Classifying occupations according to their skill requirements in job advertisements

Jyldyz Djumalieva[1], Antonio Lima[1], and Cath Sleeman[1]

[1]Nesta

March 28, 2018

Abstract

In this work, we propose a methodology for classifying occupations based on skill requirements provided in online job adverts. To develop the classification methodology, we apply semi-supervised machine learning techniques to a dataset of 37 million UK online job adverts collected by Burning Glass Technologies. The resulting occupational classification comprises four hierarchical layers: the first three layers relate to *skill specialisation* and group jobs that require similar types of skills. The fourth layer of the hierarchy is based on the offered salary and indicates *skill level*. The proposed classification will have the potential to enable measurement of an individual's career progression within the same skill domain, to recommend jobs to individuals based on their skills and to mitigate occupational misclassification issues. While we provide initial results and descriptions of occupational groups in the Burning Glass data, we believe that the main contribution of this work is the methodology for grouping jobs into occupations based on skills.

Acknowledgements

The authors are grateful for the thoughts of colleagues at Nesta, the Economic Statistics Centre of Excellence and the Office for National Statistics on this work. Particular thanks are due to Hasan Bakhshi for his comments on early drafts.

Introduction

In this work we propose a methodology for developing an occupational classification by applying Natural Language Processing methods, such as document clustering and distributed word representations, to UK online job adverts. The new occupational classification will be directly aligned with employer needs and group jobs into occupations based on similar skill requirements. Unlike the existing UK Standard Occupational Classification taxonomy, the skills based occupational classification methodology will prioritise *skill specialisation* over *skill level*. The term *skill level* refers to the amount of education and training required as well as the range of tasks performed; *skill specialisation* refers to domain-specific expertise, technology and materials used, and the products and services produced in a given occupation (International Labour Organization, 2016). The resulting classification will have the potential to enable measurement of an individual's career progression within the same skill domain, to recommend jobs to individuals based on their skills and to mitigate occupational misclassification issues.

Standard Occupational Classification (SOC) taxonomies organise jobs into meaningful groups based on work performed as well as skills, knowledge and qualifications required to competently perform typical tasks and duties. Systematic classification of occupations serves multiple purposes. First it ensures comparability of occupational data collected through various sources (Cosca and Emmel, 2010). It also lays the foundation for measuring changes over time in the distributions of workers across occupations. A wide audience, including individuals, employers, educators and policymakers, use labour market insights to support their decision making.

To provide the most value to users, SOC taxonomies should accurately reflect the nature of work and skill requirements, which change constantly due to technological, demographic and environmental shifts. This is the reason why occupational classifications are regularly revised. However, the revision process requires substantial investment of time and resources. Most SOC taxonomies have a 10-year revision cycle (Cosca and Emmel, 2010; Elias and Birch, 2010). The revision process itself takes a long time since it relies on extensive review of each occupational group by expert panels and consultation with the public. Over the course of 10 years the landscape for some occupations may change significantly, like it did for IT professionals between 2000 and 2010, necessitating the addition of new occupations to the UK SOC (Elias and Birch, 2010). Given that structural changes will continue to impact the labour market (Bakhshi, Downing, Osborne, and Schneider, 2017), there is a need to capture information on occupational dynamics in a more timely way.

Using online job adverts for occupational classification and analysis can help address this need. Traditionally, data on occupations are collected through surveys, which are restricted in their frequency and scope due to the associated costs. Unlike surveys, it is possible to efficiently collect labour market information from online job adverts in near real-time and at scale. While using online job advert data has its drawbacks, which we describe in further sections, the advantages, such as level of detail, the time and cost effectiveness of collection and increasing coverage, justify the use of this rich data source for understanding the demand side of the labour market.

Instead of mapping online vacancies to existing SOC, we propose developing an alternative occupational classification based on employer skill requirements for the following reasons. First, using employer skill requirements for organising jobs into occupations will ensure that the resulting occupational classification accurately reflects employer needs and is, therefore, immediately relevant for job seekers and people preparing to enter work. Second, the emphasis of the UK's current SOC classification principles on *skill level* (over *skill specialisation*) makes it more difficult to plan and measure individuals' career progressions since jobs with similar *skill specialisations* may be spread across different major groups. The *skill level* is also determined to a large extent by the formal qualifications required in an occupation and these requirements may change because of external factors that are unrelated to the nature of the job itself. Finally, coding online vacancies to existing UK SOC is challenging as correct assignment of a *skill level* is not easy to achieve with online job adverts.

The remainder of the paper is organised as follows. In the Related work section, we describe the advantages and drawbacks of using online job adverts as a source of labour market information. We also provide more detail on the rationale for developing a new skills based occupational classification. The datasets used and the process for generating occupational classification layers are outlined in the Data and Methodology sections respectively. The outputs of the proposed methodology are summarised in Results. In the Discussion section, we review the contributions and limitations of this work. We conclude with key takeaways and directions for future research.

Related work

As more job advertisements are moved online, real-time data on vacancies are becoming more readily available. According to some estimates, up to 70% of job openings are now posted online (Carnevale, Jayasundera, and Repnikov, 2014) and this figure is expected to rise going forward (Askitas and Zimmermann, 2015). In addition to the improving coverage of the underlying labour market, there are several other advantages of using online job adverts to analyse skill demands. First, the free text fields in job adverts allows employers to directly express their needs: job postings include specific descriptions of skills, qualifications and credentials required to perform the job. A second advantage is that the adverts provide a highly granular view on vacancies making it possible to disaggregate data geographically or by industry.

Using online job vacancy data has its limitations and occupation representativeness is one of the largest drawbacks (Carnevale et al., 2014; Kureková et al., 2015). There are alternatives to advertising vacancies online, including tender, audition, offline advertisements, and word of mouth, which are often used in some occupations. Online postings tend to be biased toward high-skilled professional occupations, and therefore estimates of vacancy levels in the economy cannot be directly inferred from online job postings. The quality of the data may also be worse than in structured surveys, as online job adverts often contain abbreviations and misspellings. Adverts may also be incomplete or a single posting may be used to advertise multiple positions. Terms used to describe job titles and skills vary to a large extent, which makes it challenging to standardise these terms across employers. While the issues of data representativeness and quality are significant, the advantages of online job adverts make it a useful source of information on labour market demand.

Online job adverts are increasingly used to enhance our understanding of the labour market. Early studies tended to examine small sample sizes and manually code advert content to identify key themes (Harper, 2012). However, as online job vacancy data became more accessible, researchers have started to apply advanced analytical techniques to process large volumes of job postings. Studies also demonstrate how skill requirements in online data can help refine economic statistics. For example, Deming and Kahn (2017) established a positive link between the requirements for social and cognitive skills mentioned in adverts and wage differences even after controlling for education, experience and geographic location. The authors also found that firms which had higher demand for both types of skills demonstrated better financial performance. The findings on both pay and firm performance show that including skill data in econometric models can add explanatory power beyond that offered by other commonly available labour market indicators.

In another study, Grinis (2017) investigated the extent to which STEM (Science, Technology, Engineering, and Mathematics) skills were in demand in non-STEM occupations. Grinis developed a machine learning approach for classifying jobs into STEM and non-STEM groups using keywords provided in job adverts. When applied to 33 million job postings, the approach showed that a large proportion of vacancies with STEM skill requirements resided in occupations traditionally considered as not requiring STEM training, such as *Product, clothing and related designers*. The findings imply that the demand for STEM skills and knowledge is underestimated.

