A Dynamic Jobs-Skills Knowledge Graph

Alejandro Seif[1,*,†], Sarah Toh[1,2,†] and Hwee Kuan Lee[3,†]

[1] *Artificial Intelligence Practice, GovTech, S117438, Singapore*
[2] *Skills Development Group, SkillsFuture Singapore, S408533, Singapore*
[3] *Bioinformatics Institute, Agency for Science, Technology and Research (A*STAR), S138671, Singapore*

Abstract

Adult learners aiming to upskill themselves are often bombarded with overwhelming and contradicting information in terms of what skills are relevant to their occupations at a given time. Companies and governments also face the same concerns as they navigate ways to optimally manage their workforce and citizens' employability. The goal of this paper is to explore the use of knowledge graphs to better understand the complex relationship between occupations and skills, aiming to provide a cohesive answer to the question of how a skill relates to an occupation. We focus on harnessing taxonomies represented as relationships within a temporal knowledge graph.

The contribution of this paper is a methodology to construct a comprehensive Jobs and Skills knowledge base that evolves dynamically over time, starting with the Singapore SkillsFuture Skills Framework as a foundational resource. Through the integration of expert knowledge and extracted entity information from labour market data, we employ data analytics techniques to refine and update the knowledge base to accurately reflect the evolving needs of the Singaporean economy. Through this, technologists in organizations working with jobs and skills can seek to build knowledge bases that incorporate expert knowledge and labour market signals so as to answer jobs and skills questions.

Keywords

adult learners, data analytics, education, graph database, job posting, knowledge graphs, labour market, occupations, online job ads, property graph, skill mismatch, taxonomies, temporal knowledge graphs

1. Introduction

The cacophony of signals coming from workforce development specialists, human resource consultants and job market advocates, including adult education offerings, can leave individuals seeking to upskill (or re-skill) themselves confused about what skills are critical for a certain occupation and which are not.

Companies express their needs by hiring individuals with specific skills. Hence, incorporating labour market data from job postings is crucial to get a direct reading of the skills demand for given occupations. Additionally, expert knowledge coming from specialists in various sectors of the economy presents contextualised views of the overall role that specific occupations and skills play. Together, the labour market insights and domain knowledge give a holistic picture of how skills and occupations are related at certain point in time.

In this paper, we will focus on unlocking the power of taxonomies represented as relationships between entities in a temporal knowledge graph to address the question of what skills are required by occupations and how those relationships evolve over time.

With the transition from traditional newspapers to online job portals, researchers initially [1] [2] turned towards natural language processing (NLP) and topic modelling techniques to automate the identification of skills from job advertisements and their categorization within specific fields or occupations. However, concerns [3] persisted regarding the generalization of skill identification methods. Some approaches to mitigate these concerns involved the creation of comprehensive skill bases that encompass diverse skill terminologies to enhance skill identification in job ads. Human resources experts have published standardized occupations and skill classifications such as ISCO [4], ESCO [5] and O*NET [6]. Some authors have further tailored published knowledge bases for the particular economic needs of their own country [7].

SkillsFuture Singapore (SSG), the national skills authority of the Singapore public service, together with employers, industry associations, institutes of higher learning and unions, has created its very own occupation and skill knowledge base, the SSG Skills Framework [8]. The SSG Skills Framework provides information on key sectors, occupations, job roles, and the required existing and emerging skills. The Singapore Department of Statistics also maintains an occupational classification known as Singapore Standard Ocupational Classification (SSOC) [9], which is used by the Skills Framework to classify Job Roles (Skills Framework) into their respective Occupations (SSOC).

Relying on rich occupational and skill bases, labour market data analytics and knowledge graph applications have emerged as pivotal tools for understanding workforce dynamics and optimizing skills matching in the job market. [10] introduces the concept of the Occupation Space, a network representation of French workers' job mobility patterns, revealing intense skill-relatedness between occupations. Leveraging O*NET, [11] highlights quantification of skill polarization in the labour market, impacting career mobility and wage distribution. [12] review explores knowledge graph-based frameworks for education and employability, enhancing skill-job matching and skill-course alignment. [3] survey of the state of the art underscores the significance of online job ads in providing large-scale insights into job market needs, fueled by advancements in artificial intelligence. [13] presents a graph-theoretic approach to skills valuation, relying on O*NET data, emphasizing skills' facilitation of occupational transitions as a measure of value. These approaches however, take existing knowledge bases as-is, even though that knowledge represents a snapshot in time when they were created and not a constantly updated source of information.

Meanwhile, [14] discusses the importance of knowledge graph refinement and the challenges in measuring and vali-

RecSys in HR'24: The 4th Workshop on Recommender Systems for Human Resources, in conjunction with the 18th ACM Conference on Recommender Systems, October 14–18, 2024, Bari, Italy.

[*] Corresponding author.

[†] These authors contributed equally.

alejandro_seif@tech.gov.sg (A. Seif); sarah_toh@ssg.gov.sg (S. Toh)
0000-0003-1275-5438 (A. Seif)

CEUR Workshop Proceedings ceur-ws.org ISSN 1613-0073

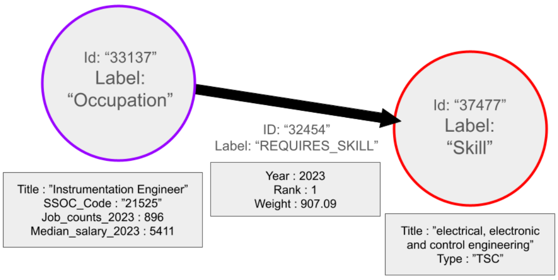


Figure 1: Example of how a property graph uses a relationship to link nodes of different types, given by node labels *Occupation* and *Skill*, and relationship label *REQUIRES_SKILL*. It is possible to include properties in the nodes and edges, such as *Median Salary (2023)* for an Occupation or *title* for a Skill.

dating such updates. In this context, [15] proposes a Skills and Occupation Knowledge Graph that leverages ISCO and ESCO, refined with job posting data, to facilitate skills-based job matching and career pathfinding. As opposed to static knowledge graphs where facts do not change with time, temporal knowledge graphs [16] incorporate new information over time to refine the relationships. It is in this area that we would like to frame our study.

Our contributions are as follows: (i) presenting a comprehensive Jobs and Skills knowledge base, and (ii) creating an approach for that knowledge base to evolve over time. Using the SSG Skills Framework as a starting point, the Jobs and Skills knowledge base combines expert knowledge with extracted entity information from large amounts of unstructured data such as job postings, course listings and other available datasets. Data analytics in the form of machine learning classifiers and sliding window weighted averages are used to refine the information extracted from unstructured data and integrate it with expert knowledge, enabling the Jobs and Skills knowledge base to evolve to reflect the needs of the ever-changing Singapore economy.

