIL KASSOU [1], AND MOUNIR GHOGHO [2], (Fellow, IEEE)
[1]ENSIAS, Mohammed V University, Rabat 10500, Morocco
[2]TicLab, College of Engineering and Architecture, International University of Rabat, Sale 11103, Morocco

Corresponding author: Imane Khaouja (imane.khaouja@uir.ac.ma)

This work was supported by the Google Africa Ph.D. Fellowship.

ABSTRACT A changing job market, influenced by different factors such as globalization and demographic growth, urges close monitoring. The digitization of the job market has given the opportunity to researchers to better understand job market needs as job postings/ads become more accessible. However, such postings are submitted in unstructured text and need further processing to identify the required skills. The aim of this survey is to review current research on skill identification from job ads and to discuss possible future research directions. In this study, we systematically reviewed 108 research articles on the topic. In particular, we evaluated and classified the prior work aiming to identify the skill bases used for analyzing job market needs; the type of extracted skills; the skill identification methods; the studied sector and the skill identification granularity. Then, we categorized the existent applications and goals of skill identification. Finally, we present key challenges and discuss recent trends.

INDEX TERMS Information extraction, job market needs, online job ads, skill identification, unstructured text.

I. INTRODUCTION

The introduction of digital transformation and the growth of the internet has deeply impacted our lives, from changing our daily basis interactions to how we look for our future jobs. The digital era made available tremendous amounts of structured and unstructured data, including videos, images, and text, which has paved the way for big data analytics. When such data is effectively and efficiently captured, processed and analyzed, entities can gain a clear and complete understanding of their business to improve efficiency and lower costs. In the same vision, analyzing online job market data can lead to further improvements in narrowing the gap between higher education and job market needs [1].

Today there are many websites dedicated to recruitment and job posting as a consequence of the digitization of the job market. Recruiters can directly post job openings on websites called job boards or job portals that can be accessed easily by candidates. The digital transformation is not only changing jobs, from the destruction to the creation of new jobs, but it is also allowing the opportunity to acknowledge better job market needs. This is done by analyzing the huge job ads posted on a daily basis on the internet in an attempt to catch job market dynamics.

Online job ads may not be representative, given that not all job openings are published online. Some recruiters prefer keeping their job opening in a closed circuit for confidentiality concerns or to reduce the time spent filtering and interviewing candidates. Others choose to contact directly specific colleges and universities to share their job openings with students and alumni. From this perspective, even if we collect all online reported job vacancies, there is a share of vacancies that remains inaccessible online, and will therefore fall outside the population sample. Consequently, it remains hard to measure the representativeness of collected job ads even if all online reported job ads are collected. This is due to the fact that the population of job vacancies and its structure is practically unknown [2]. However, job ads remain a good source to understand skill requirements, but not necessarily to estimate the number of vacancies on the market as it seems reasonable to suggest that the core aspects such as skills of a given occupation are likely to remain constant across types of establishments and companies. Using job ads from an established portal and interpreting the results with caution to avoid potential biases can be a valid and acceptable choice. Fortunately, online and internet-based job searching is likely

The associate editor coordinating the review of this manuscript and approving it for publication was Biju Issac.

to become an increasingly more prominent tool for job matching, which promises to improve the representativeness of job ads [2], [3]. As in recent years, the amount of labour market information conveyed through specialized internet portals and services have grown exponentially, encouraging and supporting the realization of many internet services and tools related to the labour market, such as job matching services, advertising of job positions, services for sharing curricula, and the creation of professional networks. Therefore, online job data shows great promise in providing real-time insights into current and ongoing workforce needs in terms of skills.

Providing such information can help different entities to make well-informed decisions and data-driven reforms. The analysis of job ads has been regarded as an established approach to illustrate job market needs, and analyze the skill mismatch/alignment, and predict possible future trends. Furthermore, the findings of job ads have provided input for curriculum development and other applications. While analyzing online job ads is a relatively new trend, it builds upon existing empirical literature based on an analysis of traditional, printed job advertisements [4].

Aside from recruitment companies (CareerBuilder, Textkernel) and professional social networks such as LinkedIn, many projects and initiatives were launched to better understand job market needs by monitoring trends in skills and evolving demands of the job market. By accessing data in real-time, policymakers can evaluate training processes and propose improvements in existing policies. Such projects and initiatives encompassed skill identification from online job ads as a way to draw and infer job market needs. Such projects and initiatives can be summarized as follows:

- CEDEFOP I and II[1]: the European Centre for the Development of Vocational Training (Cedefop) launched a call for tenders to build a system for analysis of online vacancies and to develop a system or tool to analyze vacancies and emerging skills requirements across all EU Member States, realizing a fully-fledged multilanguage (32 languages) system that collects vacancies, extracts skills, and performs real-time monitoring across all 28 EU Member States to support decision-making.
- ESSnet Big Data project[2]: In 2016, the EU and Eurostat launched the ESSnet Big Data project, involving 22 EU Member States with the aim of 'integrating Big Data in the regular production of official statistics, through pilots exploring the potential of selected Big Data sources and building concrete applications'. The aim was to analyze whether online job vacancies can be used to improve labour market statistics of national statistical institutions. The analyzed data was obtained via Web Scraping or co-operations with employment agencies, job portals or other aggregators. Several countries participated in the project and the main finding was that online job vacancies can indeed be used but instead of scraping them themselves it is better to get them from third parties.

Job ads are not only considered as a practical information source for researchers, educators, policymakers, and higher education institutions to explore the dynamic of the job market but they can also be considered as an indicator for job market health. Job ads are the primary means through which companies recruit new applicants for available positions. Therefore, a comprehensive analysis of job market needs requires an information extraction from job ads to leverage the information within these ads as they are submitted in text format.

More specifically, to identify job market needs, Natural Language Processing (NLP) techniques, including information extraction, are used to build personalized bases of skills from textual media knowledge bases, and Machine Learning (ML) based techniques, including deep neural networks and word embeddings, are utilized to capture, as precisely as possible, the skills within the textual source.

A. EXISTING SURVEYS AND CONTRIBUTION

Over the last few years, several research papers and surveys have been published covering different aspects of job market analysis. Applegate [5] examined where reasonably representative job advertisements for academic libraries may be found. This study inspected the data sources used for academic libraries job ads and how representative they are. Kim and Angnakoon [6] identified the methodological approaches and procedures employed in library and information science research that used job ads as data. They evaluated the data collection methods and analysis techniques employed. Papou *et al.* [7] provided a mapping study on knowledge extraction from online sources for the software engineering job market. They inspected the digital sources, the extracted information, the methods, stakeholders for software engineering job market analysis. However, these contributions did not review the different methods used in skill identification, did not discuss the issue of skill identification granularity, and did not cover the applications of skill identification studies done in different sectors.

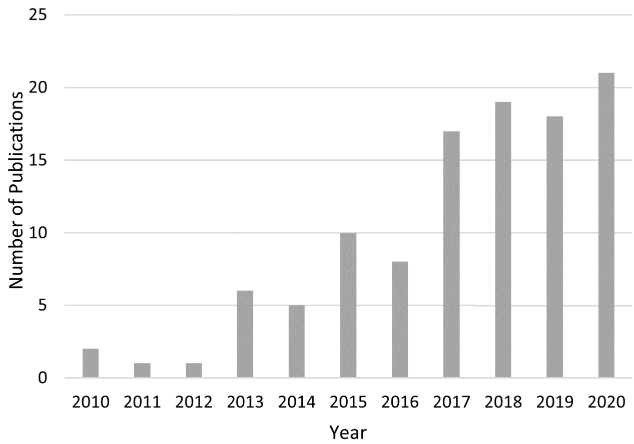
We can conclude that existing surveys have inspected the data sources and methods applied in identifying job market needs in specific fields and sectors. Thus, they only covered a part of skill identification methods in the job market. To the best of our knowledge, there is no survey that fully addresses the major aspects of skill identification including different sectors. In this regard, our paper presents a systematic review that aims to fill this gap through an in-depth analysis to cover recent advancements in skill identification of job market needs from online job ads.

B. MAIN CONTRIBUTIONS

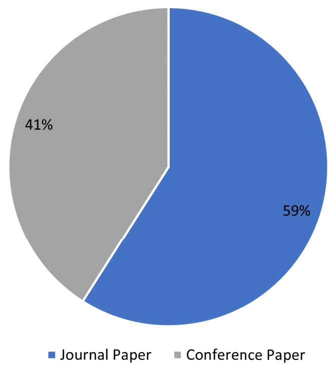
In this paper, we present a comprehensive review of related work that builds upon existing skill identification studies.

[1] https://www.cedefop.europa.eu/it/about-cedefop/public-procurement/real-time-labour-market-information-skill-requirements-setting-eu accessed on 18.01.2021

[2] https://ec.europa.eu/eurostat/cros/essnet-big-data-2_en accessed on 18.01.2021



(a) Publication trends



(b) Publication type

FIGURE 1. **Publication overview.**

More specifically, we cover recent publications in the last ten years. We focus on studies and approaches proposed to comprehend and analyze the required skills in online job ads. By examining more than 108 studies, we evaluate and classify the different approaches to identify the skill bases used for analyzing job market needs; the type of extracted skills; the skill identification methods; the skill identification granularity; the studied sectors of identified skills; the application and goals of skill identification. The main contributions of this paper can be summarized as follows:

- We present an in-depth comparative survey of research papers related to skill identification from job ads.
- We provide a comparison of the different skill bases used for skill identification and we explain methodologies applied for skill base generation.
- We pinpoint the different methodologies applied for skill identification, the granularity of the identified skills and the studied sectors
- We identify the current applications of skill identification from job ads and highlight recent trends.
- We draw insights and discuss future research direction for skill identification from job ads.

The remainder of this article is organized as follows: Section 2 systematically reviews the main available work from the perspectives of skill bases, skill identification methods, skill granularity and studied sectors. Section 3 summarizes the different applications of skill identification and presents contributions of the representative studies and recent trends. Section 4 presents speculations on the future research directions. Section 5 draws conclusions.

II. RESEARCH METHODOLOGY AND RELATED WORK

Studies on skill identification from job ads are gaining increasing interest from researchers. Fig. 1 shows the number of relevant publications in the past ten years. In the last decade, a robust body of research analyzing online job market requirements has been conducted. Nonetheless, an important number of researchers have been analyzing printed job ads in newspapers to identify job market requirements and trends. The first studies conducted to understand online job market needs and determine the required skills by inspecting job ads descriptions used manual content analysis as they dealt with a limited number of job ads. More precisely, manual content analysis was used to discover the scope of skills for a particular position or various occupations in a field.

Then, with the shift of the job advertising medium from newspapers to online job portals, some researchers have resorted to NLP techniques to automate skill identification from job ads. For example, the authors of [8], [9] applied topic modeling to a set of job ads specific to a field or occupation to extract and identify the required skills.

However, the question of the generalization of skill identification was still standing. Therefore, lately, different researchers have presented different methodologies to identify skills from job ads. Some of these researchers proposed the generation of skill bases that encompass the different terminologies of skills to better locate the skills in job ads. Furthermore, human resources experts proposed complete skill bases to better structure skills and competencies such as ESCO, ONET and e_CF. Other researchers proposed ML-based methods to identify skills and sentences containing such skills in job ads.

Once the generalization of skill identification was elucidated, researchers took different directions to better understand job market needs. On one hand, some researchers extended such generalization. In particular, different analyses were drawn for a better understanding of job market needs. Researchers also used the identified skills for job and MOOC recommendation. On the other hand, other researchers tried to weigh skills found in job ads according to their importance, and others tried to measure skill mismatch/alignment.

To identify prior studies identifying skills from job ads, systematic searches were first conducted on several

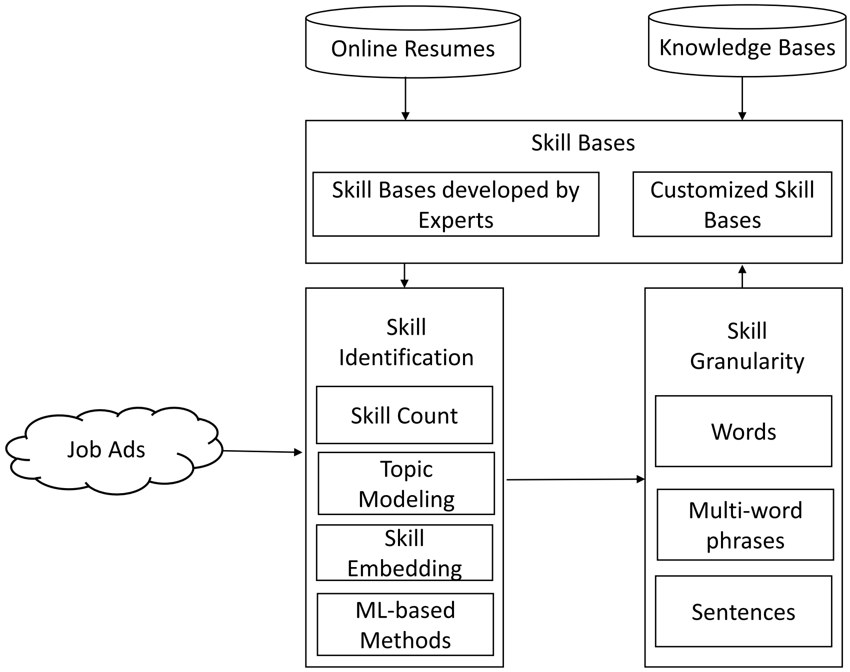


FIGURE 2. The Framework of skill identification from online job ads.

databases. The databases included are Springerlink, IEEE Xplore, ScienceDirect, ACL and Google Scholar. The searches were carried out using the keywords "job ad" with their derivative forms, 'job market' and its derivatives and 'skill' and its derivatives. Research papers were reviewed to select studies that analyze the job market through online job ads. Additionally, the references of the articles were added to ensure that all relevant studies on job ads are included in the study.