To date, researchers have mapped online vacancies to existing SOC taxonomies (Boselli et al., 2017; Gweon et al., 2017). However, we believe that a new skills based occupational classification is needed for several reasons. First, such a classification will be directly aligned with the needs of employers as expressed in adverts. This will make the classification highly relevant to job seekers and young people preparing to get their first job.

Focusing on skill requirements can also help to explore the limitations of the existing UK SOC classification principles. In the UK SOC 2010, similar to the International Standard Classification of Occupations (ISCO) and the Canadian National Occupational Classification (International Labour Organization, 2016; ESDC, 2017), the *skill level* is the primary criterion for grouping occupations into the major groups, which range from Managers, Directors and Senior Officials (major group 1) to Elementary Occupations (major group 9). Occupations are then separated based on *skill specialisation* within each major group. Because *skill level* is prioritised over *skill specialisation,* jobs which require similar skills may be assigned to completely different major groups. For example, Cost accountants can reside both in major SOC groups 2 and 4 depending on whether the employee needs a professional qualification. This approach makes it more difficult to track an individual's career progression within the same skill domain.

The UK SOC system is also susceptible to changes in qualification requirements. According to UK SOC classification principles, a formal qualification is an important criterion for assigning occupations to major groups (Thomas and Elias, 1989). When nursing became a profession, which individuals increasingly enter via degree-level route, all nurses were moved from major group 3 to major group 2 in the 2010 SOC revision (Elias and Birch, 2010). As this example illustrates, the dependence of SOC on qualifications can add volatility to the SOC structure.

The sensitivity of SOC to qualification requirements may also be exacerbated by the expansion of higher education sector in the UK. A recent report indicates that the level of under-utilisation of graduate level qualifications at the workplace is higher in the UK than in other European countries. The proportion of UK graduates entering jobs that do not require a graduate level qualification has also grown faster in the UK than in other EU countries (Brinkley and Crowley, 2017). Due to SOC's emphasis on *skill level*, an occupation might be reallocated to a different major group if an increasing share of employees hold a higher level qualification, and not necessarily as a result of a change in the actual job content or skill requirements.

While the *skill level* distinctions captured at the major level of SOC are meaningful, they pose practical challenges for coding occupations to SOC, especially in the case of automated coding. It might be difficult to capture distinctions in *skill level*, when a vacancy description does not specify qualification requirements. This issue can lead to inaccurate SOC code assignment. Belloni et al. (2014) have recently estimated that even at the 1-digit level of ISCO, in at least 33% of cases there was a discrepancy in the codes assigned by two different automated coding methods. The misclassification rates pose concerns since SOC and ISCO codes are subsequently used to measure employment and other labour market statistics. The skills based occupational classification proposed in this paper starts with *skill specialisation*, which may increase the consistency of automatic coding systems applied to online job adverts.

With regards to related work on developing occupational classifications, efforts to investigate online vacancy data from a methodological perspective have been largely concentrated in the private sector. In this space, research has been carried out by labour analytics companies, job search engines and recruitment agencies (Danger, 2016; Javed and Jacob, 2015; Posse, 2016). For these organisations the primary motivation for developing an occupational taxonomy is to improve the efficiency of matching job applicants to available opportunities. Another objective is to build commercial products on labour market intelligence, such as salary trends or dashboards on emerging skillsets (Burning Glass Technologies, 2018; Emsi, 2018). While the research published by these entities provides useful insights on analytical techniques to generate taxonomies, the resulting occupational classifications remain proprietary.

We believe that the key contribution of this paper is in providing one of the first data-driven methodologies for grouping job adverts into occupations based on the skills contained within those adverts. There is a growing recognition of the importance of taking in empirically-driven approach to analysing labour demand. For example in their recent work Turrell et al. (forthcoming) propose a *bottom-up* segmentation of the UK labour market to study the mismatch between the unemployed and job vacancies. The authors demonstrate that their data-driven solution is capable of identifying both traditional jobs as well as sub-markets not reflected in the UK SOC. Turrell et al. also show that the *bottom-up* segmentation offers explanatory power for both offered and agreed wages. The authors follow a similar approach to the one we propose, using unsupervised machine learning techniques to group online job adverts. However, they focus on the *skill specialisation* aspect of the occupations, identifying 20 occupation clusters, and do not explore the *skill level* dimension.

The methodology proposed in this work will be publicly available and will provide policymakers and researchers with a framework for analysing demands for both broad and domain-specific skills.

Data

We carried out the analysis using online job adverts provided by Burning Glass Technologies, a labour market analytics company. Every day Burning Glass scrapes and processes up to 3.4 million active job postings from thousands of web-pages (Burning Glass Technologies, 2017). Along with over 70 elements of metadata, requirements on skills, experience and qualification are extracted from job postings and standardised with the help of Burning Glass's proprietary algorithm.

The data in our sample were collected by Burning Glass over a five-year period, from January 2012 to December 2016. Each job advert contains a set of keywords extracted from the job's description, however the full job descriptions are not available. While we refer to the keywords as 'skills', these also include terms that describe personal characteristics, industry experience, knowledge and non domain-specific skills. In total, there are 36,699,666 adverts in the dataset. It is important to note that there are many adverts with missing information: only 61% of adverts contain data on offered salary, and substantially fewer mention education (19% of adverts) and experience requirements (13% of adverts).

In addition to the job adverts we also used two publicly available resources: the ONS 2010 Index (Office for National Statistics) and the European Dictionary of Skills and Competences (DISCO). The ONS Index provides a reference list of known job titles and a corpus of terms used to describe occupations. It identifies up to 30,000 alternative job titles across all occupational unit groups. We use this information in the data cleaning stage to remove non job related terms in job titles. The DISCO is a multilingual, peer-reviewed thesaurus used to classify, describe and translate skills and competences (DISCO II Portal). It has been incorporated in European classification of Skills, Competences, Occupations and Qualifications, which is a Europe 2020 initiative by the European Commission with aims to systematise skills, competences, occupations and qualifications. The DISCO divides skills into 9 non domain-specific categories and 25 domain-specific categories. Specific examples of skills from both categories were used to assign job postings to relevant skill categories.

Methodology

The proposed methodology groups occupations hierarchically, in line with existing occupational classifications. However, unlike ONS SOC or ISCO, in our classification *skill specialisation* (domain-specific expertise, knowledge of technology, materials used, products and services produced in a given occupation) is given priority over *skill level* (the measure of complexity and range of tasks performed). As shown in Figure 1, we use skills mentioned in a job advert to understand the nature of the job (*skill specialisation*) and, subsequently, we infer job seniority (*skill level*) using the data on nominal offered salaries from job adverts. Focusing first on *skill specialisation* makes the proposed taxonomy more similar to U.S. occupational taxonomy (U.S. Bureau of Labor Statistics, 2010), where the first level of the hierarchy is a set of 23 major groups, such as Business and Financial Operations Occupations, Computer and Mathematical Occupations, Architecture and Engineering Occupations, etc.