2. The Jobs and Skills Knowledge Graph (JSKG)

A knowledge graph is a structured model of information, comprising entities, relationships, and semantic descriptions [17]. The growing availability of diverse data sets has prompted researchers to explore semantic and conceptual representations, leading to the popularity of knowledge graphs (KGs). KGs serve as fundamental components for various information systems that rely on structured knowledge [18].

Within knowledge graphs, there is a sub type called property graphs (example provided in Fig. 1). These graphs represent data using nodes (entities) and relationships (edges) with properties associated with both nodes and relationships. These graphs are flexible and intuitive, making them suitable for representing complex relationships and attributes within data sets. By following the direction of the relationship, it is possible to compose sentences in English that provide the semantic information of the relationship (e.g. **Occupation** (*Instrumentation Engineer*) - REQUIRES SKILL → **Skill** (*electrical, electronic and control engineering*)")

We present the Jobs Skills Knowledge Graph (JSKG), a

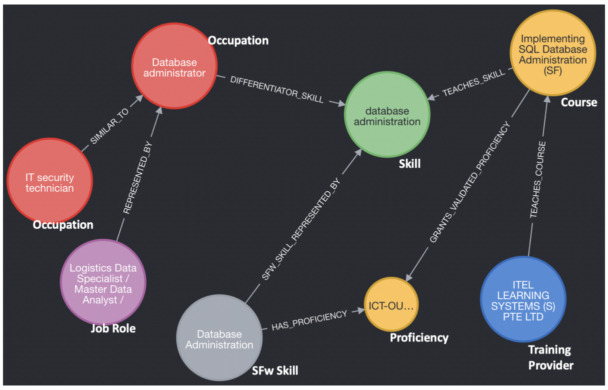


Figure 2: Example of entities and their relationships in the JSKG. Most relationships are not shown in this image to preserve ease of comprehension for the reader. Each occupation has on average 132 relationships. In this manuscript, we will focus on the Occupation-Skill relationships.

knowledge graph that can be queried based on the proximity of entities within it. By capturing the information in the form of a knowledge graph, relationships between entities such as *Occupations* and *Skills* can be easily investigated to determine commonalities and relatedness between entities. Figure 2 provides an example of various types of nodes and their relationships. The following section provides details on how the JSKG was constructed, with a focus on the *Occupation-Skill* relationships.

The JSKG itself is structured as a property graph and stored in a graph database, which can be queried through an API. A website and a large language model-enabled chatbot make use of this API to consume the results of queries between entities (e.g. two occupations) and present the information in a compelling way to non-expert users.

Details on the implementation and software tools utilised can be found in the Appendix A.

3. Data Sources

We leverage a comprehensive dataset encompassing three types of entities to examine the evolving landscape of skills demand and occupational structures in Singapore.

Firstly, the primary data source comprises of job postings collected from the internet spanning the years 2018 to 2023, offering insights into the dynamic nature of skill demand within the job market over this period. We shall refer to this as Labour Market data.

Secondly, complementing the labour market data, we incorporate expert knowledge from SkillsFuture's Singapore Skills Framework (SFw) [8], a knowledge base for job roles and skills, which also includes a mapping between job roles and the Singapore Standard Occupation Classification (SSOC) [9].

Thirdly, we also include data on courses and training providers validated by SkillsFuture Singapore to enrich the JSKG with information on options for increasing skills supply to meet market demand.

The integration of these diverse entities allows for a nuanced exploration of the intersections between job market dynamics, skill evolution, and occupational structures, contributing to a comprehensive understanding of the Singaporean workforce and continuous education landscape.

Table 1
The table showcases how the SSOC classifies occupations taxonomically. Example provided for *Instrumentation Engineer* and some of its adjacent occupations.

SSOC Code	SSOC Title
215	ENGINEERING PROFESSIONALS II
2152	Electronics Engineers
21522	Computer engineer
21523	Semi-conductor engineer
21524	Audio and video equipment engineer
21525	Instrumentation engineer

Table 2
SSOC Major Groups Classification

SSOC Code	Major Group
1	Legislators, Senior Officials and Managers
2	Professionals
3	Associate Professionals and Technicians
4	Clerical Support Workers
5	Service and Sales Workers
6	Agricultural and Fishery Workers
7	Craftsmen and Related Trades Workers
8	Plant and Machine Operators and Assemblers
9	Cleaners, Labourers and Related Workers

Figure 3 displays Occupation and Skills data sources (section 3), the tools utilised to extract entities (section 3.5) from job postings, and the methods applied on the knowledge graph entities to further derive insights.

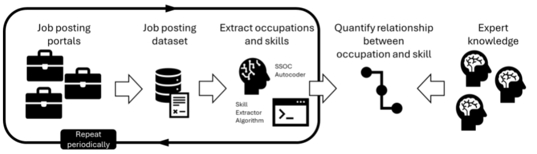

Job posting portals
Job posting dataset
Extract occupations and skills
SSOC Autocoder
Skill Extractor Algorithm
Quantify relationship between occupation and skill
Expert knowledge
Repeat periodically

Figure 3: Periodical data gathering process where job postings are collected and extracted for skills and occupations. Together with expert knowledge from SkillsFuture Singapore Skills Framework (SFw), data are combined to determine the quantified relationship between occupation and skill.

The remainder of this section will provide details on each of the specific data sources leveraged to construct the JSKG, focusing on Occupations and Skills.

3.1. Singapore Standard Occupation Classification

The JSKG uses the Singapore Standard Occupation Classification (SSOC) from the year 2020, which offers a standardized framework for classifying and organizing occupations within the Singaporean context. SSOC uses a 5 digit hierarchical classification, so that every 5 digit code is contained within a 4 digit code and so on for for fewer digits. Table 1 contains an example SSOC 3-digit, 4-digit and 5-digit:

At the highest level of abstraction, major groups of occupations are categorised at the SSOC 1D level given by Table 2.

Throughout the rest of this manuscript, we'll refer to SSOCs and Occupations interchangeably.

3.2. Labour Market Data

Job portals such as "LinkedIn" or "Indeed.com" are widely used by employers to advertise Job Postings for candidates to apply to. Given that the job market evolves continuously, periodically analysing Job Postings provides a way to measure the relative changes in demand for workers, as well as the evolving descriptions of the Job Roles they perform.

By relying on data purchased from a job aggregator company, the authors analysed over 20M job postings between years 2018 and 2023. Each job posting is defined by a title and plain text description. Given the unstructured nature of the data, additional tools were needed to classify a job posting by the occupation it represents and to identify the skills it requires. More details on the classification and extraction tools are presented below in section 3.5.

3.3. SSG Skills Data

SkillsFuture Singapore (SSG) has three variants of skills data. In this section we'll present them in greater detail.

3.3.1. SSG's Skills Framework (SFw)

SSG's Skills Framework (SFw) provides an expert view on the relationships between Occupations and Skills. It is important to note that the SFw defined a taxonomy of skills which were classified into two types: Technical Skills and Competencies (TSCs) and Critical Core Skills (CCS).