Therefore, studies that address the problem of skill identification from job ads were included. However, studies that exclusively identify skills from other sources (only resumes or social networks) were excluded. More precisely, the criteria that served as a means of judging the relevance of a study are described next.

The inclusion criteria are:

- Studies must be written in English.
- Studies must be accessible online.
- Studies must be peer-reviewed and published in journals, conferences, workshops and poster sessions of conferences.
- Studies must be related to skill extraction from job ads and the purpose of their analysis must be relevant to understand job market needs.
- Studies must be published from January 1st 2010 to January 1st 2021.

The exclusion criteria are:

- Studies that extract skills solely from resumes, social networks or candidates profiles.
- Studies that are not written in English.
- Studies that do not provide access to the online full-text.
- Books and gray literature (not published in journals, conferences, etc.).
- Studies that present extended abstracts or summaries of conferences/editorials.
- Studies published before January 1st 2010 or after January 1st 2021.

In this work, we investigated 108 research articles on skill identification from online job ads to shed light on the ongoing research and its possible future directions. First, we present the different data sources used in the skill identification context. Then, we present the existing skill bases developed by human resources experts and researchers. Finally, we categorize the representative work according to the skill identification methods, the skill identification granularity, the studied sectors. Figure 2 shows the components of skill identification from the job ads framework. In the next section, we present the applications and goals of skill identification from job ads.

A. ONLINE DATA SOURCES

The research on skill identification from the online job market mainly depends on analyzing online job ads to identify the required skills. Skill identification is also performed on resumes found in social networks or job portals, curricula and certifications. Next, we define the terms used throughout this paper:

- Job ad: a job posting, also known as a job ad, is an announcement that informs people that a certain job position is available. This announcement is written, generally in an engaging tone, and describes the job position. It has a title and a description. The description

provides details about the position, including skill requirements. This announcement is posted on recruitment websites and job portals to attract the right candidates [6], [10].
- Curriculum: the curriculum is an "academic plan" that usually includes the purpose of the curriculum (i.e., goals for student learning), the content of the curriculum, the order of the learning experience, the instructional methods and resources, the evaluation approaches, and how adjustments to the plan will be made [11], [12].
- Candidate profile: a candidate profile is a textual document that describes the different skills and knowledge acquired by the candidate. They range from candidate profiles found on professional social networks to resumes submitted by candidates in job portals.

B. SKILL BASES

A knowledge base is a collection of records in a database, which typically refers to some kind of "knowledge" about the world. Different knowledge bases were constructed such as YAGO and DBpedia. YAGO is a knowledge base obtained from Wikipedia's infoboxes and from linking Wikipedia's rich category information to the WordNet taxonomy using basic NLP techniques [13]. DBpedia [14] is a community effort to extract structured information from Wikipedia.

In the same context, different domain-specific knowledge bases were developed for specific domains. Such bases contain different terminologies and entities related to a certain domain such as skill bases. Domain-specific knowledge bases can be constructed from the extracted information or by experts. More specifically, the huge amounts of texts published online help in understanding word associations and synonyms which improves information extraction from text and knowledge base construction. In this section, we review the previous studies on skill identification from job ads in terms of the used skill bases such as predefined skill bases and customized skill bases.

In prior work, different skill bases were manually and automatically developed to serve skill identification and standardization of different occupations and sectors. In the literature, we find different terms for skill bases such as taxonomy or ontology. A taxonomy formalizes the hierarchical relationships among concepts and specifies the term to be used to refer to each; it prescribes structure and terminology. An ontology identifies and distinguishes concepts and their relationships; it describes content and relationships. Moreover, an ontology allows multidimensional relationships to be defined. Thus, ontologies and taxonomies are structured skill bases.

These skill bases contain different information about the skills and competencies of different occupations. According to [15], competence or tacit knowledge or soft skill is the "ability to use knowledge, skills and personal, social and/or methodological abilities, in work or study situations and professional and personal development". Skill or competency or implicit hard skill or explicit hard skill is the "ability to apply knowledge and use know-how to complete tasks and solve problems". Also in [16], they consider competence as something "that can demonstrate the application of a generic skill on some knowledge". Based on the two definitions, we can say that competencies and skills are used interchangeably wherein one context competencies can be soft skills or hard skills in another context.

TABLE 1. Comparison of existing skill bases developed by experts [19].

	ESCO	e-CF	O*NET	ISCO-08	ROME
Skill definition	•	•	•	–	–
Skill type	•	–	•	–	–
Occupation	•	–	•	•	•
Domain independent	•	–	•	•	•
Multilingual	•	–	–	–	–

1) SKILL BASES DEVELOPED BY EXPERTS

To structure competencies and skills, skill bases were developed by experts as a reference to human resources services. Table 1 gives a comparison between those skill bases in terms of skills definition, types, occupations, domain independence and languages. Those skill bases are as follows:
- ESCO[3] [17]: this is a project that categorized skills, occupations and other relevant competencies in different languages. The ESCO skill base encompasses 13485 skills and competencies. There is, however, no distinction between skills and competencies. Each of these concepts comes with one preferred term and several non-preferred terms in each of the 27 ESCO languages. Every concept also includes an explanation in the form of a description and a type.[4] ESCO has been developed in an open IT format, is available for use and can be accessed via the ESCO service platform.
- O*Net base [18]: this is used by the U.S. public recruitment services. The core concepts are the job categories, that are associated with generic skills as well as "tools and technology". The skills nomenclature is very coarse with few general skills.
- ROME: this is a repository that contains the definition of different occupations, developed by Pole Emploi.[5] The occupations descriptions are written in French and contains different fields such as the different titles for the occupation, the required skills and qualifications.
- ISCO-08: this is a tool developed by the International Labour Organizations for organizing jobs into a clearly defined set of groups according to the tasks and duties undertaken in the job. Its main aim is to provide a basis for the international reporting, comparison and exchange of statistical and administrative data about occupations.[6]

[3]https://ec.europa.eu/esco accessed on 18.01.2021
[4]https://ec.europa.eu/esco/portal/skill accessed on 18.01.2021
[5]https://www.pole-emploi.fr/employeur/vos-recrutements/le-rome-et-les-fiches-metiers.html accessed on 18.01.2021
[6]https://www.ilo.org/public/english/bureau/stat/isco/index.htm accessed on 18.01.2021

TABLE 2. Developed skill bases in skill identification literature.

Paper	Skill source	Method for Skill Base Generation	Targeted Skills	Sector
Malherbe et al. [27]	Candidate profiles	DBpedia aliases and stackoverflow tags	Soft and Hard skills	All sectors
Zhao et al. [28], Javed et al. [26]	Candidate profiles	Wikipedia API	Soft and Hard skills	All sectors
Bastian et al. [25]	LinkedIn profiles	Entity extraction using a threshold	Soft and Hard skills	All sectors
Khobreh et al. [32]	Job ads and Surveys	Word association	Soft skills	All sectors
Khaouja et al. [29]	Job ads	Word2vec trained on job ads combined with DBpedia wikilinks	Soft skills	All sectors
Balachander et al. [31]	Software Engineering	DBpedia ontology graph	IT skills	IT
Gugnani et al. [30]	Job ads and Candidate profiles	External sources such as Wikipedia combined with Word2vec	Soft skills	All sectors

- e-CF: this is the European e-Competence Framework based on competencies. Its purpose is to provide general and comprehensive e-Competences specified at five proficiency levels that can then be adapted and customized into different contexts from the ICT business and stakeholder application perspectives.[7]

2) CUSTOMIZED SKILL BASES

- Manually Built Skill bases:
  Different studies in the literature analyzed online job market needs by manually building their skill base or list of skills/keywords. Some studies used a predefined list of skills to identify the skills from a set of job ads. These predefined lists of skills can be harvested from prior work or the web; for instance, the work in [20] relies on the list of programming languages, found online. Other studies developed their list of skills from scratch by identifying keywords through a manual selection from the most recurring words in the collected data [21]–[23]. We can observe from Table 6 that such manual skill base construction is performed to study sector-specific job ads. Likewise, Boselli *et al.* [24] enriched the ESCO skill base by capturing the relevant n-gram expressions. An n-gram is a sequence of n words; for instance, a bi-gram is a two-word sequence. Such n-gram expressions were mapped to ESCO skill labels by leveraging strings similarities and validated by domain experts.
- Generated Skill bases from Knowledge Bases and word embeddings:
  Knowledge bases for skills are called skill bases as they contain a list of skills and/or competencies required by candidates to fulfill the requirements of a specific job position. In an attempt to understand those requirements, various skill bases were developed [25]–[28] as depicted in Table 2.
  Most of the developed skill bases mainly relied on external knowledge bases such as DBpedia and Wikipedia to enrich the skill base. Javed *et al.* [26] and Zhao *et al.* [28] examined the skill sections of resumes and Wikipedia categories to define and develop a taxonomy of professional skills. Malherbe and Aufaure [27] built a skill base that relied on page redirection in DBpedia, where skills were first extracted from candidate profiles found in various professional social media, and page redirections in Wikipedia were added to these skills as aliases or alternative labels. In [29], they built a soft skill taxonomy by combining learned word embedding from a dataset of job ads and DBpedia hyperlinks. Similarly, in [30], they used word embedding combined with an in-house named entity recognizer and POS tagging to detect relevant skills that are stored in a skill base for skill extraction task. Also in [25], they used a data-driven approach to build a comprehensive skill list. They leveraged data found in LinkedIn profiles to harvest skills. Then, using Wikipedia pages they deduplicated repeated skills, where skills having the same relationship to the same Wikipedia page are considered related to each other.
  In [31], the authors started crawling from the node ''Software Engineering'' in the DBpedia ontology graph to build an IT skills ontology/base using a depth limited Breadth-First Search algorithm. In [32], the authors relied on word association to build an ontology that bridges between job and knowledge elements found in job ads. Instead of generating a base from scratch, in [33], the authors have leveraged open knowledge bases where they used different techniques such as random walk and different algorithms to extract and infer skills from Wikipedia graph.

[7]https://www.ecompetences.eu/e-cf-overview/ accessed on 18.01.2021

C. SKILL IDENTIFICATION METHODS

In the skill identification literature, we can distinguish between four methods. The first method is based on a skill count between the skill base and the job ad. The second method is based on topic modeling to extract skills cited in job ads. The third method relies on skill embedding where skills are validated by leveraging word embedding. The fourth method is based on Machine Learning (ML) techniques such as text classifiers and deep neural networks to classify job ad sentences or tag skills. The previous studies on skill identification are summarized in Table 3.

TABLE 3. Literature comparison in terms of skill identification methods.

Category	Paper	Year	Focus			Identification Method		
			Type of skills	Sector	Experiment	Features	Model	Supervised/ Unsupervised
Skill Count	Sodhi et al. [21]	2010	Soft and Hard skills	Engineering	Yes	Job ads	Statistical Model	–
	Matturo et al. [34]	2013	Soft skills	IT	Yes	Job ads	Statistical Model	–
	Malherbe et al. [27]	2016	Soft and Hard skills	All sectors	Yes	Job ads Skill base	Statistical Model	–
	Hussain et al. [35]	2017	Hard skills	IT	Yes	Job ads Interviews	Statistical Model	–
	Sibarani et al. [36]	2017	Hard skills	IT	Yes	Job ads Skill base	Statistical Model	–
	Papoutsoglou et al. [15]	2017	Soft and Hard skills	IT	Yes	Job ads Skill base	Statistical Model	–
	Gardiner et al. [22]	2018	Hard skills	IT	Yes	Job ads	Statistical Model	–
	Florea et al. [37]	2018	Hard skills	IT	Yes	Job ads	Statistical Model	–
	Cerioli et al. [38]	2020	Soft and Hard skills	IT	Yes	Job ads Skill base	Statistical Model	–
	Calanca et al. [39]	2019	Soft skills	All sectors	Yes	Job ads Skill base	Statistical Model	–
Topic Modeling	Debortoli et al. [8]	2014	Hard skills	IT	Yes	Job ads	LSA	Unsupervised
	Junwu et al. [40]	2017	Soft and Hard skills	IT	Yes	Job ads	STM	Unsupervised
	Demauro et al. [9]	2018	Hard skills	IT	Yes	Job ads	LDA	Unsupervised
	Gurcan et al. [41], [42]	2019	Soft and Hard skills	IT	Yes	Job ads	LDA	Unsupervised
Skill Embedding	Javed et al. [26]	2017	Soft and Hard skills	All sectors	No	Job ads Skill base	Word2vec	Unsupervised
	Li et al. [43]	2020	Hard skills	All sectors	No	Job ads Skill base	Fastext	Unsupervised
	Gugnani et al. [30]	2020	Explicit and implicit Hard skills	All sectors	No	Job ads Skill base	Word2vec Doc2vec	Unsupervised
ML-based Methods	Jia et al. [44]	2018	Hard skills	IT	Yes	Job ads Labeled entities	LSTM	Supervised
	Sayfullina et al. [45]	2018	Soft skills, Soft skills sentences	All sectors	Yes	Job ads Labeled sentences	CNN+LSTM	Supervised
	Tamburi et al. [46]	2020	Sentences with Hard skills and soft skills	All sectors	No	Job ads Labeled sentences	BERT	Supervised
	Bhola et al. [47]	2020	Hard skills	All sectors	No	Job ads Extracted skills	BERT–XMLC + CAB	Supervised

Figure 3 presents the publication trends of skill identification methods during the past ten years. The skill count method has received continuous attention, most likely because of its simplicity. Recently, an increasing attention has been paid to ML-based methods.