The methodology for building a skills based classification was developed in several stages, which correspond to the layers outlined in Figure 2. While the resulting classification is hierarchical, the development of the methodology started with the second (*skill category)* layer. We take a semi-supervised approach and use an existing set of *skill specialisations* (namely the first layer of DISCO) to guide the grouping of job adverts. We found that in the Burning Glass adverts some areas, like agriculture , were underrepresented, while vacancies with requirements for business skills (i.e. sales, marketing, finance, etc.) were overrepresented. Applying an unsupervised technique to this data would prevent us from capturing distinct categories if they represent a small proportion of the job postings.

The *skill category* layer described above is the starting point for the developing the taxonomy. The very first layer of the hierarchy (*broad group* layer) is created by applying hierarchical clustering to aggregate the skill categories based on their similarity. From the *skill category* layer, we go down the hierarchy to create finer skill categories (*sub-category* layer). Finally, we form the *skill level* layer, which divides each *sub-category* layer into groups based on different salary intervals. Before describing each layer in more detail, we briefly outline the process we use to prepare the job adverts for analysis.

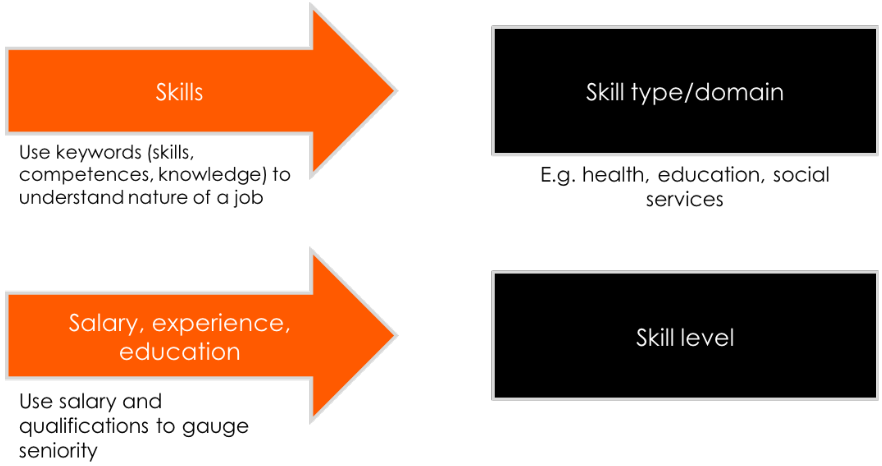

Figure 1: Using Burning Glass data to infer *skill specialisation* and *skill level*

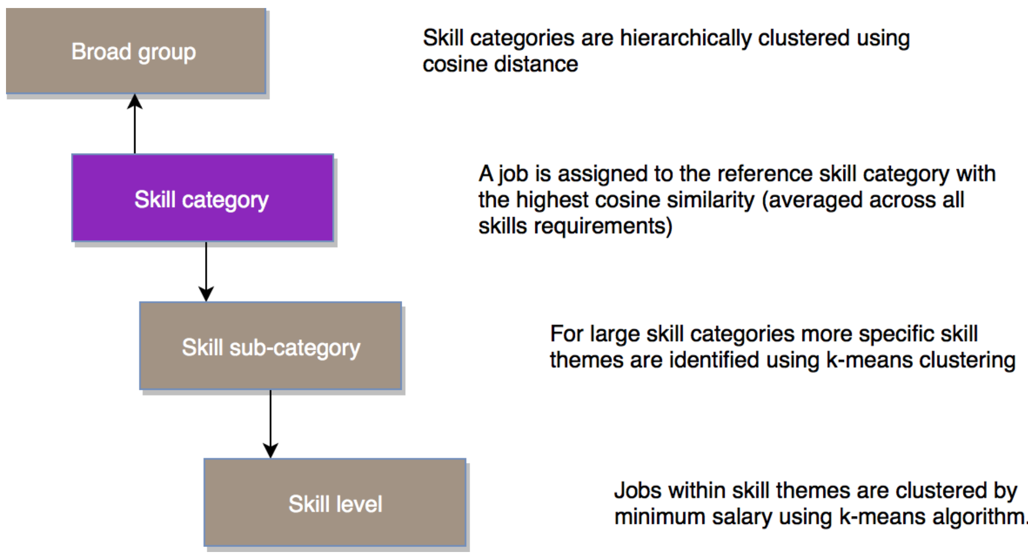

Figure 2: Layers of skills based occupational classification

Data preparation

There were a number of steps taken to prepare the data for further analysis. The job titles in online adverts often contain terms that are not directly relevant to the role, such as the job's location or the type of employment. Owing to this and other factors, job titles are highly diverse, though this diversity is often uninformative and poses challenges for identifying underlying occupations. To overcome this challenge, the job titles were processed to reduce the amount of noise. This process involved expanding abbreviations, removing words not in the ONS Index, and removing most punctuation and digits (Figure 3).

In contrast to job titles, the keywords (i.e. skills) used in adverts have been standardised by Burning Glass and are less diverse as a result. In total, there are 11,200 unique keywords mentioned across the whole dataset. To reduce noise, we removed the 438 skills that occur fewer than 3 times in all adverts. The four skills that occur most frequently in adverts (communication skills, organisational skills, planning and customer service) are also excluded to prevent them from artificially increasing the level of skill similarity in