TSCs refer to over 11,000 skills that are typically non-transferrable across occupations, such as "Digital Marketing" or "Airport Service Quality Management". CCSs refer to 16 skills typically referred as soft skills, such as "Problem Solving" or "Communication". Every Job Role in the SFw has TSCs and CCSs associated with it. In turn, every Job Role is linked to an occupation found in the SSOC (Occupations are a superset of Job Roles).

We note that the SFw accounts for similar skills belonging to different sectors as two different skills. For example, "Digital Marketing" in the "Tourism" sector is considered a different skill than "Digital Marketing" in the "Financial Services" sector. This is because the SFw is developed and validated by domain experts representing different sectors, such as other statutory boards or industry and trade associations.

Additionally, each skill in the SFw is further broken down into proficiency levels. Each proficiency level, represents a set of knowledge and abilities. An example of this is the TSC skill "Financial Modelling" which has a level 3 proficiency associated to the knowledge "Basic accounting theories of recording and reporting financial information" and a level 6 proficiency associated to the knowledge "Specialised models including real options".

3.3.2. Sector- and proficiency-agnostic skills

As mentioned in the subsection 3.3.1, the SFw labels skills using *sector* allowing for some of the skill titles to be duplicates or synonyms of each other across different sectors. This is also the case for similar skills classified into different proficiency levels. In order to obtain a more general view, SSG has created sector- and proficiency-agnostic skills. After de-duplicating and mapping, there are $\sim 2,000$ sector- and proficiency-agnostic skills (compared to $\sim 11,000$ sector-based skills). We note that this de-duplication was performed based on expert knowledge.

Table 3
This table presents a subset of the mapping of SFw Skills to Sector Agnostic Skills by removing sector-specific terms for the sector agnostic skill *big data analytics*.

Sector	SFw Skill	Sector Agnostic Skill
Healthcare	Data Analytics	big data analytics
Retail	Data Analytics	big data analytics
Food Services	Data Analytics	big data analytics
Precision Eng.	Data Synthesis	big data analytics
Food Manuf.	Data Synthesis	big data analytics
Aerospace	Big Data Analytics	big data analytics
BioPharma Manuf.	Big Data Analysis	big data analytics

An example of the mapping is presented in the Table 3.

During the extraction of skills from a job posting, there might not be any information on what sector the skill belongs to. The sector-agnostic skills are then a suitable choice for automatic extraction of skills from raw text. More details on the extraction will be provided in section 3.5.

3.3.3. Applications and Tools

The skills considered up until now are generic. For example, the skill *programming and coding* does not specify which programming language is involved. When looking for courses and job postings, specifics can be very important. For this reason, SSG has created a list of $\sim$ 1000 *Application and Tools*.

Applications and Tools are often required to deliver specific tasks of many occupations. These refer to brand specific software applications or tools such as the programming language *Python* or the 3D creation platform *Unreal Engine*. As opposed to SFw Skills, Applications and Tools are not related to a specific sector of the economy.

3.4. Courses

The third data source is courses present on the SkillsFuture Singapore portal. Each of these courses has a description of its learning goals (in plain text), alongside other information such as the fees and duration. A subset of the courses also have specific SFw skills and proficiency levels associated to them with the validation of SkillsFuture Singapore.

In order to link Labour Market and Course data with the SFw, we leveraged two different tools to identify the Skills associated with every job post and course listing, and the SSOC associated with every job post analysed. The tools were not developed by the authors, but we will provide an overview of how they work in the following section.

3.5. Tools to link Labour Market and Course data with SFw

Two key entities are required to analyse job postings and course data - Occupation associated with any job posting, and the skills required by a job or taught by a course. By extracting these entities from any job post or course listing, it is possible to aggregate and create a knowledge graph representation in which the entities are *Occupations*, *Skills* and *Courses*.

Two proprietary tools, developed by other teams within the Singapore Public Service, were utilised to achieve this linkage. These two tools rely on Natural Language Processing (NLP) and Named Entity Recognition (NER) ([19], [20] [21]) to perform extractions from unstructured text. In the context of this paper we will only provide a high level explanation of how they work, but will otherwise take them as black boxes as our contribution is centered around how the knowledge base is constructed.

First we will present the tool to extract SSOC: The SSOC Autocoder. The SSOC Autocoder is a tool that leverages transformers [22] and neural networks to determine the most likely 5-digit SSOC code for a body of unstructured text of $<$ 300 words. Its loss function is optimized so that the loss is maximized when the first digits are incorrect, and minimized when the last few digits are incorrect. In this way, if the results are not correct to the 5 digit level, it is possible to make use of the 4-digit code to retain some accuracy.

Secondly, we will present the proprietary tool to extract Skills: The Skill Extraction Algorithm (SEA). Given the possibility of highly similar skills existing in different SFw sectors, the SEA developers decided to utilise the $\sim$ 2000 sector- and proficiency-agnostic de-duplicated skills (see section 3), removing the sectoral dependence. Including the curated list of approximately 1000 Applications and Tools, the SEA is able to extract 3000 unique skills, where each skill has a type given by one of the following: 'TSC', 'CCS' and 'App/-Tools'. The algorithm uses a NER transformer model and cosine similarity to identify skills that most closely match spans extracted from unstructured text found in Labour Market job postings and course descriptions.

We highlight that if several job postings for the same occupation have skills extracted using SEA, it is then possible to count how often each skill is needed for a given occupation. This ability to quantify the presence of a skill for a given occupation will be utilised in the following section to rank the importance of a skill.

3.6. Entities in the Jobs and Skills Knowledge Graph (JSKG)

The knowledge graph is composed of entities intrinsic to the topic of jobs and skills. An example of such entities and their relationships is depicted in Fig 2. The entity types in the JSKG are comprehensively listed here:

1. *SFw Skills*: Defined through expert knowledge in the Skills Framework.
2. *Skills*: Sector- and proficiency-agnostic skills & Application and Tools, extracted using the Skills Extraction Algorithm.
3. *Occupations*: As defined by the SSOC 2020 [9]
4. *Courses*: Available on the MySkillsFuture web portal [23]
5. *Training Providers*: Providers of courses on the MySkillsFuture web portal
6. *Proficiencies*: Specific capabilities within a *SFw Skill*
7. *Job Roles*: Specific roles within an SSOC occupation

At this point we have presented the data sources and the processing done on them to extract the entities (nodes) and relevant relationships between them. In the next section we will present how the entities are linked in the JSKG.

4. Methodology

4.1. Using Job Postings to Proxy Evolving Demand for Occupations and Skills

Job Postings, found on Job Portals such as "LinkedIn", "JobStreet" and "MyCareersFuture", provide a window into the instantaneous demands of the labour market. However, we must highlight that these portals represent a proxy for the total number of jobs in the entire economy. Specific companies/sectors of the economy might rely on alternative ways to attract job seekers and it is important we highlight that a pure reliance on Job Posting portals does not give a complete picture[1].