1) SKILL COUNT

Skill count without a prior skill base remains the most reliable method for skill identification. However, it is time-consuming as job ads are read by annotators. Conversely, skill count with a skill base is simple to implement and give insights about skill requirement quickly. The advantage of skill count is to translate the requirements of employers expressed in job ads. The drawbacks of such method are its inability to capture the context of the words and may identify false-positive skills.

Skill count remains the most popular method in skill identification. Such count can be done using a skill base or rely on expert judgment and annotators to label and tag skills manually in a given job ad.

The presence of a skill term can be calculated as a boolean value, where the weight of a skill is 1 if the skill term exists in the job ad, 0 otherwise; or by the classical TF/IDF schema in which the word weight, indicating its importance to the topic, is computed in terms of a trade-off between its frequency in the job ad and in the entire corpus [10]. Once the skills are counted, statistical analysis is performed to highlight the frequency of skills within the studied dataset of job ads.

- **Skill Count using Skill Bases**
  Several studies have examined the job skills required by the job market using an exact match of skills in job ads to compute the skill frequency. In order to identify skills in job ads, prior work [15], [27], [36] relied on an exact match with the used skill base or the defined list of skills. The exact match can be done through the exact name of the skill, the alternative labels or aliases of the

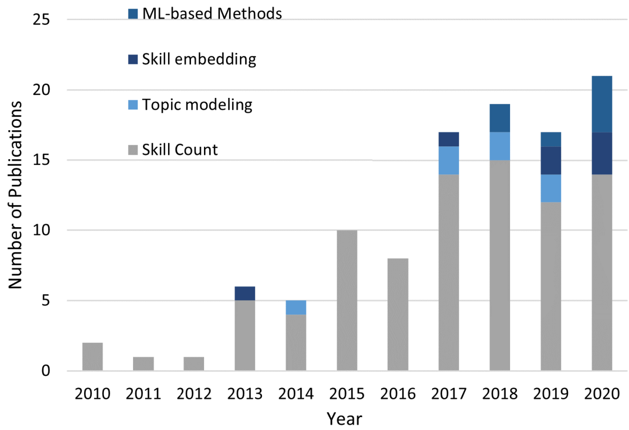


FIGURE 3. **Publication trends by skill identification methods.**

skill. Also, an exact match of skills in job ads can be performed using a simple keyword search or combined with regular expressions.

- **Skill Count without Skill Bases**
  Skill count without skill bases relies mainly on expert judgment. Such count is performed either by manually analyzing the content of the written job ads or by curing and inspecting the list of the most frequent terms in the corpora of job ads. Manual content analysis is a data analysis technique used in qualitative research. Content analysis determines the presence of certain themes or concepts within a given qualitative data to quantify and analyze the presence, meanings and relationships of such themes or concepts.
  Traditionally, content analysis is performed manually by reading the content of each document in the study. More recently, computerized content analysis has been proposed to sift and organize the content.
  Analyzing the list of the most frequent words is performed by experts. They start by identifying the most frequent terms using computerized solutions. Then they clean and identify the skills within this list.
  In the context of analyzing the job market needs, several studies used content analysis to identify the required skills such as [21]–[23], [34], [35], [37], [48]–[50]. Generally, content analysis is used to compute the frequency of skills in specific sectors or domains as depicted in Table 6. Manual analysis can be superior to computerized analysis in terms of the accuracy of the results because words have different meanings in different contexts. However, it is much more time-consuming than computerized analysis and may not be effective for periodic analysis of large amounts of text information [21].

2) TOPIC MODELING

Topic modeling is an unsupervised method that learns the set of underlying topics, in terms of word distributions, from a set of documents. Formally, topics contain the important keywords cited in the text. Topic modeling has seen a growing use for various applications, including direct applications to the unsupervised analysis of textual corpora [51]. However, such methods require a final interpretation of the results by domain experts. Thus, the main studies that applied such methods focused on one area, sector or occupation as the interpretation of the final result is time-consuming. The advantage of such methods resides in their unsupervised nature, where no skill base is required.

Different studies identified the required skills by using topic modeling algorithms. They consider the topics of each job ad as the top required skills. Then, they analyze the keywords of the most cited topics and capture the cited skills. Various topic modeling methods were applied. In [8], the authors used Latent Semantic Analysis (LSA), which is comprised of three phases, to identify the most sought skills in a job ad. The first phase of LSA consists of transforming a collection of documents into a term document matrix. In the second phase, the term-document matrix undergoes a Singular Value Decomposition (SVD) to reduce the dimensionality of the term-document matrix without losing essential information by identifying groups of highly correlated terms (i.e., terms that co-occur together in documents) and highly correlated documents (i.e., documents that contain similar terms). The result of the SVD is a set of factors (topics) with associated high-loading terms and documents. Together, they form patterns of words that represent topics in the underlying collection of documents. The extracted word patterns are interpreted in the third phase, which usually involves additional statistical analyses and, most importantly, expert judgment [52]. LSA depends heavily on SVD which is computationally intensive and hard to update as new documents appear [53].

Similarly, in [9], [41], [42], the authors used Latent Dirichlet Allocation (LDA) to identify the most popular keywords referring to skills in job ads. LDA is an unsupervised generative probabilistic method for modeling a corpus. LDA assumes that each document can be represented as a probabilistic distribution over latent topics, and that topic distribution in all documents share a common Dirichlet prior. Each latent topic in the LDA model is also represented as a probabilistic distribution over words and the word distributions of topics share a common Dirichlet prior as well [51].

3) SKILL EMBEDDING

Many studies leveraged skill embeddings to disambiguate skills and create clusters of skills that appear together (see Table 3). The skill embedding methods were proposed to respond to the limitations of skill count method. Moreover, skill embedding offers the possibility to tag skills from different sectors. The term Skill Embedding refers to word embedding models trained on a collection of job ads. The training aims to produce vector representations of skills such that similar skills and co-occurring skills are close to each other in the vector space. The distance between vectors is measured by metrics such as the Cosine or Jaccard similarity coefficients. Such methods reduce the false positives when performing skill identification. However, this method may increase the number of false negatives. By filtering the identified skills using skill embedding, we may exclude some true extracted skills if they do not pertain to the cluster of skills.

Word embeddings are a class of techniques where individual words are represented as real-valued vectors in a vector space. Each word is mapped to one vector and the vector values are learned from the text.

Once the word embedding model is trained, different similarities between skills can be computed to validate the presence of the skill in the job ad as skills pertaining to the same cluster of skills tend to be close to other skills. Different word embeddings were computed from job ads. For instance, in [26], [28], the authors trained the word2vec model on a collection of job ads to obtain a vector representation of the context of skills. These vectors are then used as input to a clustering algorithm such that in each cluster the aggregated contexts (represented by vectors) can be used to determine a skill sense. For example, such vectors help in differentiating between the term 'BI' that refers to business intelligence and the same term that is also used to refer to the Bank of Indonesia.

Similarly, [54] used word2vec embeddings [55] to compute the similarity between skills cited in job ads and professional standards. The latter are a set of practices, ethics, and behaviors to which members of a particular professional group must adhere. Word2vec was trained on job ads corpus to learn from the context of a job ad and to cluster skills according to their occurrence in job ads.

In [43], the authors used skill embedding (Fastext trained on job ads) to guarantee the coherence of extracted skills, i.e. the vectors of the extracted skills should be close to each other in the skill embeddings vector space. FastText [56], is an extension of word2vec for scalable word representation and classification. One of its major contributions is considering sub-word information by representing each word as the sum of its n-gram vectors. This approach handles out of vocabulary cases since it can generate a representation of a vector close to the original word despite misspellings. One other advantage over word2vec is that rarely occurring words in the corpus are better represented in the word embedding space.

In [30], after identifying explicit skills, the authors infer implicit skills for a job ad using Doc2vec where similar job descriptions that share common features such as location or company share similar skills. Implicit skills are skills that are not explicitly cited but are required by the position. Doc2Vec [57] computes a feature vector for every document in the corpus while Word2Vec computes a feature vector for every word in the corpus. Doc2vec model is an extension of word2Vec to learn document-level embeddings. Based on this similarity, the authors add inferred skills to the skills extracted from the job ad.

4) ML-BASED METHODS

Recently, a series of breakthrough advances in artificial neural networks and ML-based methods has resulted in considerable success in several areas in NLP. There are two promising techniques in the context of skill identification from job ads. The first is Named Entity Recognition (NER) using deep learning. NER is the computerized procedure of recognizing and labeling entities in given texts. NER is also known as (named) entity identification, entity chunking, and entity extraction. It is a subtask of information extraction that seeks to locate and classify named entities mentioned in unstructured text into pre-defined categories such as person names, organizations, locations, medical codes, time expressions, quantities, monetary values, percentages, etc. Considering that the text is a sequential data format, Long-short Term Memory (LSTM) deep learning methods are commonly used to tackle NER problem.

In the skill identification context, typical entities are the different skills cited within job ads. These deep learning models allow detecting skills by parsing the job ad text. For example, in [44], the authors used an LSTM model, pre-trained for skill NER [58], to recognize skills from a job ad text.

The second technique is text classification, which is the process of assigning tags or categories to text according to its content. It is one of the fundamental tasks in NLP with broad applications such as sentiment analysis and spam detection. It was applied by [45], [46] to classify sentences that contain skills in the job ad description. The authors labeled a huge dataset containing sentences of job ads. Once the sentence is identified as containing a skill, the skill cited within is extracted. Sifting such sentences can help in accurately identifying skills within job ads and prevent false extracted skills. In [45], the authors compared Convolutional Neural Network (CNN) and LSTM models for sentence classification. CNN model allows taking into account the word order by applying a fixed-size window on the input array composed of words and their corresponding word embedding dimensions [59]. LSTM leverages the sequential nature of the text, handles long-term dependencies and allows making predictions on a variable-length input.

Similarly, in [46], the authors leveraged state-of-the-art language modeling approaches such as BERT to classify sentences found in the job ads. BERT is designed to pre-train deep bidirectional representations from the unlabeled text by jointly conditioning on both left and right context in all layers. As a result, the pre-trained BERT model can be fine-tuned with just one additional output layer to create models for a wide range of tasks, such as question answering and text classification, without substantial task-specific architecture modifications [60].

Moreover, instead of using NER, the authors of [47] used multi-label text classification to label each job description with the skills within. Instead of classifying the type of each word appearing in the job description, the authors considered job descriptions as the evidence for the binary relevance of thousands of individual skills. To do so, they leveraged BERT encoder and added another layer to perform multi-label classification. They also added a Correlation Aware Bootstrapping process(CAB) that encompasses structured semantic representation of skills and their co-occurrences to acknowledge the missing (implicit) skills cited within the job ads by increasing the number of training examples.

Deep Learning-based methods are very powerful at capturing the hidden relationships between words and showed promising results in different NLP tasks. However such methods remain data greedy and require a significant labeled dataset with rigorous fine-tuning to give good results in the skill identification context.

D. EVALUATION METRICS

The performance of methods in identifying skills within job ads is evaluated using various metrics according to the method.

For skill count and skill embedding methods, studies evaluated the extracted skills from job ads by computing true and false positives as well as true and false negatives from the extracted skills. Recall, precision and F1-score are then computed on the basis of the evaluation of experts. Precision is the fraction of correctly extracted skills out of all the predictions of a particular class. Recall is the fraction of correctly extracted skills out of all actual members of the class. F-Measure is the harmonic mean of precision and recall. To compute such metrics, studies use either domain experts or sampling-based users survey to evaluate the performance of skill identification methods. More specifically, these experts and users are invited to verify the accuracy of the extracted skills within the job ad and add the missing skills.

For topic modeling methods, no evaluation is needed as the method is unsupervised and learns from the words within the job descriptions. Once the results are generated, experts inspect and interpret the generated topics. For ML-based methods, most studies focus primarily on binary classification task. Thus, studies rely on precision, recall and F-score to evaluate their models. For the multi-label classification task, the authors of [47] used Mean Reciprocal Rank (MRR), Normalized Discounted Cumulative Gain(nDCG) and recall to evaluate the performance of skill extraction. MRR indicates the position of first true positive in the predicted list of skills. It yields a score between 0 and 1 depending on the position of true extracted skills. To rank the relevance of the returned skills, they also used nDCG score that discounts the true positive skills that occur later in the extraction rankings.

Many studies depend on the implication of experts for evaluating skill identification since publicly available labelled datasets for skill identification from job ads do not exist. More recently, in [47], the authors released a labeled dataset of job descriptions with the required skills. Such dataset could facilitate the evaluation and comparison of skill identification methods.

E. SKILL IDENTIFICATION GRANULARITY

In skill identification literature, we can distinguish between three approaches of the identified skill identification granularity. The first approach identifies skills as single words or multi-word phrases. The second approach identifies skills expressed in sentences. The third approach combines the two approaches where skills are identified as sentences and n-grams. In this section, we review the previous studies on skill identification of the job market in terms of skill identification granularity and summarize them in Table 4.