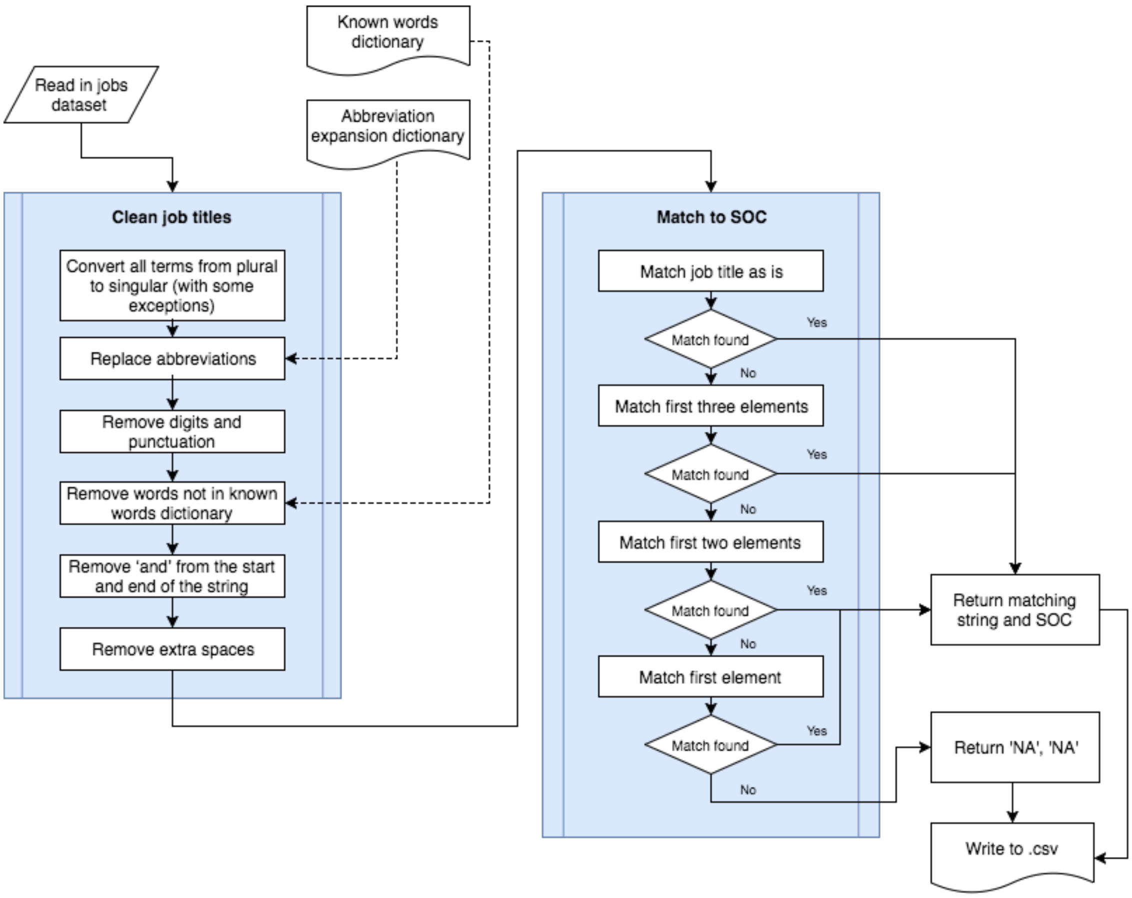

Figure 3: Job title cleaning and initial matching to ONS SOC

different jobs. As shown in Figure 4, pre-processing of skills involves collapsing the case and removing most punctuation characters, digits and extra spaces.

Classification layers

Skill category

At the *skill category* classification layer, jobs are assigned to skill categories based on cosine similarity between reference skill categories and skill requirements provided in the job advert. We chose to use the first layer of DISCO because of its extensive vocabulary of skill terms and phrases. However, the same methodology can be applied with a different skills taxonomy, such as a ONET or a new taxonomy developed in the future.

There is little overlap between the skill terms used in Burning Glass data and in the DISCO skills taxonomy. There are 11,200 skills in the Burning Glass data and over 5,900 skills across all levels of DISCO skills taxonomy listed in an online tool (DISCO II Portal). Only 400 skills (checked using exact spelling) existed in both Burning Glass and DISCO. For this reason, we use word embeddings, a Natural Language Processing technique, which captures semantic similarities of terms based on their distribution in large text corpora. While there are different word embeddings approaches to mapping words to their distributed representation, the resulting output is typically a numeric vector with length 300, where dimensions represent implicit semantic concepts (Mikolov, Sutskever, Chen, Corrado, and Dean, 2013). Word embeddings are more flexible than bag-of-words techniques, which represent documents as multisets of words ignoring word ordering and

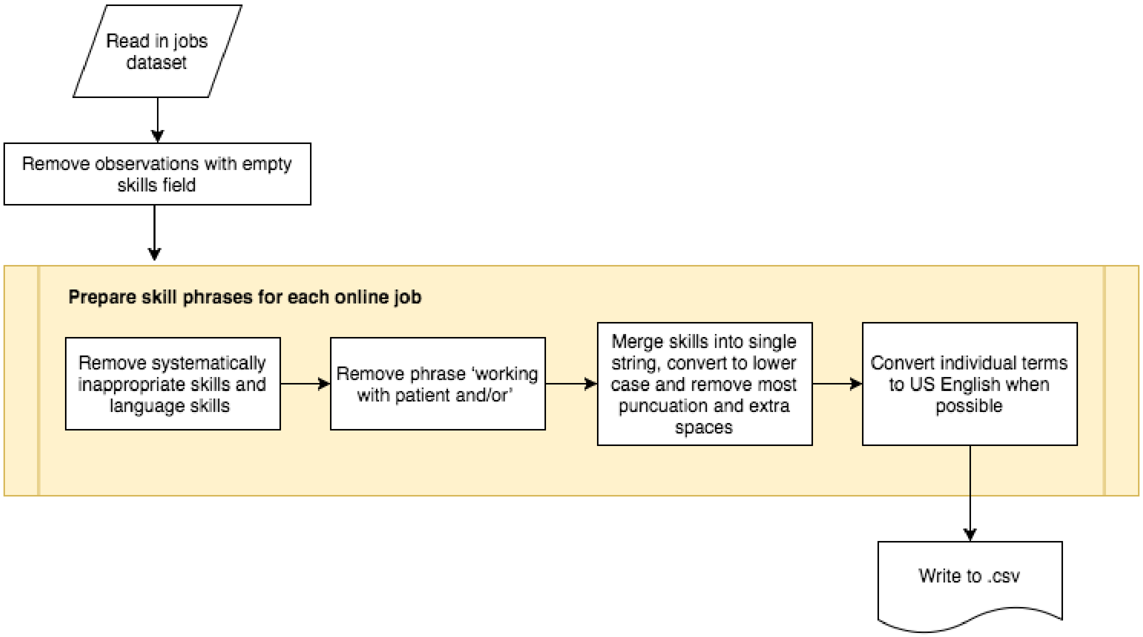

Figure 4: Pre-processing of skills

semantics (Jurafsky and Martin, 2008). Using word embeddings allows for comparing similarity of documents (i.e. job adverts, skill descriptions) that contain terms, which are semantically similar, but not exactly the same. There are publicly available pre-trained word embeddings models. We use a GloVe model, which contains a vocabulary of 2.2 million words and was trained using word to word co-occurrences in a Common Crawl corpus (Pennington, Socher, and Manning, 2014). The Common Crawl is an organisation that crawls the web and contains up to 1.81 billion webpages (as of 2015) in its archives. It would have been preferable to train our own word embeddings model on an occupation-specific corpus to extract more domain-specific semantic word representations. However, since a large investment of resources would be needed to curate such a corpus, we have decided to use a pre-trained model.

In order to assign job adverts to reference skill categories, we first convert unique skills in our dataset to vector representation using the GloVe pre-trained word embeddings model (Figure 5). We then generate 39 reference skill vectors from DISCO's 33 domain-specific and 6 non domain-specific categories (Figure 6). Several DISCO non domain-specific categories (*basic action verbs*, *driving licenses* and *materials, tools, products and software)* are very broad and are not included. We also re-organise the domain-specific DISCO categories merging some categories together: 3 manufacturing related categories are grouped into a single M*anufacturing and processing* category; L*ife*, *physical* and *social* sciences are also merged. Other categories are split: P*ersonal services* are divided into *Personal services*, *Food preparation*, *Leisure and sport* and *Travel and events*. We also use the second layer of *Business and administration* category instead of the first, because otherwise this category is very large and would contain over 52% of all job adverts.

Each DISCO skill category description contains multiple skill terms and in order to generate a single reference vector we average word embedding vectors of individual skill terms. This method is one of the common approaches for extending the word embeddings technique to multiple word use cases. Lau and Baldwin (2016) found that simple averaging of word embedding vectors performed reasonably well in comparison to other document-level embedding approaches.

Skills that fall under non domain-specific categories (artistic, personal, social and communication, managerial and organisational, basic computer skills and competences) are automatically dropped from the list of skill requirements mentioned in a given job advert. Only jobs with fewer than 20 domain-specific skills were included in further analysis, because we have previously found that job adverts that exceeded this threshold tended to represent several separate vacancies that have been incorrectly merged during the process of collection.