Additionally, the retrieval at scale of Job Posting from such portals can be a source of inaccuracy in itself. The portal's changes in popularity by employers posting jobs, server downtime and errors in data processing can lead to fluctuating sampling of Job Postings. The outcome of these variations can be large differences in the volume of job postings for two different time periods that are not solely due to an observed economic shift in the workforce, but rather potentially attributable to duplicated postings, re-postings and intermittent accessibility to portals, just to name some causes. For example, the year of 2018 could have a total of ~ 1.5 million job postings, whereas a different time period (the year of 2019) can have up to ~ 2.3 million job postings [2]

This immediately leads us to realize that comparisons on extensive quantities (e.g. job post counts across years) are unreliable. A specific occupation could garner double its job post count from one period to the next, but that could be a data sampling problem rather than a reflection of a stark change in the economy.

To mitigate this problem, we opt to represent and compare quantities derived from Job Postings in terms of intensive quantities (e.g. ranking), which are less sensitive to sampling fluctuations.

Particularly for the case of quantifying the relationship between an occupation and a skill, we can compare the top N most popular skills for a given occupation across the years analysed without having to be as concerned about the absolute number of job postings. The popularity of the skill, for a given occupation, is given by counts of skill presence when extracting job postings associated to a specific occupation. By ranking these counts for a given occupation O, we can then compute a ranking $R(s, O, t)$ of skills s in each time period t. Below, we will present the notation to represent the relationships between occupations and skills.

4.1.1. Determining skill rankings using counts of skill presence

For each skill s and occupation O relationship ($s - O$), we compute a ranking $R(s, O, t)$ in terms of the weight between $s - O$ for a given time period t. The weights are based on the confidence of the SEA extraction, the frequency with which a job posting mentions a skill, and the frequency with which a skill is required for an occupation. Additionally, we can also rank the $s - O$ relationships derived from expert knowledge in the SFw, which is represented as $R_{SFw}(s, O)$.

In section 4.2 we'll use the $1/R(s, O)$ as a score for every $s - O$ relationship. The same is applied to $R_{SFw}(s, O)$ for $s - O$ relationship derived from the SFw. In this way, we can aggregate scores over time periods and incorporate expert knowledge and labour market data. More importantly, it also enables us to deal with the issue of missing $s - O$ relationships during certain periods. By working uniquely with rank, we would be forced to assign a rank value to a missing skill. Operating with inverse values allows us to assign $1/R(s, O) = 0$ for missing skills. Table 4 depicts an example for the Occupation *Data Scientist* on selected skills and year.

4.2. Refinement of the Occupation-Skill relationships

The temporal aspect of the JSKG requires relationships to be updated as new information comes in. In particular, this affects the relationships between occupations and skills. Our goal is to compute a score that can be updated over time to better reflect the reality of the changes in the job market.

Every time period, a new dataset of newly posted job postings is analysed. Aggregated relationships between skills and occupations are thus computed for that period. Although we keep relationships labeled by year, having an **overall consolidated** relationship between occupations and skills is best for analysts that need holistic information, while retaining information of the evolution of the relationship.

Using the SFw as a starting point, we can use each new set of job postings to update or reinforce the original expert knowledge (if any).

There are three considerations at this point:

1. Skills of type *App/Tools*, which are not present in the SkillsFuture's Singapore Skills Framework (SFw). Their starting weight derived from expert knowledge is zero.
2. The SFw had by design 6 skills associated to a given occupation. The SEA is unbounded and able to retrieve any number of skills from a single Job Posting. This implies that skills in SFw will have a boosted weight, as compared to those that are not in the SFw.
3. Expert knowledge (SFw) on TSCs and CCSs is considered ever-green. The weight contribution from the SFw should not decay over time.

To address these considerations, we make a judgment call to take the SFw as a starting point, from which variations are made based on signals from labour market data. Secondly, we compare *App/Tools* separately from TSCs and CCSs, to prevent these from consistently having weaker relationships (by virtue of having zero representation in the SFw).

To control the influence of the ever-green SFw, we can introduce parameter $0 < \epsilon_{\text{SFw}} \leq 1$ to regulate how strong the influence of expert knowledge is on the computed score from labour market signals (job postings). The more time passes before the SFw is updated with expert knowledge, the more ϵ_{SFw} should approach 0.

We propose the mathematical formula given by Eq 2 to compute an evolving score for the weight W of the relationship between an occupation O and a skill s. R_{LM} represents the ranking of the values of a sliding window of length Δy in which weights over the years are computed using

[1] We'd like to note that this study solely focuses on Job Postings listed on Singapore.

[2] There is no observable macroeconomic indicator that signaled that Singapore experienced a $\sim 53\%$ growth in demand of workers, more likely, this increase in job postings is due to a variation in sampling.

Table 4
We present the ranking and scores for Occupation *Data Scientist* and selected skills over the years. We can observe the values of $R(s, O, y)$ and $1/R(s, O, y)$ for $y = 2018$ and the expert knowledge from the Singapore SkillsFuture Skills Framework (SFw), handling the missing skill as 0 contribution.

Skill	R(SFw)	R(2018)	$\frac{1}{R(SFw)}$	$\frac{1}{R(2018)}$
Data and Statistical Analysis	1	2	1.0	0.5
Programming and Coding	2	5	0.5	0.2
Mathematical Concepts Application	missing	38	0	0.026

a weighted average approach. The formula is presented below:

$$R_{\text{LM}}(s, O, y) = R\left(\sum_{y=y_{\max}-\Delta y}^{y_{\max}} \frac{(1 + y_{\max} - y)^{-\frac{1}{2}}}{R(s, O, y)}\right) \quad (1)$$

$$W(s, O) = \frac{\epsilon_{\text{SFw}}}{R_{\text{SFw}}(s, O)} + \frac{1}{R_{\text{LM}}(s, O, y)} \quad (2)$$

Once $W(s, O)$ is computed for every skill and occupation pair, we can once again perform a ranking of W and then arrive at the *Consolidated Skill Requirement* relationship which provides an updated view of the importance of various skills given a specific occupation.

4.3. Computing differentiator skills for an occupation

While the frequency with which a skill is required by an occupation is an indicator of its importance for the tasks the occupation entails, it is also useful to be able to identify skills that differentiate one occupation from another, especially within a cluster of similar occupations. To achieve this, similar to [12], we adopted Term Frequency - Inverse Document Frequency (TF-IDF) given by Eq. 3 to identify differentiator skills for each occupation relative to the full set of SSOC 2020 occupations, as well as clusters defined by the SSOC 2020 hierarchical taxonomy.