Identifying skills as n-grams is popularly applied to analyze skill requirements from job ads. Due to the complexity of dealing with sentences, we notice a small fraction of studies that deal with such skills.

1) SKILLS AS N-GRAMS

Identifying skills as n-grams deals with job descriptions as a bag-of-words model. Skills are identified as single words or multi-word phrases, from bi-grams up to 4 grams to identify composed skills. First studies inspecting skills from job ads identified skills as n-grams [21], [61]. Through the different skill identification methods as depicted in Table 4, researchers captured skills as n-grams from job ads. In skill count, researchers performed naive skill matches of single words or multi-word skill terms. In a similar way, the emerging topics from topic modeling were composed of words or multi-word that express skills. Then, researchers resolved to skill embedding [26], [43] in order to reduce the number of false extracted skills expressed as uni-grams or multi-grams. Such identification works well on identifying programming languages, natural languages, concept words and even soft skills.

2) SKILLS AS SENTENCES

In natural language, the full meaning of a sentence is determined by the words and the syntax. The disadvantage of tagging skills as single words or multi-word phrases is the loss of the structural relations between words in the sentence, which can limit the skill identification. For example, sentences describing the company may cause false positives extracted skills. Therefore, some researchers have taken a further step by detecting sentences containing the skill term as a first step, where deep learning models were used to label job description sentences. For instance, [45], [46] identifies the sentences containing the soft skills and hard skills consecutively, then skills are identified as n-grams from a predefined skill base or list of skills.

Another limitation of skill identification using n-grams resides in its inability to distinguish between sentences like ''Follow the schedule'' and ''Schedule tasks'' where matching with words will fail and identify 'scheduling' as a skill in both tasks. In [62], the authors use pattern mining to identify emerging skills expressed as sentences with action verbs. Based on stemmed sentences, they examine the occurrences of such sentences to highlight the emergent competencies. Moreover, some researchers have resorted to extracting manually skills from job ads to better identify skills expressed in all formats (sentences and n-grams) [37].

TABLE 4. Literature comparison in terms of skill identification granularity.

Category	Reference	Year	Focus			Identification Methods		
			Type of skills	Sector	Experiment	Skill base	Skill Identification	Manual/ Automatic
Sentences	[45]	2018	Soft skills sentences	All sectors	Yes	No	ML-based Method	Automatic
	[46]	2020	Sentences with skills	All sectors	No	No	ML-based Method	Automatic
N-grams	[61]	2010	IT skills	IT	Yes	Yes	Skill count	Automatic
	[21]	2010	Hard skills	Engineering	Yes	Yes	Skill count	Automatic
	[63]	2012	Soft and Hard skills	IT	yes	yes	skill count	Automatic
	[34]	2013	Soft skills	IT	Yes	No	Skill count	Manual
	[64]	2013	Soft and Hard skills	Engineering	Yes	Yes	Skill count	Manual
	[28]	2015	Soft and Hard skills	All sectors	No	Yes	Skill count	Automatic
	[27]	2016	Soft and Hard skills	All sectors	Yes	Yes	Skill count	Automatic
	[26]	2017	Soft and Hard skills	All sectors	No	Yes	Skill embedding	Automatic
	[65]	2017	Soft skills	IT	Yes	No	Skill count	Manual
	[66]	2017	Hard skills	IT	Yes	Yes	Skill count	Manual
	[15]	2017	Soft and Hard skills	IT	Yes	Yes	Skill count	Automatic
	[35]	2017	Soft and Hard skills	IT	Yes	Yes	Skill count	Manual
	[40]	2017	Soft and Hard skills	IT	yes	no	Topic modeling	Automatic
	[36]	2017	Hard skills	IT	Yes	Yes	Skill count	Automatic
	[20]	2018	Hard skills	IT	Yes	Yes	Topic modeling	Automatic
	[67]	2018	Soft and Hard skills	Engineering	Yes	Yes	Skill count	Manual
	[68]	2018	Soft and Hard skills	IT	Yes	Yes	Skill count	Automatic
	[44]	2018	Hard skills	IT	Yes	No	ML-based Method	Semi-automatic
	[23]	2018	Soft and Hard skills	IT	Yes	Yes	Skill count	Semi-automatic
	[22]	2018	Soft and Hard skills	IT	Yes	No	Topic modeling	Semi-automatic
	[48]	2019	Soft skills	Engineering	Yes	Yes	Skill count	Manual
	[39]	2019	Soft skills	All sectors	Yes	Yes	Skill count	Semi-automatic
	[41]	2019	Soft and Hard skills	IT	Yes	No	Topic modeling	Automatic
	[43]	2020	Hard skills	All sectors	No	Yes	Skill embedding	Automatic
Sentences & N-grams	[62]	2015	Hard skills	All sectors	Yes	No	Skill count	Automatic
	[37]	2018	Hard skills	Sector specific	Yes	No	Skill count	Manual
	[69]	2019	Soft and Hard skills	Sector specific	Yes	No	Skill count	Manual

F. STUDIED SECTORS

Existing studies identify the required skills from job ads at different levels and sectors. The skill identification start by analyzing skills from a single sector or an occupation. Such a choice was made in order to keep the number of job ads manageable. Recently, the growing number of complex data collection technologies has increased the demand for large-scale and real-time skill identification in the job market. Therefore, researchers have addressed the issue of skill identification from all online job ads without distinction. In this section, we review the previous studies on skill identification of the job market in terms of the studied sectors i.e., sector-specific and multiple sectors and summarize them in Table 5.

1) SKILL IDENTIFICATION FROM SPECIFIC SECTORS

Different sectors were analyzed in the literature. In Table 5, we can see the different studied sectors such as IT, healthcare, renewable energy, operation research and engineering. However, the majority of the studies targeted IT to better acknowledge the soft and hard skill requirements of different IT occupations through the analysis of online job ads. One explanation of such interest is the large volume and availability of IT job ads, which represent the majority of posted job ads [3].

2) SKILL IDENTIFICATION FROM MULTIPLE SECTORS

Recent studies proposed to identify skills from multiple sectors for a comprehensive skill identification and to shed light on skill requirements from multiple sectors without exception. *Such studies deal with a more important job ad collection where the identification is automatically performed to deal with the huge number of job ads.* To do so, researchers relied on skill bases to generalize the skill identification from multiple sectors. In [26]–[28], the authors generated a comprehensive skill base relying mainly on Wikipedia knowledge base. Besides, customized skill bases were manually tailored for the same purpose. In [24], [70], the authors enriched the ESCO skill base by capturing the relevant n-gram expressions. Then, they tagged skills from job ads of different sectors.

TABLE 5. Studied sectors in skill identification literature.

Sector	Subsector/Occupation	Reference
IT	–	[15], [16], [20], [31], [54], [61], [66], [70], [72]–[77]
	AI	[44]
	Cybersecurity	[23], [78], [79]
	Requirement Engineers	[65], [80]
	Software Development	[34], [41], [68]
	Big Data and BI	[8], [9], [22], [40], [42], [81], [82]
	Software Testing	[37], [38]
	DevOps	[35], [83]
	Developing computers and IS	[49], [63], [69]
	Data science	[19], [36], [84]–[87]
	Mobile Applications	[88]
	Bioinformatics	[89]
Healthcare	–	[90], [91]
	General surgery	[92]
	Nursery	[32], [93]
	Epidemiology	[94]
	Radiologist	[95], [96]
	Pharmacist	[97]
Engineering	–	[48], [64]
	GIS	[98]
	Electrical	[99]
	Operational Research	[21]
	Renewable Energy	[67], [100]
	Biomedical	[101]
	BIM	[102]
	Automotive	[103]
	Industry 4.0	[104], [105]
Logistics	Warehouse managers	[106]
	Logistic managers	[107]
Information Science	Librarians	[108]–[112]
	Records management	[113]
Tourism	–	[114]
Offshoring & Outsourcing	–	[115], [116]
Multiple Sectors	–	[24], [26]–[30], [33], [39], [43], [45]–[47], [62], [70], [117]–[121]
	Entry-level public relations	[122]
	HR	[123]
	Project Managers	[124]
	Sales professionals	[125]
	Social media managers	[126]

Moreover, Skill embedding were leveraged by different studies [30], [43], [71] to disambiguate skills of the different sectors as depicted in Table 3. Furthermore, the authors of [47] used a multi-label classifier to tag skills from job ads without any distinction of the sector related to the job ad.

III. APPLICATION AND EXPERIMENT OF SKILL IDENTIFICATION

Skill identification through online job ads is performed for different applications and goals. In particular, understanding the job market needs is a direct application, where the required skills are extracted for different occupations, sectors and industries. Other applications have been derived such as skill mismatch, Job & MOOC recommendation, skill demand prediction and skill salience measurement. In this section, we review the previous studies on skill identification of the job market in terms of its applications and goals, i.e., skill extraction and job market analysis, curricula design, job & MOOC recommendation, skill mismatch & alignment, skill demand prediction and skill salience measurement; these studies are summarized in Table 6.

A. EXISTING APPLICATIONS OF SKILL IDENTIFICATION

1) SKILL EXTRACTION AND JOB MARKET ANALYSIS

Job market analysis is incontestably a valuable input and a resourceful asset for many entities, e.g. Human Resources departments to better acknowledge their current and future

TABLE 6. Literature comparison in terms of skill identification applications.

Category	Data source	Reference	Type of skills	Sector
Skill Extraction & Job Market Analysis	Job ads	[20], [21], [36], [37], [44]	Hard skills	Sector specific
		[15], [22], [23], [40], [41], [63], [64], [67]	Soft and Hard skills	Sector specific
		[70], [120]	Soft and Hard skills	All sectors
		[34], [48], [65]	Soft skills	Sector specific
	Job ads and Resumes	[27]	Soft and Hard skills	All sectors
		[45]	Soft skills sentences	All sectors
		[28]	Soft and Hard skills	All sectors
		[26]	soft and Hard skills	All sectors
	Job ads and Interviews	[35]	Soft and Hard skills	Sector specific
	LinkedIn resumes	[33]	Soft and Hard skills	All sectors
Curricula Development	Job ads	[69]	soft and hard skills	sector specific
		[68]	Soft and hard skills	Sector specific
	Job ads and University programs	[89]	Hard skills	Sector specific
Job & MOOC Recommendation	Job ads	[43]	Hard skills	All sectors
	Job ads and Resumes	[30]	Hard skills	All sectors
	Job ads and MOOC videos	[121]	Hard skills	Sector specific
Skill Mismatch & Alignment	Job ads	[117]	Hard skills	All sectors
	Job ads and Curricula	[61]	IT skills	Sector specific
	Job ads, Graduate profiles and Certification schemes	[66]	Hard skills	Sector specific
	Job ads, Curricula and Certification schemes	[76]	Hard skills	Sector specific
Competitive Intelligence & Talent Search	Job ads	[8], [9]	Hard skills	Sector specific
		[40], [42], [127]	Soft and Hard skills	Sector specific
	Job ads and Candidate profiles	[128]	Hard skills	Sector specific
Skill Demand Prediction	Job ads	[62]	Hard skills	All sectors
		[129]	Hard skills	Sector specific

needs, and educators, professionals and job seekers to meet the industry needs.

In an attempt to understand job market needs, several studies extracted skills from job ads and other data sources. The requirements of employers and recruiters are extracted and identified through the posted job ads. Some studies focused on understanding hard skills requirements, while others examined soft skill needs. Such studies varied in scope; [22], [23], [65] inspected specific sectors, while [25]–[28] proposed skill extraction systems that deal with job ads from multiple sectors in order to draw a comprehensive job market analysis.

Moreover, in [20], the authors studied multi-perspective popularity of skills; they identified popular skills by examining the generated graph of co-occurring skills from different perspectives (salary-wise or company-wise). Furthermore, Giabelli *et al.* [70] generated a graph from online job ads and developed skill bases of different countries to compare and evaluate country-based labor market dynamics for supporting policy and decision-making activities at the European level. The graph encompassed different nodes such as skills, occupations, cities and sectors.

2) CURRICULA DEVELOPMENT

One direct application of the identification of the skills required by the job market is the development or redesign of the curriculum. A curriculum that responds to the job market needs will offer graduates the necessary skills to find their first job. The purpose of [68], [69], [89] was to provide necessary information for curriculum development by identifying the skills that should be encompassed in the program. These studies were mainly conducted for the IT sector to identify the technologies and programming languages sought after in the job market.

3) JOB AND MOOC RECOMMENDATION

The main idea of a job recommendation system is to provide a set of job recommendations in response to a user's current profile. In these systems, the users can typically upload their skills, resumes or their job search criteria; similarly, the employers or their agents can upload a job ad or skill sets needed along with other information such as location, position and other job-specific details. More recently, MOOC recommendation systems were proposed. Such systems find and recommend relevant MOOC materials to master the skills required by the job market.

Many studies built recommendation systems for jobs and MOOCs using the skills identified from job ads. In [130], the authors presented a data-driven job search engine. This consists of comprehensive search filters including user skill set-focused attributes and various company attributes. The returned jobs are filtered according to the extracted skills. The authors of [131] proposed Jobsense, a framework that gathers and integrate jobs, skills and careers data from multiple data sources such as job ads. In [30], the authors extracted explicit and implicit skills from job ads and candidate resumes for an accurate job recommendation. For the same purpose, in [43], the authors built a job understanding model to improve the representation of jobs and proposed an improvement for the job posting flow in LinkedIn. In [132], the authors built a job recommender that jointly learns the representation of the jobs and skills in the shared k-dimensional latent space of job transition network, job-skill network, and skill co-occurrence network. In a similar way, the authors of [133] proposed a recommender system that, starting from a set of users' skills, identifies the most suitable jobs as they emerge from a large dataset of online IT job ads, which were processed and represented as a graph of occupations and skills.