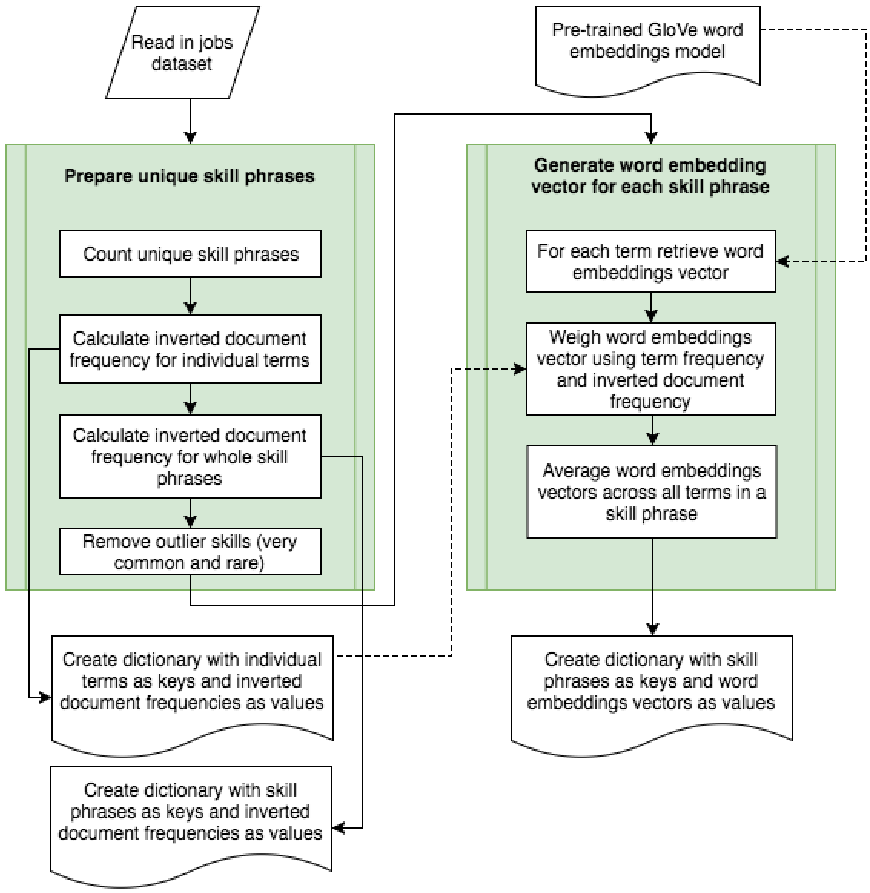

Figure 5: Generating word embedding vectors for skill phrases

For each job we measure the similarity between individual skills and each of the 33 DISCO domain-specific skill category vectors, using cosine similarity. We then calculate the element-wise mean of resulting vectors of cosine similarities and assign the job to the category with the highest average similarity (Figure 7).

Several corrections are made to re-assign certain job adverts from automatically assigned skill categories to more appropriate ones. For example, jobs requiring *Child protection* and *Information security* are automatically assigned to the *Security services* category due to the strong semantic links between the terms *protection* and *security*. These jobs are manually re-assigned to *Social services* and *Computing* respectively. We carry out corrections to a total of 1.89% of the sample.

Broad group

There are a number of skills that appear frequently in multiple skill categories, such as *Computer Aided Design (CAD)* and *Project management*. This indicates that some skill categories might be closely related to each other. To identify these relationships, we take samples of job adverts from every skill category assigned using the method outlined in the previous section. The size of the samples is determined as follows: if the skill category contains fewer than 100,000 adverts, all job adverts are used; in case of larger categories

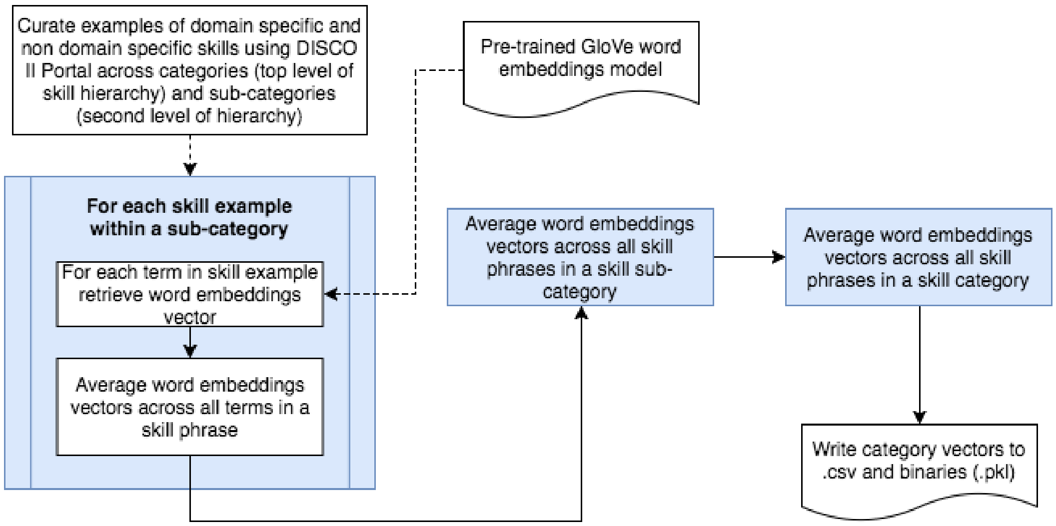

Figure 6: Generating word embedding vectors for reference DISCO skill categories

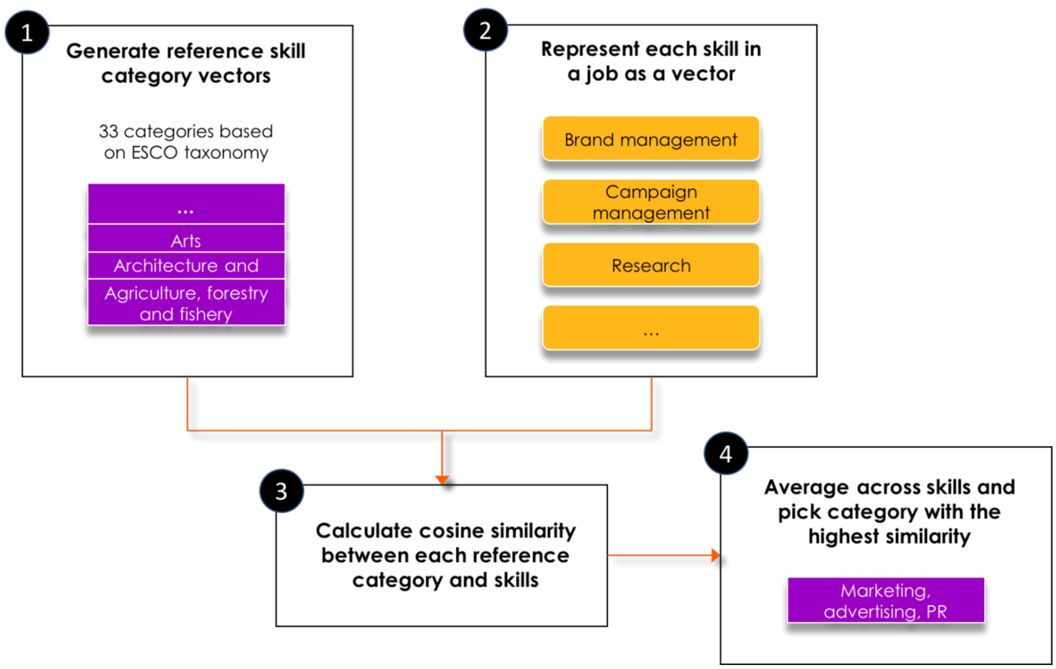

Figure 7: Steps to assign a job advert to a skill category

a random sample of 100,000 is selected. For each skill category, we calculate a representative skill vector by taking the element-wise mean of all skill vectors in the sample. We then hierarchically cluster the resulting skill vectors using Ward's method and cosine distance. The resulting dendrogram (Figure 8) demonstrates that there are broad groups amongst the skill categories. These insights are useful in assessing the potential for misclassifying jobs since it is more likely to involve similar skill categories. Grouping skill categories into fewer broad groups at the top of the hierarchy also makes it easier to work with the occupational classification structure.