Using the *consolidated skills requirement* as the relationship between occupations and skills, differentiator skill scores are calculated using this formula:

$$\text{differentiator score} = \frac{1}{R_{\text{LM}}} \times \left(\log\left(\frac{1 + \text{N}}{1 + \text{n}}\right) + 1\right) \quad (3)$$

The terms in formula are:

- R_{LM} is the rank of the *Consolidated Skill Requirement* detailed in the section above.
- N refers to the total number of occupations within the specified cluster
- n refers to the number of occupations within the cluster requiring the skill in question

This formula represents an adaptation of TF-IDF, in which we substitute the inverse rank of *consolidated skills requirement* relationships for TF. The inverse rank gives a measure of how important a skill is to a specific occupation. This substitution is necessary because *consolidated skills requirement* relationships do not have a direct measure of the frequency with which a skill is needed for a given occupation, which is a more direct parallel for TF.

Table 5 presents example rankings for occupation *Instrumentation Engineer* where top 5 skills are presented using consolidated skills requirement ("CONSOLIDATED") and differentiator skills ("DIFFERENTIATOR").

4.4. Evaluating skill-occupation relationships through Large Language Models

The recent advances in Large Language Models (LLMs) [24], particularly applied to tasks involving Retrieval Augmented Generation (RAG) [25], provide us with an opportunity to perform an otherwise time-consuming human-supervised evaluation of the skills associated to occupations (SSOC), while minimising the effect of hallucinations by providing contextual information. In this case, the team leveraged the AI tool "Open AI *GPT-4o*" [26] function calling, together with a team of Jobs Skills Analysts from SkillsFuture SG to evaluate the labeling of skills-to-occupations relationships produced by GPT-4o (in the form of confidence labels). The way the RAG was utilised to minimise hallucination, was in terms of providing the SSOC Occupation definition (a paragraph of text) given by [9] and the top 50 skills associated to an SSOC 5D through *consolidated skills requirement.* The output of GPT-4o was two labels for every such relationship between skill and occupation - *True / False* for whether the skill was relevant to the occupation, and a confidence level for this evaluation - *high, medium* or *low*, where *high* implies high confidence in the evaluation.

Below is the methodology followed:

1. Take top $N = 50$ TSC and CCS skills associated to an SSOC given by *consolidated skills requirement*
2. Take the title and paragraph definition of SSOC 5D definition given by [9]
3. Prompt OpenAI's GPT-4o to label every selected TSC and CCS skill linked to an SSOC 5D given its occupation definition. Each skill-occupation relationship is to be labeled *True* or *False* with confidence labels *high, medium, low.*
4. Leverage on a team of human Job Skills Analysts in SkillsFuture Singapore to evaluate a sample of the relationship labels produced by GPT-4o

The Jobs Skills Analysts evaluated a sample of $\sim$ 400 SSOC 5D across the total 1002 SSOC 5D in the classification. In that sample, they found that 67% of relationships labeled *False* with *high* confidence should be dropped, as they deemed that the skills are unrelated to the occupations. For skill-to- SSOC relationships labeled *False* by GPT-4o with *medium* and *low* confidence, the human evaluation indicated that most relationships are not unrelated. For those with *medium* confidence, $< 5\%$ were deemed to be unrelated, while for those with high confidence, 0% was deemed to be unrelated. For relationships labeled *True*, for those with *low* confidence, $< 1\%$ were deemed to be unrelated and 0% of those labeled *mid, high* were found to be unrelated. In favour of leaning towards less but true values, all relationships labelled *False* with *high* confidence by GPT-4o were dropped across all SSOC 5D. The downstream

Table 5
Comparison of Consolidated and Differentiator Ranks for Instrumentation Engineer. Although all skills are relevant, Differentiator Skills ranks higher the skills specific to the tasks of an Instrumentation Engineer.

CONSOLIDATED	RANK	DIFFERENTIATOR	RANK
electrical, electronic and control engineering	1	electrical, electronic and control engineering	1
communication	2	instrumentation and control design engineering management	2
failure analysis	3	engineering support management	3
operation management	4	communication	4
decision making	5	instrumentation and control system design	5

implication is that now the top 50 ranked skills might be less than 50 distinct skills on several SSOC 5D. The following section presents the results obtained in applying these methodologies.

5. Results: Bringing it all together

5.1. The JSKG schema

By leveraging the entities presented in section 3.6 we can get a full picture of how occupations and skills come together. In Figure 4 we present the example of the *Instrumentation Engineer* mentioned in Table 5.

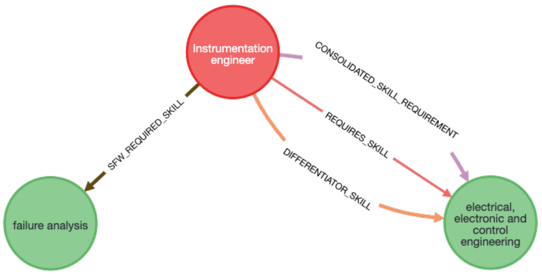


Figure 4: In the case of Instrumentation Engineer, we present the $\#1$ ranking skill for the following relationships: *REQUIRED_SKILL (year:2023)*, *SFW_REQUIRED_SKILL*, *CONSOLIDATED_SKILL_REQUIREMENT* and *DIFFERENTIATOR_SKILL*. All relationships co-exist within the JSKG, allowing analysts to select the most pertinent relationship to the study at hand.

Given that the focus of our discussion has been regarding the refinement/updating of the SSOC-Skill relationship, we'd like to present two illustrative use cases of how the JSKG can be used to derive value for two distinct types of users.

5.2. Evaluation of the Labour Market Inferred Skills

In order to provide an evaluation of the methods to extract skills across all SSOC, we consider the SFw Skill-Occupation linkage as *ground truth* and run a comparison with the skills associated purely through the labour market data as *inference* between the years 2020 to 2023. Borrowing from machine learning approaches, we take the top N skills extracted by those two approaches and compare how these skills overlap using precision and recall metrics in terms of true values (Top N SFw TSC+CCS Skills) and inferred values (Top N Labour Market TSC+CCS Skills). The evaluation can be made in terms of the confusion matrix presented in Table 6. *App/Tools* must be excluded from the comparison, as these are not present in the Skills Framework.

Table 6
The Confusion Matrix is given by the comparison of presence of SFw TSC+CCS Skills and Labour Market (LM) TSC+CCS Skills

	SFw Skill Presence	**SFw Skill Absence**
LM Presence	True Positive (TP)	False Positive (FP)
LM Absence	False Negative (FN)	True Negative (TN)

Precision and recall are given by the following formulas: precision $prec = \frac{TP}{TP+FP}$ and recall $rec = \frac{TP}{TP+FN}$. We can contextualise our findings in terms of the domains of SSOC 1D and by the median amount of job postings analysed per SSOC 5D, which is captured as *Job Density* also on Table 7.

Table 7 also includes the amount of skills-occupation relationships dropped using the approach presented in 4.4, aggregated at the SSOC 1D level. We can observe that high precision is correlated with lower drop rates of occupation-skill relationships. Definitions for each of the major groups in SSOC 1D are given by Table 2.