Likewise, [121] built a MOOC recommender based on skill requirements found in job ads. This was developed using a machine learning model to predict whether a given video fits a skill extracted and required by the job market. Similarly, the authors of [134] examined the dynamics between online learning platforms and online hiring platforms in the software programming profession. By combining four data sources together, Stack Overflow, Google Trends, Udemy, a platform offering skill-based Massively Open Online Courses (MOOCs), and Stack Overflow Jobs. One important finding of the study is that it takes only a few months between the first Stack Overflow appearance of a new skill and its first appearance on Udemy or Stack Overflow Jobs.

4) SKILL MISMATCH AND ALIGNMENT

Skill mismatch is a prevailing issue around the world. There is a growing concern that skills available among the workforce cannot meet the fast-changing demands of the economy, thus creating a major barrier to growth and development. A fast-changing job market, affected by several factors and trends such as globalization and demographic change, gives an impression of an expanding skill gap and brings greater urgency to policy implementation. There are different types of skill mismatch, covering many forms of job market asymmetry. We can cite skill gaps and skill shortages. Skill gaps are when workers lack the skills necessary to do their jobs effectively. Skill shortages are when employers cannot find enough professionals with the right qualifications and skills [135]. They both express an asymmetry between skill supply and skill demand, which made their comparison a focus of different studies.

A skill shortage tool was proposed in [117]; the tool detects skill shortages by correlating different features in the job ad such as its location, salary level and the job ad period being advertised. In [85], they developed a data-driven solution to detect skill shortages from online job ads data. To do so, they identified skills specific to data science and analysis, to capture the labor trends of such an occupation. Then, they inspected the evolution of such skill set in a collection of job ads through different dimensions such as salary levels, education requirements, required experience and posting frequency. Moreover, the authors of [16] tested the alignment of academic profiles and job advertisements, thus detecting the academic topics with which the job offers are most aligned.

We can also cite [61], which is one of the early studies that examined skill mismatch through skill identification from job ads. The study identified current needs and requirements of IT skills through online job ads. Moreover, the authors inspected universities' online courses and module descriptors of UK academic institutions to shed light on the skill mismatch between higher education and job market needs. In [66], based on IT job ads, IT graduate profiles, IT certification schemes and IT national job classification, the authors compared the skills needed in each data source to measure the mismatch.

5) COMPETITIVE INTELLIGENCE AND TALENT SEARCH

Staying attentive to the job market needs does not only benefit job seekers, it is also considered a valuable intelligence for companies to stay competitive in the market. By acknowledging the trending skills, companies can not only adapt and offer innovative services, it can also help companies to keep track of the most sought-after skills, thus guiding organizations in composing effective portfolios. In particular, sector-specific studies focused on identifying the popular and the most sought-after skills to better understand job skills. In [8], [9], [40], [41], [127], the authors inspected the skills in data-related positions, such as business intelligence and big data roles, in order to build a classifier of required skills in such roles. This classification provides a framework that organizations can use to inventory their existing workforce competencies and establish clear strategies for the acquisition and the development of the right skills needed to become more data-driven.

Attracting the best candidates and profiles remains a prevailing concern for most companies. Therefore, companies are constantly innovating in recruitment practices, by boosting the employer brand, hiring headhunters, and are very active on social networks. In [128], the authors proposed a skill identification tool that serves talent search. By identifying both skills cited in the job ad and candidate profiles, the tool can find the best candidates available in resume databanks. In [136], the authors proposed a person-job fit neural network that measures the fitting between job requirements and candidate resume.

6) SKILL DEMAND PREDICTION

Anticipating future trends is beneficial to make decisions following the evolution of the environment. In the job market context, anticipating future needs will allow job seekers to adapt rapidly. A fast-changing job market, affected by several factors and trends, has fast-changing requirements that need close monitoring. Therefore, the identification of any new trend regarding required skills in the job market will enable job seekers, education and training organizations better respond to such needs. In [62], the authors inspected such needs by identifying emerging competencies through pattern mining of skills expressed with action verbs in job ads. In [129], the authors used time series to discover high-demanded skills. Using co-occurring skills observed in job ads, they generate a skill graph representing skills as nodes and denoting edges as the co-occurrence appearance. Then they inspected the evolution of the clusters of skills over time.

B. RECENT APPLICATIONS OF SKILL IDENTIFICATION

1) SKILL SALIENCE

Skill salience focuses on identifying the most significant skills to the core job function and occupation. The term ‘Salience’ was introduced in [137], where the authors argue that not all skills are required equally, as some skills may be more important than others to the core job function. They relied on different signals to identify and measure the importance of skills. Such salience is computed through labeled segments and sentences of the job ad. Similarly, the authors of [133] computed a measure of skill importance for each occupation in each country, using the Revealed Comparative Advantage (RCA) measure, inspired by the work in [138]. This enables to focus on skills that are over-expressed in occupations.

A prior study [10] has tried to quantify skill relevance to job titles using the term-frequency-inverse-document frequency (TF-IDF) model while computing the frequency of skills. The authors argue that if a skill is required by many job ads having the same job title, it is likely to be an important skill.

2) GENDER BIAS

Examining gender bias in job ads has emerged as a new application of skill identification. The purpose of this application is to measure gender bias and inequality while writing the job ad. It could also shed light on gender bias in the job market and promote fairness by reducing inequality.

In [39], the authors examined soft skills required by women and men in job ads and found that soft skills are associated with gender segregation across occupations and reinforce wage inequalities between men and women by rewarding typically ‘‘male’’ characteristics and penalizing ‘‘female’’ traits. In [139], the authors examined how job seekers react to requirements in job ads. More precisely, they investigated the impact of job requirements on women’s job attraction and decision to apply. To tackle these inequalities, a tool was proposed by [140] to detect keywords that discourage women from applying and improve gender-neutrality in job ads.

IV. DIRECTIONS FOR FUTURE WORK

In the past decade, many studies have been conducted to identify skills from job ads. The availability of job ads online has unlocked many horizons for researchers and gave them the opportunity to examine job market needs at a large scale. Moreover, with the increasing power of artificial intelligence techniques, especially the ability to digest unstructured text, extracting more valuable and accurate information from Web media and capturing job market signals and needs with greater sensitivity and precision have become possible. We believe that the study of job market needs, and more specifically skill identification from job ads, will continue for a long time to come. In this section, we briefly describe areas or aspects that we believe are in need of future research and advancement.

- Skill Identification using Deep Learning: Some studies have already explored the use of deep learning for skill tagging from job ads, e.g. [44]–[46]. However, to the best of our knowledge, a comprehensive skill recognizer that leverages deep learning to detect skills expressed in their different terminologies is still lacking. Therefore, investigating the use of recent deep learning techniques for skill identification is an interesting direction for researchers interested in uncovering the job market needs.
- Skill Identification through Graph Embedding: Uncovering skill connections through graph embedding seems to be a promising area of research. Some studies have already explored the use of graphs in examining skills in IT job ads [20], [129]. Generalizing such studies to other sectors would be interesting.
- Understanding Skill Demand Evolution: Predicting skill demand can help better acknowledge future job market needs. Such information is valuable to different entities as it sheds light on a constantly changing job market. Some work has already examined skill demand prediction (see Section III-A6). Such prediction could be performed through a temporal analysis of skill demand such as time series analysis and emergence detection. To perform such a task, historical data is required to detect the variability in skill demand and then build predictive models. In particular, combining such data with other sources such as scientific papers can help shed light on the evolution of new skills from their first introduction to the job market to their surging demand.
- Level of expertise: It would be interesting to examine and measure the importance of skills and level of expertise required in job ads. Such a task could be performed by examining the context of the skills where optional skills should be given less attention than the required skills. More specifically, such scaling of skills could be done through classifying skills as either ‘important’,

'required', or 'optional', and expertise as 'high expertise', medium expertise' and 'basic knowledge'.

- Skill Bases Generation: Many studies constructed skill bases for skill extraction from job ads. However, it would be interesting to construct a skill base that encompasses a complete occupation description with the required tasks, skills and education requirements. Such evolving skill bases would be valuable in detecting emergent skills and jobs.
- Open Datasets for the Skill Identification Task: Open datasets of job ads with the required skills mapped to the different skill bases can help improve the skill identification methods. Moreover, such datasets would structure the task of skill identification from job ads by evaluating the performance of different identification methods on the same benchmark datasets.

V. CONCLUSION

In this study, we systematically reviewed 108 research articles on skill identification from job ads published between 2010 and 2020. We presented, in particular, a comprehensive survey on skill identification.

We provided a framework for dividing the challenge of skill identification into three main issues, i.e., skill base generation, skill identification method, and skill identification granularity, to clarify the paths for further improvements.

We detailed the different applications of skill identification to clarify the importance of such information extraction. These include the identification of skill mismatch, which is a prevailing issue around the world, job recommendation, and job demand prediction. Finally, we provided insightful suggestions for future studies.

ACKNOWLEDGMENT

The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied of Google.

REFERENCES

[1] P. Comyn, O. Strietska-Ilina, and International Labour Office. (2019). *Skills Jobs Mismatches Low-Middle-Income Countries*. [Online]. Available: https://www.ilo.org/wcmsp5/groups/public/—ed_emp/documents/publication/wcms_726816.pdf

[2] L. M. Kureková and Z. Žilinčíková, "Are student jobs flexible jobs? Using online data to study employers' preferences in Slovakia," *IZA J. Eur. Labor Stud.*, vol. 5, no. 1, p. 20, Dec. 2016, doi: 10.1186/s40174-016-0070-5.

[3] A. P. Carnevale, T. Jayasundera, and D. Repnikov, "Understanding online job ads data," Georgetown Univ., Washington, DC, USA, Center Educ. Workforce, Tech. Rep., Apr. 2014.

[4] B. Fabo, M. Beblavý, and K. Lenaerts, "The importance of foreign language skills in the labour markets of central and eastern Europe: Assessment based on data from online job portals," *Empirica*, vol. 44, no. 3, pp. 487–508, Aug. 2017, doi: 10.1007/s10663-017-9374-6.

[5] R. Applegate, "Job ads, jobs, and researchers: Searching for valid sources," *Library Inf. Sci. Res.*, vol. 32, no. 2, pp. 163–170, Apr. 2010. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0740818810000071

[6] J. Kim and P. Angnakoon, "Research using job advertisements: A methodological assessment," *Library Inf. Sci. Res.*, vol. 38, no. 4, pp. 327–335, Oct. 2016. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S074081881630322X

[7] 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. 157595–157613, 2019. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/8884193

[8] S. Debortoli, O. Müller, and J. vom Brocke, "Comparing business intelligence and big data skills," *Bus. Inf. Syst. Eng.*, vol. 6, no. 5, pp. 289–300, Oct. 2014. [Online]. Available: https://link.springer.com/article/10.1007/s12599-014-0344-2

[9] A. De Mauro, M. Greco, M. Grimaldi, and P. Ritala, "Human resources for big data professions: A systematic classification of job roles and required skill sets," *Inf. Process. Manage.*, vol. 54, no. 5, pp. 807–817, 2018. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0306457317300018

[10] W. Zhou, Y. Zhu, F. Javed, M. Rahman, J. Balaji, and M. McNair, "Quantifying skill relevance to job titles," in *Proc. IEEE Int. Conf. Big Data (Big Data)*, Dec. 2016, pp. 1532–1541.

[11] L. R. Lattuca and J. S. Stark, *Shaping the College Curriculum: Academic Plans in Context*. Hoboken, NJ, USA: Wiley, 2011.

[12] C. Constantinov, P. S. Popescu, C. M. Poteras, and M. L. Mocanu, "Preliminary results of a curriculum adjuster based on professional network analysis," in *Proc. 19th Int. Conf. Syst. Theory, Control Comput. (ICSTCC)*, Oct. 2015, pp. 860–865.

[13] H. Bast, B. Buchhold, and E. Haussmann, "Semantic search on text and knowledge bases," *Found. Trends Inf. Retr.*, vol. 10, nos. 2–3, pp. 119–271, Jun. 2016, doi: 10.1561/1500000032.

[14] S. Auer, C. Bizer, G. Kobilarov, J. Lehmann, R. Cyganiak, and Z. Ives, "DBpedia: A nucleus for a web of open data," in *The Semantic Web* (Lecture Notes in Computer Science). Berlin, Germany: Springer, 2007, pp. 722–735, doi: 10.1007/978-3-540-76298-0_52.

[15] M. Papoutsoglou, N. Mittas, and L. Angelis, "Mining people analytics from StackOverflow job advertisements," in *Proc. 43rd Euromicro Conf. Softw. Eng. Adv. Appl. (SEAA)*, Aug. 2017, pp. 108–115. [Online]. Available: https://ieeexplore.ieee.org/document/8051336

[16] A. González-Eras and J. Aguilar, "Determination of professional competencies using an alignment algorithm of academic profiles and job advertisements, based on competence thesauri and similarity measures," *Int. J. Artif. Intell. Educ.*, vol. 29, no. 4, pp. 536–567, Dec. 2019, doi: 10.1007/s40593-019-00185-z.