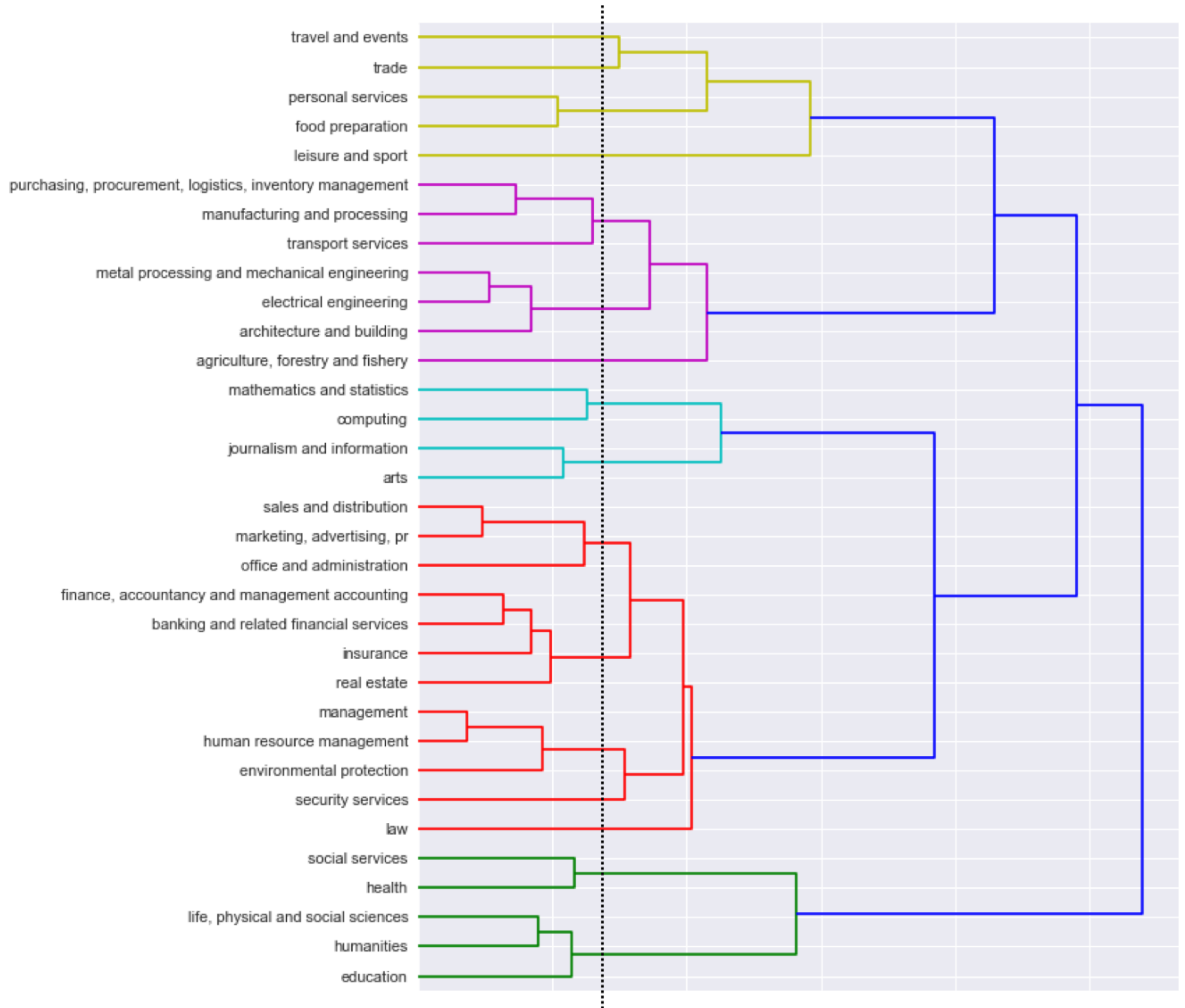

Figure 8: Dendrogram of skill categories grouped using Ward's method and cosine distance

Skill sub-category

Once jobs are allocated to skill categories, the next step is to identify more specific sub-categories for the largest skill categories (those with at least 5% of job adverts). For each advert, a single skill requirement vector is calculated as a weighted average of individual word embeddings skill vectors mentioned in a posting. We use term frequency - inverted document frequency (tf-idf) to weight skill vectors. Tf-idf measures the importance of terms in a corpus (Jurafsky and Martin, 2008). This statistic is often used to discount ubiquitous terms that occur in many documents. By using tf-idf to weight skill vectors we limit the contribution of very common skills to the overall skill vector. This prevents jobs appearing to be similar to each other simply because they mention one common skill.

The skill requirement vectors are clustered using the k-means algorithm. The optimal number of clusters, k, is determined based on the cluster stability. As demonstrated by Hennig (2007), when the right number of clusters is chosen, observations are likely to be consistently assigned to the same cluster over multiple runs of the algorithm. Conversely, if the inappropriate value of k is used, the membership of clusters is expected to vary between the runs. The stability of cluster membership is measured using the Jaccard coefficient; a cluster is considered to be stable if the Jaccard coefficient is over 0.75 for 100 iterations of algorithm with bootstrapping. We use this approach on random samples of 100,000 adverts and select the number of clusters that is associated with the highest mean value of the Jaccard coefficient.

As an alternative method for identifying skill sub-categories, we have explored the Latent Dirichlet Algorithm (LDA) for topic modelling. In principle, LDA might be a more appropriate technique for our use case, because it yields 'soft' groupings where a given job advert can be assigned to more than one topic and, therefore, can help better capture instances where a job combines two or more distinct skill sub-categories. However, this method appeared to produce less stable results, especially for diverse skill categories such as *Health*. It is likely that the short and sometimes sparse nature of the keywords in the Burning Glass dataset was a limiting factor and made LDA less suitable for unsupervised grouping of the job adverts.

Skill level

We use the k-means algorithm to partition each skill sub-category into clusters based on nominal offered salaries mentioned in job adverts that had been placed into those sub-categories. In our dataset 61% of adverts provide information on salary, this proportion varies from 44% to 70% across skill sub-categories. For the skill categories that are not partitioned into sub-categories, the salary clustering is performed on all the jobs in the skill category. The salary data are first log-transformed to address the large positive skew in the original values and then standardised prior to clustering. Applying the elbow method we found that the proportion of variance explained by cluster membership plateaus rapidly after 3 clusters, which means that adding more clusters will not substantially improve the clustering.