Across all SSOCs found in both the SFw and labour market data, 71% of the top 10 skills are the same. This decreases to 59% for the top 20 skills and 37% for the top 50 skills. Some divergence in the skills signals for the labour market and the SFw is expected due to changes over time, so these figures suggest that the skill-occupation relationships in the knowledge graph are most useful when considering the most important skills for each occupation. This is particularly so for jobs that are advertised more commonly in Singapore (high job post count). Possible ways to improve the rest of the skill-occupation relationships further are covered in section 7.1.

5.3. Illustrative Case Studies

Next, we illustrate the use of the JSKG by providing two different customer use cases.

Mid-career Switcher. The user is a mid-career marketing manager who is considering their options to exit their current occupation in favour of one that is more rewarding monetarily and that brings some new challenges. The user needs to be able to take stock of their existing Skills and expected pay in their occupation, and use that information to gauge adjacent occupations in terms of Skill similarity and salary increment.

The user is able to tap on a web portal linked to the JSKG to perform a comparison in terms of skill similarity, obtaining the occupations that have the most skills in common, while providing an increase in pay. The user can then retrieve the skill gap through an occupation comparison and

Table 7
Comparison of SSOC1D Precision, Recall, Job Counts, Total number of Job Posts and LLM Drop Rate. We observe that SSOC 1D where occupations have larger amounts of job post counts, the precision and recall is higher. We take note that the Job Density seems uncorrelated from LLM Drop Rate, whereas high precision correlates with low drop rate.

SSOC1D	prec@N=10	prec@N=20	prec@N=50	rec@N=10	rec@N=20	rec@N=50	Jobs Density	Job Posts	Drop Rate
1	0.81	0.74	0.53	0.5	0.51	0.5	777	469944	1.00%
2	0.73	0.61	0.38	0.45	0.45	0.44	283	824103	0.70%
3	0.64	0.52	0.3	0.36	0.36	0.36	125	387261	3.30%
4	0.6	0.46	0.26	0.36	0.38	0.38	46	80610	8.70%
5	0.64	0.47	0.26	0.37	0.37	0.37	175	85006	4.60%
6	0.53	0.33	0.16	0.31	0.31	0.31	4	76	5.80%
7	0.46	0.3	0.15	0.36	0.36	0.36	105	18697	2.10%
8	0.54	0.37	0.21	0.38	0.38	0.38	15	28731	2.90%
9	0.26	0.22	0.22	0.08	0.08	0.08	159	70548	11.70%
All	0.71	0.59	0.37	0.42	0.43	0.43	717	1964976	2.60%

find optimal courses that will equip them with the required skills for the best cost and least amount of time. In this example, the marketing manager can identify the occupation of *Business Development Manager* as an occupation with 9 skills in common and a skills gap of 9 skills. The skill gap can then be fulfilled by identifying the optimal course set required to achieve such skills for the minimum cost, duration, and wait time to earliest opportunity for training.

At this point, the user can get a sense of the difficulty involved in making the career switch, but can also quantify the monetary and time commitment required to achieve basic literacy to bridge the skill gap and achieve their goals of switching careers.

The Learning and Development HR analyst. Some of the key tasks for a L&D analyst involve three aspects: (i) analyzing the gaps between the current skills of employees and the skills required for their roles or for future organizational needs; (ii) Designing and developing training programs and learning materials tailored to address identified skill gaps;(iii) Implementation of Training Programs to re-skill or up-skill the workforce to achieve organizational needs organically.

The L&D analyst, might be tasked to identify roles for which demand is declining to train individuals to move into adjacent occupations with higher demand, such as training secretaries to become compliance officers. By comparing occupations, the analyst identifies the usual skill gaps between the roles. In this way, it is possible to determine the optimal course set to obtain the missing differentiator skills to achieve basic literacy in the new roles. By considering the costs in terms of fees for training, the analyst can also determine the optimal training strategy and the programs to carry out such transformation.

6. Lessons Learned and Challenges

One of the primary challenges encountered was with the Skills Extractor Algorithm (SEA), which at times provided incorrect skill extractions. This issue arose from the need for the tool to operate quickly enough to ensure computational feasibility. Consequently, skills were sometimes omitted or incorrectly extracted based on the phrasing of job descriptions. To curb the incorrectly extracted skills, we found that ranking extracted skills by their popularity typically relegated incorrect extractions to lower tiers. Furthermore, relying on ground truth SkillsFuture (SFw) skills and updating them with labour market data helped to purge incorrect extractions. Additionally, using the GPT-4 Large Language Model (LLM) to prune skills, based on its general knowledge, further reduced the number of incorrect extractions, albeit at the cost of occasionally removing some correct ones.

We also learned that rankings, rather than absolute values, provide more intuitive insights to users. Given the wide variation in the popularity of occupations, absolute numbers often do not convey meaningful information. For instance, in our job posting dataset for 2018, the occupation "Marketing Manager" had 2039 extractions, while "Management Consultant" had only 13. In that same year, the skill "Business Opportunities Development" appeared 314 times for "Marketing Manager," making it the 13th most popular skill for that occupation, whereas for "Management Consultant," the same skill appeared 10 times, ranking as the most popular skill for that occupation.

We faced the dilemma of whether to feature jobs that are not typically offered in the job market, such as "Legislator," through synthetic methods such as large language models (LLMs). The absence of such jobs in the data is likely due to the nature of hiring for those occupations rather than a lack of data. However, given the definitions of these occupations and their associated tasks, it is feasible to include them in the knowledge graph through synthetic means.

Another challenge was dealing with changes in skill or occupation taxonomy. Should the SEA extraction tool adopt a new list of skills (e.g., new Technical Skills and Competencies (TSCs) included) or if a new version of the SSOC is released, a re-extraction of SSOCs or skills across all historical job postings would be necessary. This re-extraction is required for consistency and to ensure fair comparison.

Finally, Section 4.2 discusses the use of $1/R(s, O)$ as a score to compute the aggregated effect within a sliding time window and differentiator scores. We found that using a succession of $\frac{1}{n}$ provides an over-estimation of the difference between the highest-ranking skills compared to directly comparing rankings $R(s, O)$. However, the variation in the number of elements across rankings makes the direct use of $R(s, O)$ challenging, particularly when a skills-occupation pair is absent in certain time periods. This $\frac{1}{n}$ approach helps address the problem of missing rankings, offering a practical solution.

7. Conclusion

This research contributes to a deeper understanding of workforce dynamics and supports timely and informed decision-making in skills matching and career planning initiatives. The creation of the Jobs Skills Knowledge Graph delineates a dynamic repository of occupational roles, skills, and associated entities within the realm of continuous adult education. Leveraging occupation and skill classifiers serves as foundational mechanisms for constructing a knowledge base of occupations and competencies. Central to this endeavour is the ongoing refinement and updating of relationships between occupations and skills, which serves as cornerstone for deriving ever-green graph analytics insights. The collaboration of LLMs and human experts further refines the end-result. An evaluation of the results of this methodology is presented in terms of precision and recall, after the human-LLM layered drops the more unrelated skill associations. The evaluation results throw light on which occupations major groups (SSOC 1D) are more accurately represented by the methodology proposed.