[17] J. De Smedt, M. L. Vrang, and A. Papantoniou, "ESCO: Towards a semantic web for the European labor market," in *Proc. LDOW WWW*, May 2015. [Online]. Available: http://ceur-ws.org/Vol-1409/paper-10.pdf

[18] National Research Council, Division of Behavioral and Social Sciences and Education, Committee on National Statistics, and Panel to Review the Occupational Informatio Network (O*NET), *A Database for a Changing Economy: Review of the Occupational Information Network (O*NET)*. Washington, DC, USA: National Academies, Jun. 2010.

[19] A.-S. Dadzie, E. M. Sibarani, I. Novalija, and S. Scerri, "Structuring visual exploratory analysis of skill demand," *J. Web Semantics*, vol. 49, pp. 51–70, Mar. 2018. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1570826817300690

[20] T. Xu, H. Zhu, C. Zhu, P. Li, and H. Xiong, "Measuring the popularity of job skills in recruitment market: A multi-criteria approach," in *Proc. AAAI Conf. Artif. Intell.*, vol. 32, 2018.

[21] M. S. Sodhi and B.-G. Son, "Content analysis of OR job advertisements to infer required skills," *J. Oper. Res. Soc.*, vol. 61, no. 9, pp. 1315–1327, Sep. 2010, doi: 10.1057/jors.2009.80.

[22] A. Gardiner, C. Aasheim, P. Rutner, and S. Williams, "Skill requirements in big data: A content analysis of job advertisements," *J. Comput. Inf. Syst.*, vol. 58, no. 4, pp. 374–384, Mar. 2018, doi: 10.1080/08874417.2017.1289354.

[23] N. G. Brooks, T. H. Greer, and S. A. Morris, "Information systems security job advertisement analysis: Skills review and implications for information systems curriculum," *J. Educ. Bus.*, vol. 93, no. 5, pp. 213–221, 2018, doi: 10.1080/08832323.2018.1446893.

[24] R. Boselli, M. Cesarini, S. Marrara, F. Mercorio, M. Mezzanzanica, G. Pasi, and M. Viviani, "WoLMIS: A labor market intelligence system for classifying web job vacancies," *J. Intell. Inf. Syst.*, vol. 51, no. 3, pp. 477–502, Dec. 2018, doi: 10.1007/s10844-017-0488-x.

[25] M. Bastian, M. Hayes, W. Vaughan, S. Shah, P. Skomoroch, H. Kim, S. Uryasev, and C. Lloyd, "LinkedIn skills: Large-scale topic extraction and inference," in *Proc. 8th ACM Conf. Recommender Syst. (RecSys)*. New York, NY, USA: ACM, 2014, pp. 1–8, doi: 10.1145/2645710.2645729.

[26] F. Javed, P. Hoang, T. Mahoney, and M. McNair, "Large-scale occupational skills normalization for online recruitment," in *Proc. 29th IAAI Conf.*, 2017.

[27] E. Malherbe and M.-A. Aufaure, "Bridge the terminology gap between recruiters and candidates: A multilingual skills base built from social media and linked data," in *Proc. IEEE/ACM Int. Conf. Adv. Social Netw. Anal. Mining*, Aug. 2016, pp. 583–590.

[28] M. Zhao, F. Javed, F. Jacob, and M. McNair, "SKILL: A system for skill identification and normalization," in *Proc. 29th AAAI Conf. Artif. Intell. (AAAI)*. Palo Alto, CA, USA: AAAI Press, 2015, pp. 4012–4017. [Online]. Available: http://dl.acm.org/citation.cfm?id=2888116.2888273

[29] I. Khaouja, G. Mezzour, K. M. Carley, and I. Kassou, "Building a soft skill taxonomy from job openings," *Social Netw. Anal. Mining*, vol. 9, no. 1, p. 43, Dec. 2019. [Online]. Available: https://link.springer.com/article/10.1007/s13278-019-0583-9

[30] A. Gugnani and H. Misra, "Implicit skills extraction using document embedding and its use in job recommendation," in *Proc. AAAI*, 2020, pp. 13286–13293. [Online]. Available: https://ojs.aaai.org/index.php/AAAI/article/view/7038

[31] Y. Balachander and T.-S. Moh, "Ontology based similarity for information technology skills," in *Proc. IEEE/ACM Int. Conf. Adv. Social Netw. Anal. Mining (ASONAM)*, Aug. 2018, pp. 302–305.

[32] 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 Trans. Emerg. Topics Comput.*, vol. 4, no. 3, pp. 462–473, Jul. 2016.

[33] I. Kivimäki, A. Panchenko, A. Dessy, D. Verdegem, P. Francq, H. Bersini, and M. Saerens, "A graph-based approach to skill extraction from text," in *Proc. TextGraphs-8 Graph-based Methods Natural Lang. Process.*, 2013, pp. 79–87.

[34] G. Matturro, "Soft skills in software engineering: A study of its demand by software companies in Uruguay," in *Proc. 6th Int. Workshop Cooperat. Human Aspects Softw. Eng. (CHASE)*, May 2013, pp. 133–136.

[35] W. Hussain, T. Clear, and S. MacDonell, "Emerging trends for global DevOps: A New Zealand perspective," in *Proc. IEEE 12th Int. Conf. Global Softw. Eng. (ICGSE)*, May 2017, pp. 21–30. [Online]. Available: https://openrepository.aut.ac.nz/bitstream/10292/10659/4/ICGSE_2017_paper_19.pdf

[36] E. M. Sibarani, S. Scerri, C. Morales, S. Auer, and D. Collarana, "Ontology-guided job market demand analysis: A cross-sectional study for the data science field," in *Proc. 13th Int. Conf. Semantic Syst.* New York, NY, USA: ACM, Sep. 2017, pp. 25–32, doi: 10.1145/3132218.3132228.

[37] R. Florea and V. Stray, "Software tester, we want to hire you! An analysis of the demand for soft skills," in *Agile Processes in Software Engineering and Extreme Programming* (Lecture Notes in Business Information Processing), J. Garbajosa, X. Wang, and A. Aguiar, Eds. Cham, Switzerland: Springer, 2018, pp. 54–67. [Online]. Available: https://link.springer.com/chapter/10.1007/978-3-319-91602-6_4

[38] M. Cerioli, M. Leotta, and F. Ricca, "What 5 million job advertisements tell us about testing: A preliminary empirical investigation," in *Proc. 35th Annu. ACM Symp. Appl. Comput.* New York, NY, USA: ACM, Mar. 2020, pp. 1586–1594, doi: 10.1145/3341105.3373961.

[39] F. Calanca, L. Sayfullina, L. Minkus, C. Wagner, and E. Malmi, "Responsible team players wanted: An analysis of soft skill requirements in job advertisements," *EPJ Data Sci.*, vol. 8, no. 1, p. 13, Apr. 2019.

[40] J. Wu, H. Shi, and J. Yang, "Are big data talents different from business intelligence expertise?: Evidence from text mining using job recruitment advertisements," in *Proc. Int. Conf. Service Syst. Service Manage.*, Jun. 2017, pp. 1–6. [Online]. Available: https://ieeexplore.ieee.org/document/7996289/

[41] F. Gurcan and S. Sevik, "Expertise roles and skills required by the software development industry," in *Proc. 1st Int. Inform. Softw. Eng. Conf. (UBMYK)*, Nov. 2019, pp. 1–4. [Online]. Available: https://ieeexplore.ieee.org/document/8965571/

[42] F. Gurcan and N. E. Cagiltay, "Big data software engineering: Analysis of knowledge domains and skill sets using LDA-based topic modeling," *IEEE Access*, vol. 7, pp. 82541–82552, 2019.

[43] S. Li, B. Shi, J. Yang, J. Yan, S. Wang, F. Chen, and Q. He, "Deep job understanding at LinkedIn," in *Proc. 43rd Int. ACM SIGIR Conf. Res. Develop. Inf. Retr.* New York, NY, USA: ACM, Jul. 2020, pp. 2145–2148, doi: 10.1145/3397271.3401403.

[44] S. Jia, X. Liu, P. Zhao, C. Liu, L. Sun, and T. Peng, "Representation of job-skill in artificial intelligence with knowledge graph analysis," in *Proc. IEEE Symp. Product Compliance Eng. Asia (ISPCE-CN)*, Dec. 2018, pp. 1–6. [Online]. Available: https://ieeexplore.ieee.org/document/8805749/

[45] L. Sayfullina, E. Malmi, and J. Kannala, "Learning representations for soft skill matching," in *Analysis of Images, Social Networks and Texts* (Lecture Notes in Computer Science), W. M. P. van der Aalst, V. Batagelj, G. Glava, D. I. Ignatov, M. Khachay, S. O. Kuznetsov, O. Koltsova, I. A. Lomazova, N. Loukachevitch, A. Napoli, A. Panchenko, P. M. Pardalos, M. Pelillo, and A. V. Savchenko, Eds. Cham, Switzerland: Springer, 2018, pp. 141–152, doi: 10.1007/978-3-030-11027-7_15.

[46] D. A. Tamburri, W.-J. V. D. Heuvel, and M. Garriga, "DataOps for societal intelligence: A data pipeline for labor market skills extraction and matching," in *Proc. IEEE 21st Int. Conf. Inf. Reuse Integr. Data Sci. (IRI)*, Aug. 2020, pp. 391–394, doi: 10.11092FIRI49571.2020.00063.

[47] A. Bhola, K. Halder, A. Prasad, and M.-Y. Kan, "Retrieving skills from job descriptions: A language model based extreme multi-label classification framework," in *Proc. 28th Int. Conf. Comput. Linguistics*. Barcelona, Spain, Dec. 2020, pp. 5832–5842. [Online]. Available: https://www.aclweb.org/anthology/2020.coling-main.513

[48] H. Chaibate, A. Hadek, S. Ajana, S. Bakkali, and K. Faraj, "Analyzing the engineering soft skills required by Moroccan job market," in *Proc. 5th Int. Conf. Optim. Appl. (ICOA)*, Apr. 2019, pp. 1–6. [Online]. Available: https://ieeexplore.ieee.org/document/8727712

[49] F. Niederman and M. Sumner, "Resolving the skills paradox: A content analysis of a jobs database," in *Proc. Comput. People Res. Conf.* New York, NY, USA: ACM, Jun. 2019, pp. 164–167, doi: 10.1145/3322385.3322395.

[50] J. A. Rios, G. Ling, R. Pugh, D. Becker, and A. Bacall, "Identifying critical 21st-century skills for workplace success: A content analysis of job advertisements," *Educ. Researcher*, vol. 49, no. 2, pp. 80–89, Mar. 2020, doi: 10.3102/0013189X19890600.

[51] H. Jelodar, Y. Wang, C. Yuan, X. Feng, X. Jiang, Y. Li, and L. Zhao, "Latent Dirichlet allocation (LDA) and topic modeling: Models, applications, a survey," *Multimedia Tools Appl.*, vol. 78, no. 11, pp. 15169–15211, Jun. 2019, doi: 10.1007/s11042-018-6894-4.

[52] T. K. Landauer, P. W. Foltz, and D. Laham, "An introduction to latent semantic analysis," *Discourse Process.*, vol. 25, nos. 2–3, pp. 259–284, 1998.

[53] B. V. Barde and A. M. Bainwad, "An overview of topic modeling methods and tools," in *Proc. Int. Conf. Intell. Comput. Control Syst. (ICICCS)*, Jun. 2017, pp. 745–750.

[54] D. Botov, J. Klenin, A. Melnikov, Y. Dmitrin, I. Nikolaev, and M. Vinel, "Mining labor market requirements using distributional semantic models and deep learning," in *Business Information Systems* (Lecture Notes in Business Information Processing), W. Abramowicz and R. Corchuelo, Eds. Cham, Switzerland: Springer, 2019, pp. 177–190.

[55] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, and J. Dean, "Distributed representations of words and phrases and their compositionality," in *Proc. Adv. Neural Inf. Process. Syst.*, 2013, pp. 3111–3119.

[56] P. Bojanowski, E. Grave, A. Joulin, and T. Mikolov, "Enriching word vectors with subword information," *Trans. Assoc. Comput. Linguistics*, vol. 5, pp. 135–146, Dec. 2017, doi: 10.1162/tacl_a_00051.

[57] Q. Le and T. Mikolov, "Distributed representations of sentences and documents," in *Proc. Int. Conf. Mach. Learn.*, 2014, pp. 1188–1196.

[58] G. Lample, M. Ballesteros, S. Subramanian, K. Kawakami, and C. Dyer, "Neural architectures for named entity recognition," in *Proc. Conf. North Amer. Chapter Assoc. Comput. Linguistics, Hum. Lang. Technol.*, 2016, pp. 260–270.

[59] Y. Kim, "Convolutional neural networks for sentence classification," in *Proc. Conf. Empirical Methods Natural Lang. Process. (EMNLP)*, Oct. 2014, pp. 1746–1751. [Online]. Available: https://www.aclweb.org/anthology/D14-1181

[60] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," 2018, *arXiv:1810.04805*. [Online]. Available: http://arxiv.org/abs/1810.04805

[61] A. Capiluppi and A. Baravalle, "Matching demand and offer in on-line provision: A longitudinal study of monster.com," in *Proc. 12th IEEE Int. Symp. Web Syst. Evol. (WSE)*, Sep. 2010, pp. 13–21. [Online]. Available: http://citeseerx.ist.psu.edu/viewdoc/download;jsessionid=33030E3A13BB901FF51C858ACDB1563A?doi=10.1.1.426.528&rep=rep1&type=pdf

[62] Y. Abbound, A. Boyer, and A. Brun, "Predict the emergence: Application to competencies in job offers," in *Proc. IEEE 27th Int. Conf. Tools Artif. Intell. (ICTAI)*, Nov. 2015, pp. 612–619. [Online]. Available: https://ieeexplore.ieee.org/document/7372190/

[63] C. Litecky, A. J. Igou, and A. Aken, "Skills in the management oriented IS and enterprise system job markets," in *Proc. 50th Annu. Conf. Comput. People Res. (SIGMIS-CPR)*. New York, NY, USA: ACM, 2012, pp. 35–44, doi: 10.1145/2214091.2214104.