We also investigated Gaussian Mixture Model (GMM) as an alternative approach to grouping jobs based on salary. The advantage of the GMM is that it identifies clusters based on the density of salaries. The disadvantage is that the recommended number of clusters under this method is consistently over four, which might be impractical for an occupational classification.

Results

The resulting occupational classification comprises 16 broad groups, 33 skill categories, 50 skill sub-categories and 150 skill levels (Figure 9).

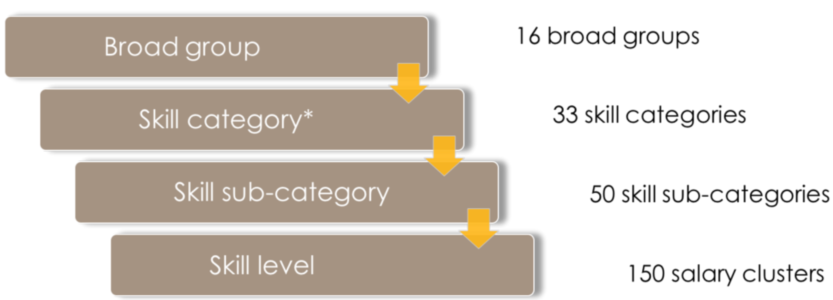

Figure 9: Skills based occupational classification

Broad groups

We use cosine distance to group skill categories hierarchically; the resulting hierarchy is shown in Figure 8. The dendrogram is dissected in such a way as to yield clusters of categories with low within cosine distance (i.e. skill categories that join relatively early in the dendrogram). This gives 16 broad groups, each comprising between one and four skill categories. Six of the skill categories are relatively distinct from the others, and so have their own broad groups. The membership of the broad groups is shown in Figure 10.

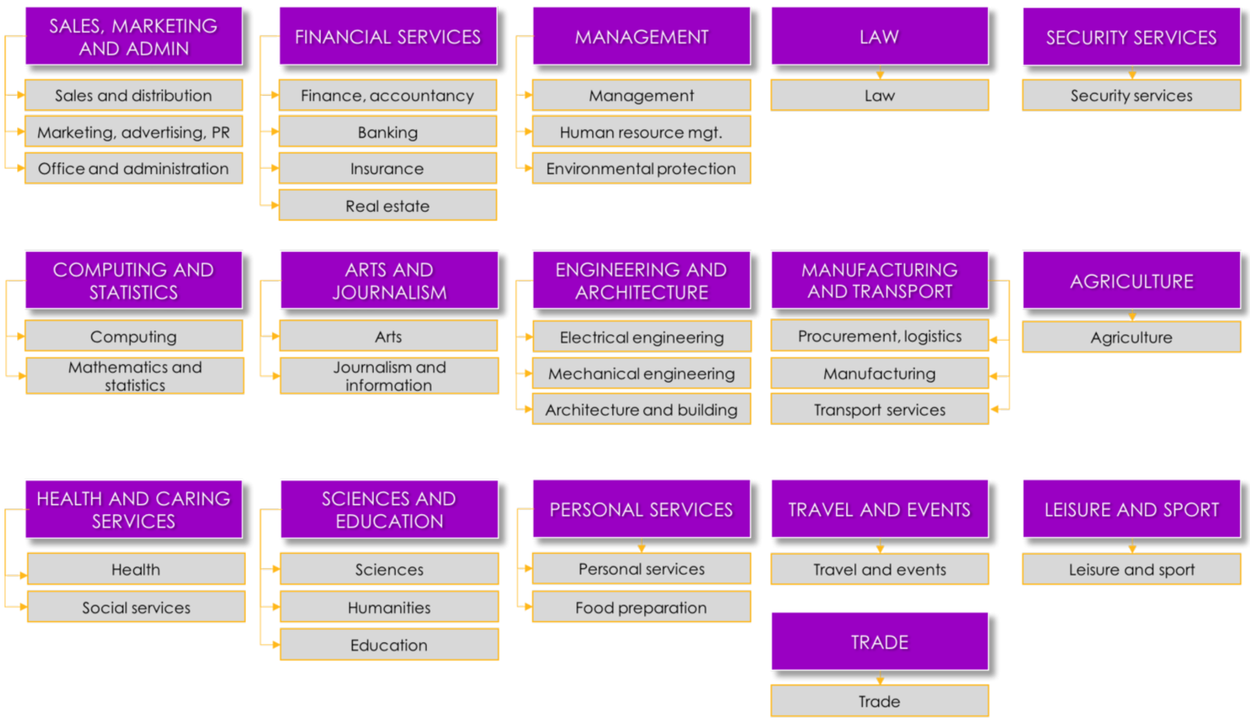

Figure 10: Composition of the broad groups

Skill categories

As described in the methodology section, job adverts are aligned with 33 DISCO based domain-specific skill categories. The skill categories with the largest proportion of job adverts are *Sales and distribution*, *Computing*, *Finance, accountancy*, and *Management* (Figure 11).

The Appendices provide more detail on each of the skill categories, including the proportion of job adverts assigned, the most important skills and the most common job titles.

Skill sub-categories

Eight of the skill-categories are divided into sub-categories. The identified sub-categories are shown in Figure 12. They are labelled by identifying the common themes amongst the most important skills for each cluster (i.e. words with the highest weight in the tf-idf matrix). For example, important skills for one of the sub-categories within *Office and administration* included *Calendar management*, *Typing*, *Secretarial skills*, and *Travel arrangements*. Based on these skills, we label the sub-category *'Secretarial'*.

Skill level

We divide each of the 50 skill sub-categories into 3 salary clusters based on the minimum salary mentioned in job adverts that have been placed into those sub-categories. Table 1 shows median Minimum Salary, Maximum Salary, Years of experience and Years of education for each cluster. The *Banking*, *Management* and all *Computing* sub-categories appear to contain the highest paid jobs. As shown by Figure 13, the lowest paid jobs are in in *Personal services*, *Agriculture*, *Food preparation* and *Office and administration* sub-categories.

The summary statistics in Table 1 were calculated for a limited subset of adverts that contain information on offered salary, education and experience requirements. Thus, for small skill categories, the summary statistics may not be representative.