7.1. Future Work

Through the linkages derived between occupations and skills, it is possible to derive a network purely of skills (or occupations). Such a network allows us to investigate the inter-relationship of skills. This is particularly relevant for the linkage between Technical Skill Competencies (TSCs) and the ever growing list of Applications/Tools that allow the skill to be applied.

Investigating the network using centrality metrics or clustering methods could enable us to identify key clusters of skills that are highly transferable or that are often required together. These can serve as strong recommendations for training providers to adapt their offerings to the ever-changing landscape of the needs of the economy.

Robust evaluation of the results remains only partially addressed. Although the precision and recall results in Table 7 give a quantified figure of how well the occupation categories (SSOC 1D) match expert knowledge, this does not tell us if the difference is due to an evolution in the job definition in Singapore, or due to limitations in the approach. Potential verifications could involve leveraging datasets from other countries (which would need to be contextualised in the taxonomies of occupation and skills used in this paper) or leveraging on another round of costly Labour Market experts to validate the findings with updated validated relationships. Both approaches are highly taxing in terms of effort and cost, however, rounds of human validation of small samples can give indications of the benefits of the method for smaller costs.

Lastly, recent growing interest in applications between knowledge graphs and large language models, underline the importance of reliable knowledge bases to power such models through graph retrieval-augmented generation (GraphRAG) [27]. Areas of further refinement lie in the creation of triplets for the knowledge graph from unstructured text [28] [29], to the alleviation of hallucinations in Large Language Models (LLMs) [30] through the facts captured in knowledge graphs.

Acknowledgments

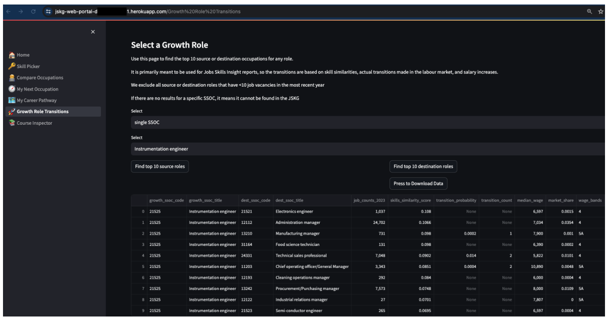
The authors would like to gratefully acknowledge the support of SkillsFuture Singapore in facilitating the research and application of the Jobs and Skills Knowledge Graphs. The authors also extend their sincere thanks to Ong Meng Hwee Victor for his valuable comments and help with the review of the manuscript.

Disclaimer: The content of this paper do not imply that the research outcomes are currently deployed or planned for deployment by SkillsFuture Singapore at any given time. The views and opinions expressed in this paper are solely those of the authors and do not necessarily represent those of SkillsFuture Singapore, Government Technology Agency of Singapore or A-STAR.

A. Software tools utilised to build and deploy a proof of concept JSKG

The graph database utilized was *Neo4j* version 5.3, selected due to its efficiency and ease in handling relationships within complex datasets. The application programming interface (API) implemented for seamless communication and data exchange was *FastAPI*.

The web portal presented in Figure 5 was crafted using *Streamlit*, a user-friendly and interactive framework for data visualization and exploration. To ensure the accessibility and availability of the developed API and web portal, both were deployed on the *Heroku* platform, offering a scalable and reliable cloud infrastructure. For the hosting of the Neo4j graph database, the *AuraDB* service provided by Neo4j was utilized, facilitating a cloud-native, fully managed database-as-a-service solution.



jskg-web-portal-d[redacted]1.herokuapp.com/Growth%20Role%20Transitions

Select a Growth Role

Use this page to find the top 10 source or destination occupations for any role.

It is primarily meant to be used for Jobs Skills Insight reports, so the transitions are based on skill similarities, actual transitions made in the labour market, and salary increases.

We exclude all source or destination roles that have <10 job vacancies in the most recent year

If there are no results for a specific SSOC, it means it cannot be found in the JSKG

Select: single SSOC

Select: Instrumentation engineer

	growth_ssoc_code	growth_ssoc_title	dest_ssoc_code	dest_ssoc_title	job_counts_2023	skills_similarity_score	transition_probability	transition_count	median_wage	market_share	wage_bands
0	21525	Instrumentation engineer	21521	Electronics engineer	1,037	0.108	None	None	6,597	0.0015	4
1	21525	Instrumentation engineer	12112	Administration manager	24,702	0.1066	None	None	7,034	0.0354	4
2	21525	Instrumentation engineer	13210	Manufacturing manager	731	0.098	0.0002	1	7,900	0.001	5A
3	21525	Instrumentation engineer	31164	Food science technician	131	0.098	None	None	6,390	0.0002	4
4	21525	Instrumentation engineer	24331	Technical sales professional	7,048	0.0902	0.014	2	5,822	0.0101	4
5	21525	Instrumentation engineer	11203	Chief operating officer/General Manager	3,343	0.0851	0.0004	2	10,890	0.0048	5A
6	21525	Instrumentation engineer	12193	Cleaning operations manager	292	0.084	None	None	6,000	0.0004	4
7	21525	Instrumentation engineer	13242	Procurement/Purchasing manager	7,573	0.0748	None	None	8,000	0.0109	5A
8	21525	Instrumentation engineer	12122	Industrial relations manager	27	0.0701	None	None	7,807	0	5A
9	21525	Instrumentation engineer	21523	Semi-conductor engineer	265	0.0695	None	None	6,597	0.0004	4

Figure 5: Screenshot of the web portal powered by the JSKG. The experimental widget displayed allows users to search for potential next occupations in terms of skill similarity, job availability, and wage increase.

References

[1] S. Debortoli, O. Müller, J. vom Brocke, Comparing business intelligence and big data skills: A text mining study using job advertisements, Business Information Systems Engineering 6 (2014) 289–300. URL: http://dx.doi.org/10.1007/s12599-014-0344-2. doi:10.1007/s12599-014-0344-2.

[2] A. De Mauro, M. Greco, M. Grimaldi, P. Ritala, Human resources for big data professions: A systematic classification of job roles and required skill sets, Information Processing & Management 54 (2018) 807–817. URL: http://dx.doi.org/10.1016/j.ipm.2017.05.004. doi:10.1016/j.ipm.2017.05.004.

[3] I. Khaouja, I. Kassou, M. Ghogho, A survey on skill identification from online job ads, IEEE Access 9

(2021) 118134–118153. URL: http://dx.doi.org/10.1109/ACCESS.2021.3106120. doi:10.1109/access.2021.3106120.