[64] F. Steinmann, K.-I. Voigt, and T. Schaeffler, "Engineering competences: A content analysis of job advertisements," in *Proc. Technol. Manage. IT-Driven Services (PICMET)*, Jul. 2013, pp. 1925–1929. [Online]. Available: https://ieeexplore.ieee.org/document/6641759

[65] M. Daneva, C. Wang, and P. Hoener, "What the job market wants from requirements engineers? An empirical analysis of online job ads from The Netherlands," in *Proc. ACM/IEEE Int. Symp. Empirical Softw. Eng. Meas. (ESEM)*, Nov. 2017, pp. 448–453. [Online]. Available: https://ieeexplore.ieee.org/document/8170133

[66] H. B. Santoso and P. O. H. Putra, "Bridging the gap between IT graduate profiles and job requirements: A work in progress," in *Proc. 7th World Eng. Educ. Forum (WEEF)*, Nov. 2017, pp. 145–148. [Online]. Available: https://ieeexplore.ieee.org/document/8467146

[67] B. Baruah, T. Ward, N. Jackson, and A. Gbadebo, "Addressing the skills gap for facilitating renewable energy entrepreneurship—An analysis of the wind energy sector," in *Proc. Majan Int. Conf. (MIC)*, Mar. 2018, pp. 1–7. [Online]. Available: https://ieeexplore.ieee.org/document/8363156

[68] C. Hiranrat and A. Harncharnchai, "Using text mining to discover skills demanded in software development jobs in Thailand," in *Proc. 2nd Int. Conf. Educ. Multimedia Technol.* New York, NY, USA: ACM, Jul. 2018, pp. 112–116, doi: 10.1145/3206129.3239426.

[69] J. C. Nwokeji, R. Stachel, T. Holmes, and R. O. Orji, "Competencies required for developing computer and information systems curriculum," in *Proc. IEEE Frontiers Educ. Conf. (FIE)*, Oct. 2019, pp. 1–9. [Online]. Available: https://ieeexplore.ieee.org/document/9028613

[70] A. Giabelli, L. Malandri, F. Mercorio, and M. Mezzanzanica, "GraphLMI: A data driven system for exploring labor market information through graph databases," *Multimedia Tools Appl.*, pp. 1–30, Jun. 2020, doi: 10.1007/s11042-020-09115-x.

[71] F. Javed, Q. Luo, M. McNair, F. Jacob, M. Zhao, and T. S. Kang, "Carotene: A job title classification system for the online recruitment domain," in *Proc. IEEE 1st Int. Conf. Big Data Comput. Service Appl.*, Mar. 2015, pp. 286–293.

[72] D. Azahar, N. A. Manaf, N. Dharmarajan, and W. A. W. Adnan, "Malaysia IT career opportunity analysis based on classified advertisement in the New Straits Times (2000–2008)," in *Proc. Int. Conf. Inf. Retr. Knowl. Manage. (CAMP)*, Mar. 2010, pp. 348–352.

[73] T. Janotík, "Current labor market trends in the area of information technology in Slovakia," in *Proc. 13th Int. Conf. Emerg. eLearning Technol. Appl. (ICETA)*, Nov. 2015, pp. 1–8.

[74] I. A. Wowczko, "Skills and vacancy analysis with data mining techniques," in *Informatics*, vol. 2. Basel, Switzerland: Multidisciplinary Digital Publishing Institute, 2015, pp. 31–49.

[75] M. Stevens and R. Norman, "Industry expectations of soft skills in IT graduates: A regional survey," in *Proc. Australas. Comput. Sci. Week Multiconference*. New York, NY, USA: ACM, Feb. 2016, pp. 1–9, doi: 10.1145/2843043.2843068.

[76] M. A. Mardis, J. Ma, F. R. Jones, C. R. Ambavarapu, H. M. Kelleher, L. I. Spears, and C. R. McClure, "Assessing alignment between information technology educational opportunities, professional requirements, and industry demands," *Educ. Inf. Technol.*, vol. 23, no. 4, pp. 1547–1584, Jul. 2018, doi: 10.1007/s10639-017-9678-y.

[77] P. Pillai and D. Amin, "Understanding the requirements of the Indian IT industry using web scrapping," *Procedia Comput. Sci.*, vol. 172, pp. 308–313, Jan. 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1877050920313739

[78] L. E. Potter and G. Vickers, "What skills do you need to work in cyber security?: A look at the Australian market," in *Proc. ACM SIGMIS Conf. Comput. People Res.* New York, NY, USA: ACM, Jun. 2015, pp. 67–72, doi: 10.1145/2751957.2751967.

[79] I. Rahhal, I. Makdoun, G. Mezzour, I. Khaouja, K. Carley, and I. Kassou, "Analyzing cybersecurity job market needs in Morocco by mining job ads," in *Proc. IEEE Global Eng. Educ. Conf. (EDUCON)*, Apr. 2019, pp. 535–543.

[80] C. Wang, Y. Tang, P. Liang, M. Daneva, and M. van Sinderen, "What industry wants from requirements engineers in China?: An exploratory and comparative study on RE job ads," in *Proc. 14th ACM/IEEE Int. Symp. Empirical Softw. Eng. Meas. (ESEM)*. New York, NY, USA: ACM, Oct. 2020, pp. 1–11, doi: 10.1145/3382494.3410672.

[81] L. A. Leon, K. C. Seal, Z. H. Przasnyski, and I. Wiedenman, "Skills and competencies required for jobs in business analytics: A content analysis of job advertisements using text mining," *Int. J. Bus. Intell. Res.*, vol. 8, no. 1, pp. 1–25, Jan. 2017, doi: 10.4018/IJBIR.2017010101.

[82] W. Poonnawat, E. Pacharawongsakda, and N. Henchareonlert, "Jobs analysis for business intelligence skills requirements in the ASEAN region: A text mining study," in *Advances in Intelligent Informatics, Smart Technology and Natural Language Processing*, T. Theeramunkong, R. Kongkachandra, M. Ketcham, N. Hnoohom, P. Songmuang, T. Supnithi, and K. Hashimoto, Eds. Cham, Switzerland: Springer, 2019, pp. 187–195.

[83] N. Kerzazi and B. Adams, "Who needs release and devops engineers, and why?" in *Proc. Int. Workshop Continuous Softw. Evol. Del.* New York, NY, USA: ACM, May 2016, pp. 77–83, doi: 10.1145/2896941.2896957.

[84] F. Pires, J. Barbosa, and P. Leitão, "Data scientist under the Da.Re perspective: Analysis of training offers, skills and challenges," in *Proc. IEEE 16th Int. Conf. Ind. Informat. (INDIN)*, Jul. 2018, pp. 523–528.

[85] N. Dawson, M.-A. Rizoiu, B. Johnston, and M.-A. Williams, "Adaptively selecting occupations to detect skill shortages from online job ads," in *Proc. IEEE Int. Conf. Big Data (Big Data)*, Dec. 2019, pp. 1637–1643.

[86] P. N. Y. Le, L. V. Nguyen, T. H. Nguyen, K. M. Vo, and S. N. Hoang, "Toward a smart platform for data science career," in *Proc. 11th Int. Conf. Knowl. Syst. Eng. (KSE)*, Oct. 2019, pp. 1–6.

[87] M. Della Volpe and F. Esposito, "The data scientist job in Italy: What companies require," in *Advances P2P, Parallel, Grid, Cloud and Internet Computing* (Lecture Notes in Networks and Systems), L. Barolli, P. Hellinckx, and J. Natwichai, Eds. Cham, Switzerland: Springer, 2020, pp. 894–903.

[88] J. Jia, Z. Chen, and X. Du, "Understanding soft skills requirements for mobile applications developers," in *Proc. IEEE Int. Conf. Comput. Sci. Eng. (CSE), IEEE Int. Conf. Embedded Ubiquitous Comput. (EUC)*, vol. 1, Jul. 2017, pp. 108–115.

[89] H. Wu and A. Palani, "Bioinformatics curriculum development and skill sets for bioinformaticians," in *Proc. IEEE Frontiers Educ. Conf. (FIE)*, Oct. 2015, pp. 1–7. [Online]. Available: https://ieeexplore.ieee.org/document/7344321/

[90] J. M. Keralis, B. L. Riggin-Pathak, T. Majeski, B. A. Pathak, J. Foggia, K. M. Cullinen, A. Rajagopal, and H. S. West, "Mapping the global health employment market: An analysis of global health jobs," *BMC Public Health*, vol. 18, no. 1, p. 293, Feb. 2018, doi: 10.1186/s12889-018-5195-1.

[91] A. Alibasic, M. C. E. Simsekler, T. Kurfess, W. L. Woon, and M. A. Omar, "Utilizing data science techniques to analyze skill and demand changes in healthcare occupations: Case study on USA and UAE healthcare sector," *Soft Comput.*, vol. 24, no. 7, pp. 4959–4976, Apr. 2020, doi: 10.1007/s00500-019-04247-1.

[92] M. R. Decker, N. W. Bronson, C. C. Greenberg, J. P. Dolan, K. C. Kent, and J. G. Hunter, "The general surgery job market: Analysis of current demand for general surgeons and their specialized skills," *J. Amer. College Surgeons*, vol. 217, no. 6, pp. 1133–1139, Dec. 2013.

[93] V. B. Kobayashi, S. T. Mol, H. A. Berkers, G. Kismihók, and D. N. Den Hartog, "Text mining in organizational research," *Organizational Res. Methods*, vol. 21, no. 3, pp. 733–765, Jul. 2018.

[94] C. L. Frankenfeld, "Trends in employer postings for epidemiology jobs: An analysis of PublicHealthJobs.net data from 2003 to 2016," *Annals Epidemiol.*, vol. 27, no. 9, pp. 553–557, Sep. 2017. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1047279717306646

[95] M. Nellamattathil, J. Flug, J. C. Hoffmann, S. Weiss, and I. Amber, "Characteristics of musculoskeletal radiology job postings to guide radiology trainees," *Clin. Imag.*, vol. 61, pp. 54–57, May 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0899707120300279

[96] D. H. Hoffman and A. B. Rosenkrantz, "Defining the abdominal radiologist based on the current U.S. Job market," *Abdominal Radiol.*, vol. 43, no. 11, pp. 3184–3187, Nov. 2018, doi: 10.1007/s00261-018-1577-1.

[97] J. A. Draime, E. C. Wicker, Z. J. Krauss, J. L. Sweeney, and D. C. Anderson, "Description of position ads for pharmacy technicians," *Pharmacy*, vol. 8, no. 2, p. 88, May 2020. [Online]. Available: https://www.mdpi.com/2226-4787/8/2/88

[98] J. E. Hong, "Identifying skill requirements for GIS positions: A content analysis of job advertisements," *J. Geography*, vol. 115, no. 4, pp. 147–158, Jul. 2016, doi: 10.1080/00221341.2015.1085588.

[99] L. K. H. Yanaze and R. D. D. Lopes, "Transversal competencies of electrical and computing engineers considering market demand," in *Proc. IEEE Frontiers Educ. Conf. (FIE)*, Oct. 2014, pp. 1–4. [Online]. Available: https://ieeexplore.ieee.org/document/7044169

[100] A. Ward, A. Gbadebo, and B. Baruah, "Using job advertisements to inform curricula design for the key global technical challenges," in *Proc. Int. Conf. Inf. Technol. Based Higher Educ. Training (ITHET)*, Jun. 2015, pp. 1–6.

[101] A. E. Ward, B. Baruah, A. Gbadebo, and N. J. Jackson, "What do job adverts tell higher education about the 'shape' of biomedical engineering graduates?" in *Proc. Int. Conf. Advancements Med. Health Care Technology*, S. Vlad and N. M. Roman, Eds., Cluj-Napoca, Romania. Cham, Switzerland: Springer, Oct. 2017, pp. 43–48.

[102] M. Uhm, G. Lee, and B. Jeon, "An analysis of BIM jobs and competencies based on the use of terms in the industry," *Autom. Construct.*, vol. 81, pp. 67–98, Sep. 2017. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S092658051630259X

[103] I. Makdoun, G. Mezzour, K. M. Carley, and I. Kassou, "Analyzing the needs of the automotive job market in Morocco," in *Proc. 13th Int. Conf. Comput. Sci. Educ. (ICCSE)*, Aug. 2018, pp. 1–6.

[104] M. Pejic-Bach, T. Bertoncel, M. Meško, and Ž. Krstić, "Text mining of industry 4.0 job advertisements," *Int. J. Inf. Manage.*, vol. 50, pp. 416–431, Feb. 2020. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0268401218313677

[105] J. Mena, M. Torrent-Moreno, D. González, L. Portell, O. Pujol, and J. Vitrià, "Analysis of vocational education and training and the labour market in catalonia. A data-driven approach," in *Machine Learning and Knowledge Discovery in Databases* (Communications in Computer and Information Science), P. Cellier and K. Driessens, Eds. Cham, Switzerland: Springer, 2020, pp. 522–537.