[4] International Labour Office, The International Standard Classification of Occupations (ISCO-08) Companion Guide, Geneva, 2023.

[5] J. D. Smedt, M. le Vrang, A. Papantoniou, Esco: Towards a semantic web for the european labor market, in: LDOW@WWW, 2015. URL: https://api.semanticscholar.org/CorpusID:14184714.

[6] National Academies Press, 2010. URL: http://dx.doi.org/10.17226/12814. doi:10.17226/12814.

[7] M.-I. Dascalu, I. Marin, I. V. Nemoianu, I.-F. Puskás, A. Hang, An ontology for educational and career profiling based on the romanian occupation classification framework: Description and scenarios of utilisation, in: ICERI Proceedings, ICERI2022, IATED, 2022. URL: http://dx.doi.org/10.21125/iceri.2022.1881. doi:10.21125/iceri.2022.1881.

[8] Skills frameworks, 2023. URL: https://www.skillsfuture.gov.sg/skills-framework.

[9] Singapore standard occupational classification ssoc 2024, 2024. URL: https://www.singstat.gov.sg/standards/standards-and-classifications/ssoc.

[10] C. Joyez, C. Laffineur, The occupation space: network structure, centrality and the potential of labor mobility in the french labor market, Applied Network Science 7 (2022). URL: http://dx.doi.org/10.1007/s41109-022-00453-3. doi:10.1007/s41109-022-00453-3.

[11] A. Alabdulkareem, M. R. Frank, L. Sun, B. AlShebli, C. Hidalgo, I. Rahwan, Unpacking the polarization of workplace skills, Science Advances 4 (2018). URL: http://dx.doi.org/10.1126/sciadv.aao6030. doi:10.1126/sciadv.aao6030.

[12] Y. Fettach, M. Ghogho, B. Benatallah, Knowledge graphs in education and employability: A survey on applications and techniques, IEEE Access 10 (2022) 80174–80183. doi:10.1109/ACCESS.2022.3194063.

[13] A. Vista, Data-driven identification of skills for the future: 21st-century skills for the 21st-century workforce, SAGE Open 10 (2020) 215824402091590. URL: http://dx.doi.org/10.1177/2158244020915904. doi:10.1177/2158244020915904.

[14] H. Paulheim, Knowledge graph refinement: A survey of approaches and evaluation methods, Semantic Web 8 (2016) 489–508. URL: http://dx.doi.org/10.3233/SW-160218. doi:10.3233/sw-160218.

[15] M. de Groot, J. Schutte, D. Graus, Job posting-enriched knowledge graph for skills-based matching, ArXiv abs/2109.02554 (2021). URL: https://api.semanticscholar.org/CorpusID:237420688.

[16] S. Ji, S. Pan, E. Cambria, P. Marttinen, P. S. Yu, A survey on knowledge graphs: Representation, acquisition, and applications, IEEE Transactions on Neural Networks and Learning Systems 33 (2022) 494–514. URL: http://dx.doi.org/10.1109/TNNLS.2021.3070843. doi:10.1109/tnnls.2021.3070843.

[17] S. Ji, S. Pan, E. Cambria, P. Marttinen, P. S. Yu, A survey on knowledge graphs: Representation, acquisition, and applications, IEEE Transactions on Neural Networks and Learning Systems 33 (2022) 494–514. doi:10.1109/TNNLS.2021.3070843.

[18] X. Zou, A survey on application of knowledge graph, Journal of Physics: Conference Series 1487 (2020) 012016. URL: http://dx.doi.org/10.1088/1742-6596/1487/1/012016. doi:10.1088/1742-6596/1487/1/012016.

[19] M. Zhang, K. Jensen, R. van der Goot, B. Plank, Skill extraction from job postings using weak supervision (2022). doi:10.48550/arXiv.2209.08071.

[20] S. Fareri, N. Melluso, F. Chiarello, G. Fantoni, Skillner: Mining and mapping soft skills from any text, Expert Systems with Applications 184 (2021) 115544. URL: http://dx.doi.org/10.1016/j.eswa.2021.115544. doi:10.1016/j.eswa.2021.115544.

[21] N. H. N. Minh, N. K. Doan, P. Q. Huy, K. X. Loc, H. N. Vu, H. Nguyen, H. C. Phap, Information Technology Skills Extractor for Job Descriptions in vku-ITSkills Dataset Using Natural Language Processing, Springer Nature Switzerland, 2023, p. 250–261. URL: http://dx.doi.org/10.1007/978-3-031-36886-8_21. doi:10.1007/978-3-031-36886-8_21.

[22] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, I. Polosukhin, Attention is all you need, in: Proceedings of the 31st International Conference on Neural Information Processing Systems, NIPS'17, Curran Associates Inc., Red Hook, NY, USA, 2017, p. 6000–6010.

[23] My skillsfuture portal, 2024. URL: https://www.myskillsfuture.gov.sg/.

[24] S. Minaee, T. Mikolov, N. Nikzad, M. Chenaghlu, R. Socher, X. Amatriain, J. Gao, Large language models: A survey, 2024. URL: https://arxiv.org/abs/2402.06196. doi:10.48550/ARXIV.2402.06196.

[25] X. Wang, Z. Wang, X. Gao, F. Zhang, Y. Wu, Z. Xu, T. Shi, Z. Wang, S. Li, Q. Qian, R. Yin, C. Lv, X. Zheng, X. Huang, Searching for best practices in retrieval-augmented generation, 2024. URL: https://arxiv.org/abs/2407.01219. doi:10.48550/ARXIV.2407.01219.

[26] Gpt-4 turbo (gpt-4o), 2024. URL: https://platform.openai.com/docs/models/gpt-4o.

[27] Y. Hu, Z. Lei, Z. Zhang, B. Pan, C. Ling, L. Zhao, Grag: Graph retrieval-augmented generation, 2024. arXiv:2405.16506.

[28] S. Carta, A. Giuliani, L. Piano, A. S. Podda, L. Pompianu, S. G. Tiddia, Iterative zero-shot llm prompting for knowledge graph construction, arXiv preprint arXiv:2307.01128 (2023).

[29] A. Chepurova, A. Bulatov, Y. Kuratov, M. Burtsev, Better together: Enhancing generative knowledge graph completion with language models and neighborhood information, in: Findings of the Association for Computational Linguistics: EMNLP 2023, Association for Computational Linguistics, 2023. URL: http://dx.doi.org/10.18653/v1/2023.findings-emnlp.352. doi:10.18653/v1/2023.findings-emnlp.352.

[30] W. Fan, S. Wang, J. Huang, Z. Chen, Y. Song, W. Tang, H. Mao, H. Liu, X. Liu, D. Yin, Q. Li, Graph machine learning in the era of large language models (llms), 2024. URL: https://arxiv.org/abs/2404.14928. doi:10.48550/ARXIV.2404.14928.