[106] I. Szabó, K. Ternai, and S. Fodor, "Competence mining to improve training programs," in *Innovative Technologies and Learning* (Lecture Notes in Computer Science), T.-C. Huang, T.-T. Wu, J. Barroso, F. E. Sandnes, P. Martins, and Y.-M. Huang, Eds. Cham, Switzerland: Springer, 2020, pp. 147–157.

[107] H. Kotzab and S. Wünsche, "Competences and qualifications of logistic managers—An empirical analysis of job postings in Germany," in *Logistics Management* (Lecture Notes in Logistics), J. Dethloff, H.-D. Haasis, H. Kopfer, H. Kotzab, and J. Schönberger, Eds. Cham, Switzerland: Springer, 2015, pp. 517–527.

[108] M. L. Gold and M. G. Grotti, "Do job advertisements reflect ACRL's standards for proficiencies for instruction librarians and coordinators?: A content analysis," *J. Academic Librarianship*, vol. 39, no. 6, pp. 558–565, Nov. 2013. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0099133313000773

[109] J. M. Nielsen, "The blended business librarian: Technology skills in academic business librarian job advertisements," *J. Bus. Finance Librarianship*, vol. 18, no. 2, pp. 119–128, Apr. 2013, doi: 10.1080/08963568.2013.768849.

[110] J. Xia and M. Wang, "Competencies and responsibilities of social science data librarians: An analysis of job descriptions," *College Res. Libraries*, vol. 75, no. 3, pp. 362–388, May 2014.

[111] H.-L. Chen and Y. Zhang, "Educating data management professionals: A content analysis of job descriptions," *J. Acad. Librarianship*, vol. 43, no. 1, pp. 18–24, Jan. 2017. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0099133316303263

[112] Y. Zhang, F. Su, and B. Hubschman, "A content analysis of job advertisements for digital humanities-related positions in academic libraries," *J. Academic Librarianship*, vol. 47, no. 1, Jan. 2021, Art. no. 102275. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S009913332030166X

[113] L. Gibbons and J. Douglas, "Markers of professional identity: Records management jobs advertisements in Australia," *Records Manage. J.*, vol. 31, no. 1, pp. 1–17, May 2021, doi: 10.1108/RMJ-03-2020-0007.

[114] I. Szabó and G. Neusch, "Dynamic skill gap analysis using ontology matching," in *Electronic Government and the Information Systems Perspective* (Lecture Notes in Computer Science), A. Ko and E. Francesconi, Eds. Cham, Switzerland: Springer, 2015, pp. 231–242.

[115] I. Khaouja, I. Rahhal, M. Elouali, G. Mezzour, I. Kassou, and K. M. Carley, "Analyzing the needs of the offshore sector in Morocco by mining job ads," in *Proc. IEEE Global Eng. Educ. Conf. (EDUCON)*, Apr. 2018, pp. 1380–1388. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/8363390

[116] P. Alpar, L. Osterbrink, and E. Klein, "Outsourcing and offshoring to individuals," in *Digital Technologies for Global Sourcing of Services* (Lecture Notes in Business Information Processing), I. Oshri, J. Kotlarsky, and L. P. Willcocks, Eds. Cham, Switzerland: Springer, 2020, pp. 93–109.

[117] C. Terblanche and P. Wongthongtham, "Ontology-based employer demand management," *Softw., Pract. Exper.*, vol. 46, no. 4, pp. 469–492, Apr. 2016.

[118] A. Phaphuangwittayakul, S. Saranwong, S.-N. Panyakaew, P. Inkeaw, and J. Chaijaruwanich, "Analysis of skill demand in Thai labor market from online jobs recruitments websites," in *Proc. 15th Int. Joint Conf. Comput. Sci. Softw. Eng. (JCSSE)*, Jul. 2018, pp. 1–5. [Online]. Available: https://ieeexplore.ieee.org/document/8457393

[119] R. B. K. Mbah, M. Rege, and B. Misra, "Using spark and scala for discovering latent trends in job markets," in *Proc. 3rd Int. Conf. Compute Data Anal.* New York, NY, USA: ACM, Mar. 2019, pp. 55–62, doi: 10.1145/3314545.3314566.

[120] E. Colombo, F. Mercorio, and M. Mezzanzanica, "AI meets labor market: Exploring the link between automation and skills," *Inf. Econ. Policy*, vol. 47, pp. 27–37, Jun. 2019. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0167624518301318

[121] M. Tavakoli, S. Hakimov, R. Ewerth, and G. Kismihok, "A recommender system for open educational videos based on skill requirements," in *Proc. IEEE 20th Int. Conf. Adv. Learn. Technol. (ICALT)*, Jul. 2020, pp. 1–5. [Online]. Available: https://ieeexplore.ieee.org/document/9155881

[122] S. Meganck, J. Smith, and J. P. D. Guidry, "The skills required for entry-level public relations: An analysis of skills required in 1,000 PR job ads," *Public Relations Rev.*, vol. 46, no. 5, Dec. 2020, Art. no. 101973. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0363811120301004

[123] P. Poba-Nzaou, S. Uwizeyemungu, and C. Clarke, "Patterns underlying required HR and IT competencies: A content and cluster analysis of advertisements of HR manager positions," *Int. J. Hum. Resource Manage.*, vol. 31, no. 16, pp. 2065–2088, Sep. 2020, doi: 10.1080/09585192.2018.1424019.

[124] A. Atolagbe and S. Floyd, "Diversifying the next generation of project managers: Skills project managers must have in the digital age," in *Sustainable Digital Communities* (Lecture Notes in Computer Science), A. Sundqvist, G. Berget, J. Nolin, and K. I. Skjerdingstad, Eds. Cham, Switzerland: Springer, 2020, pp. 665–676.

[125] K. Robson, S. Beninger, and D. Hall, "Attraction of marketing and sales professionals in the financial services industry: An analysis of pre-, during and post-financial crisis job postings," *J. Financial Services Marketing*, vol. 19, no. 2, pp. 85–93, Jun. 2014, doi: 10.1057/fsm.2014.6.

[126] R. Pharswan, P. V. Ilavarasan, and A. K. Kar, "What makes a social media manager? Insights from the content analysis of job advertisements in India," in *Re-Imagining Diffusion and Adoption of Information Technology and Systems: A Continuing Conversation* (IFIP Advances in Information and Communication Technology), S. K. Sharma, Y. K. Dwivedi, B. Metri, and N. P. Rana, Eds. Cham, Switzerland: Springer, 2020, pp. 391–403.

[127] P. G. Lovaglio, M. Cesarini, F. Mercorio, and M. Mezzanzanica, "Skills in demand for ICT and statistical occupations: Evidence from web-based job vacancies," *Stat. Anal. Data Mining, Data Sci. J.*, vol. 11, no. 2, pp. 78–91, 2018, doi: 10.1002/sam.11372.

[128] O. Manad, M. Bentounsi, and P. Darmon, "Enhancing talent search by integrating and querying big HR data," in *Proc. IEEE Int. Conf. Big Data (Big Data)*. Seattle, WA, USA: IEEE, Dec. 2018, pp. 4095–4100.

[129] E. M. Sibarani and S. Scerri, "SCODIS: Job advert-derived time series for high-demand skillset discovery and prediction," in *Database Expert Systems Applications* (Lecture Notes in Computer Science), S. Hartmann, J. Küng, G. Kotsis, A. M. Tjoa, and I. Khalil, Eds. Cham, Switzerland: Springer, 2020, pp. 366–381.

[130] R. Muthyala, S. Wood, Y. Jin, Y. Qin, H. Gao, and A. Rai, "Data-driven job search engine using skills and company attribute filters," Sep. 2017, *arXiv:1709.05050*. [Online]. Available: http://arxiv.org/abs/1709.05050

[131] X. J. S. Ashok, E.-P. Lim, and P. K. Prasetyo, "JobSense: A data-driven career knowledge exploration framework and system," in *Proc. IEEE Int. Conf. Data Mining Workshops (ICDMW)*, Nov. 2018, pp. 1411–1416.
[132] V. S. Dave, B. Zhang, M. Al Hasan, K. AlJadda, and M. Korayem, "A combined representation learning approach for better job and skill recommendation," in *Proc. 27th ACM Int. Conf. Inf. Knowl. Manage. (CIKM)*. New York, NY, USA: ACM, Oct. 2018, pp. 1997–2005, doi: 10.1145/3269206.3272023.
[133] A. Giabelli, L. Malandri, F. Mercorio, M. Mezzanzanica, and A. Seveso, "Skills2Job: A recommender system that encodes job offer embeddings on graph databases," *Appl. Soft Comput.*, vol. 101, Mar. 2021, Art. no. 107049. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S156849462030987X
[134] R. Yazdanian, R. West, and P. Dillenbourg, "Keeping up with the trends: Analyzing the dynamics of online learning and hiring platforms in the software programming domain," *Int. J. Artif. Intell. Educ.*, pp. 1–44, Dec. 2020, doi: 10.1007/s40593-020-00231-1.
[135] U. Dadush and H. Saoudi. *Inequality in Morocco: An International Perspective*. Accessed: Feb. 11, 2020. [Online]. Available: https://www.policycenter.ma/publications/inequality-morocco-international-perspective#
[136] C. Qin, H. Zhu, T. Xu, C. Zhu, L. Jiang, E. Chen, and H. Xiong, "Enhancing person-job fit for talent recruitment: An ability-aware neural network approach," in *Proc. 41st Int. ACM SIGIR Conf. Res. Development Inf. Retr. (SIGIR)*. New York, NY, USA: ACM, Jun. 2018, pp. 25–34, doi: 10.1145/3209978.3210025.
[137] B. Shi, J. Yang, F. Guo, and Q. He, "Salience and market-aware skill extraction for job targeting," in *Proc. 26th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining (KDD)*. New York, NY, USA: ACM, Aug. 2020, pp. 2871–2879, doi: 10.1145/3394486.3403338.
[138] A. Alabdulkareem, M. R. Frank, L. Sun, B. AlShebli, C. Hidalgo, and I. Rahwan, "Unpacking the polarization of workplace skills," *Sci. Adv.*, vol. 4, no. 7, Jul. 2018, Art. no. eaao6030.
[139] L. Wille and E. Derous, "When job ads turn you down: How requirements in job ads may stop instead of attract highly qualified women," *Sex Roles*, vol. 79, nos. 7–8, pp. 464–475, Oct. 2018.
[140] S. Böhm, O. Linnyk, J. Kohl, T. Weber, I. Teetz, K. Bandurka, and M. Kersting, "Analysing gender bias in IT job postings: A pre-study based on samples from the German job market," in *Proc. Comput. People Res. Conf.*, Jun. 2020, pp. 72–80.

IMANE KHAOUJA received the engineering degree in computer science from ENSIAS, Rabat, Morocco, in 2014, with a major in business intelligence. She is currently pursuing the Ph.D. degree with Mohammed V University, Rabat, with a co-direction with the International University of Rabat. Her research interests include NLP and information extraction. She was a recipient of the Google Ph.D. Fellowship, in 2018.

ISMAIL KASSOU received the master's degree, in 1988, and the Ph.D. degree in computer science from the University of Rouen, in 1992. He has been a Professor with Mohammed V University, Rabat, Morocco, since 1993. He was the Deputy Director in charge of research and cooperation of ENSIAS and the Director of the ST2I Doctoral Studies Center, Mohammed V University. He has a wide experience in project management and international cooperation. He is currently the Vice President of research and cooperation with Mohammed V University. He is the coauthor of two patents and about 63 scientific articles. His research interests include knowledge management, text mining, web mining, and information retrieval.

MOUNIR GHOGHO (Fellow, IEEE) received the M.Sc. and Ph.D. degrees from the National Polytechnic Institute of Toulouse, France, in 1993 and 1997, respectively. He was an EPSRC Research Fellow with the University of Strathclyde, Scotland, from September 1997 to November 2001. In December 2001, he joined the School of Electronic and Electrical Engineering, University of Leeds, England, where he was promoted to a Full Professor, in 2008. While still affiliated with the University of Leeds, in 2010, he joined the International University of Rabat, Morocco, where he is currently the Dean of the Doctoral College and the Director of TICLab (ICT Research Laboratory). He held an invited scientist/professor positions at Telecom Paris-Tech (France), NII (Japan), BUPT (China), University Carlos III of Madrid (Spain), ENSICA (Toulouse), Darmstadt Technical University (Germany), and Minnesota University (USA). He is also a Co-Founder and the Co-Director of the CNRS-Associated International Research Lab DataNet. His research interests include signal processing, machine learning, wireless communication, and their applications. He was a member of the IEEE Signal Processing Society SPCOM Technical Committee, the IEEE Signal Processing Society SPTM Technical Committee, and the IEEE Signal Processing Society SAM Technical Committee. He is currently a member of the Steering Committee of the IEEE TRANSACTIONS OF SIGNAL AND INFORMATION PROCESSING OVER NETWORKS. He was awarded U.K. Royal Academy of Engineering Research Fellowship, in 2000, and the IBM Faculty Award, in 2013. He organized many conferences and workshops, including the 2019 Intelligent Environments Conference, the 2018 ITCities Workshop, the 2013 European Signal Processing Conference (EUSIPCO), and the 2010 IEEE Workshop on Signal Processing for Advanced Wireless Communications (SPAWC). In the past, he served as an Associate Editor for the *Signal Processing Magazine*, the IEEE TRANSACTIONS ON SIGNAL PROCESSING, the IEEE SIGNAL PROCESSING LETTERS, and the *Digital Signal Processing* journal (Elsevier).