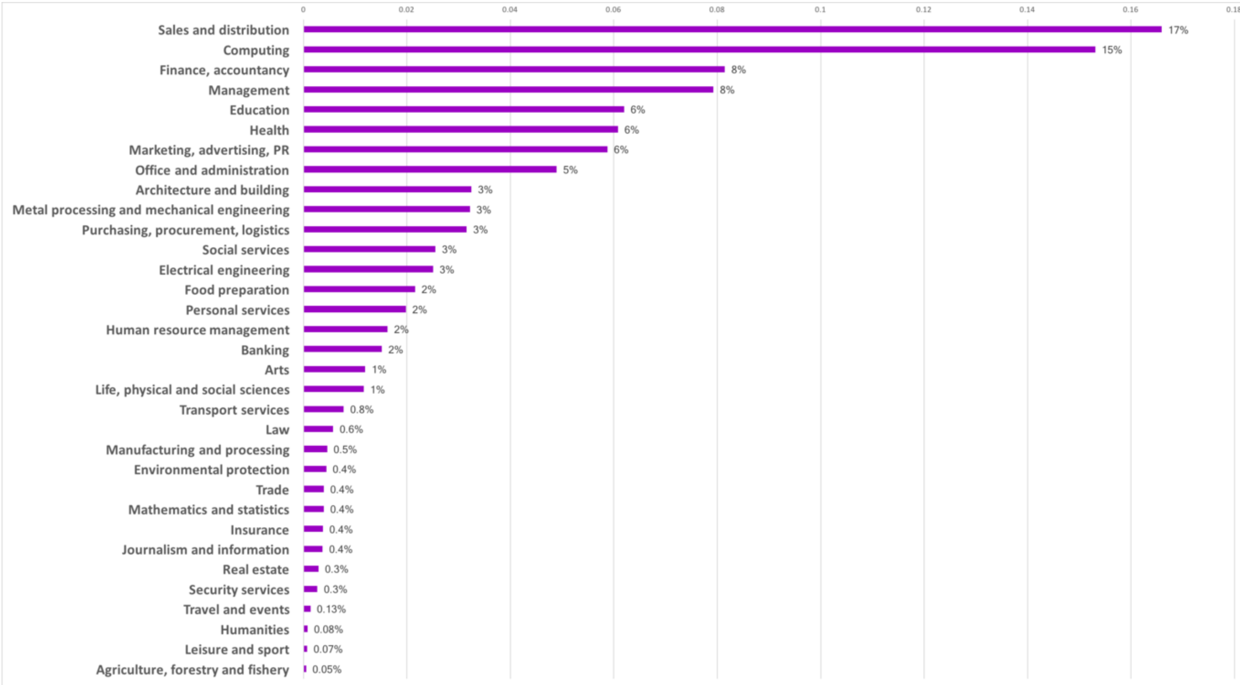

Figure 11: Proportion of job adverts in each skill category

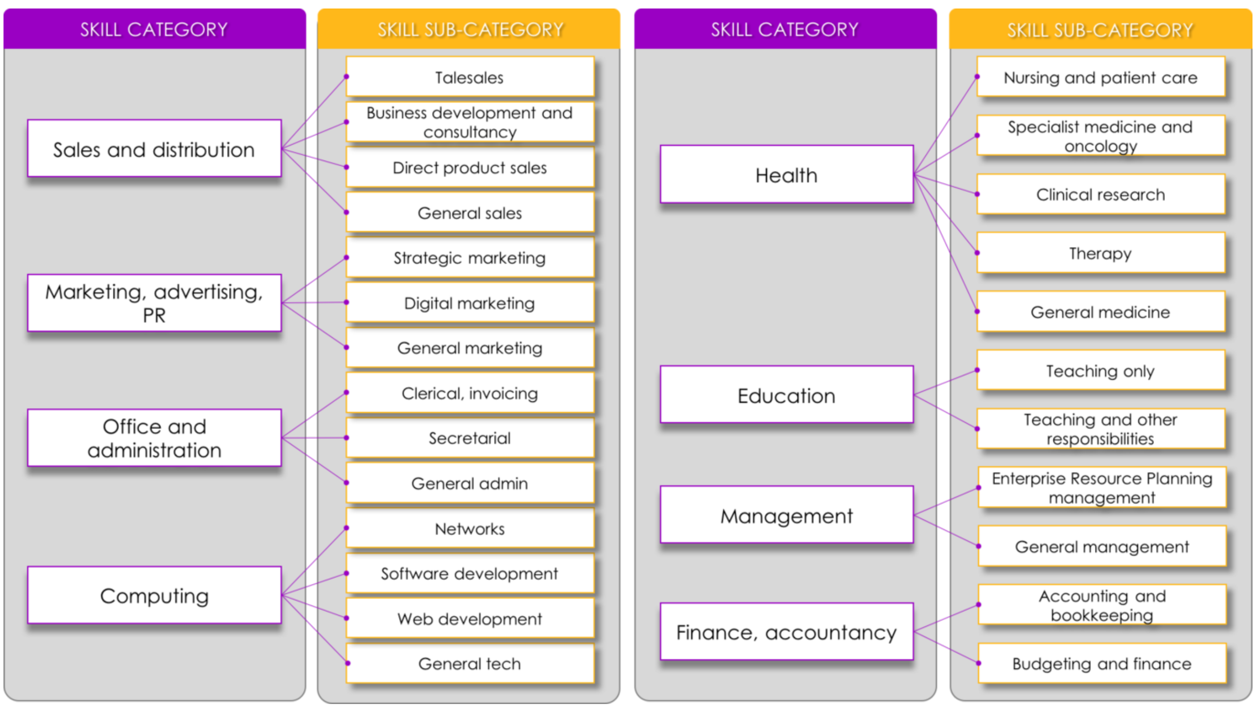

Figure 12: Skill sub-categories

Table 1: Overview of skill level groups

Skill sub-category	Skill level	Min Salary (median)	Max Salary (median)	Years of experience (median)	Years of education (median)	Proportion
Agriculture, forestry and fishery	Lower	£14,287	£14,976	1	11	45%
	Mid	£18,000	£19,859	2	12	36%
	Upper	£27,000	£30,000	3	14	19%
Arts	Lower	£18,000	£20,800	2	13	37%
	Mid	£28,000	£32,000	3	16	44%
	Upper	£45,000	£52,000	5	16	19%
Journalism and information	Lower	£18,000	£20,000	1.5	16	35%
	Mid	£29,000	£32,295	2	16	47%
	Upper	£46,911	£55,000	3	16	18%
Networks	Lower	£25,000	£30,000	2	16	33%
	Mid	£45,000	£50,000	3	16	43%
	Upper	£83,200	£96,200	3	16	24%
Software development	Lower	£26,000	£35,000	2	16	27%
	Mid	£45,000	£52,000	3	16	50%
	Upper	£80,000	£97,500	3	16	23%
Web development	Lower	£25,000	£35,000	2	16	42%
	Mid	£40,000	£45,000	3	16	43%
	Upper	£70,000	£78,000	3	16	15%
General tech	Lower	£20,000	£24,000	2	14	36%
	Mid	£38,000	£45,000	3	16	47%
	Upper	£78,000	£90,000	4	16	18%
Mathematics and statistics	Lower	£20,000	£24,908	2	16	35%
	Mid	£35,000	£40,000	2	16	43%
	Upper	£65,000	£75,000	4	16	21%
Metal processing and mechanical engineering	Lower	£18,200	£20,800	2	12	29%
	Mid	£26,000	£30,000	3	13	48%
	Upper	£40,000	£45,000	5	14	24%
Electrical engineering	Lower	£18,720	£20,800	2	13	25%
	Mid	£29,120	£32,000	3	13	52%
	Upper	£45,000	£50,000	5	16	23%
Architecture and building	Lower	£18,720	£20,800	2	12	27%
	Mid	£28,500	£31,616	3	13	50%
	Upper	£45,000	£52,000	5	16	23%
Accounting and bookkeeping	Lower	£18,000	£20,000	2	11	49%
	Mid	£26,000	£30,000	2	16	34%
	Upper	£46,000	£55,000	3	16	17%
Budgeting and finance	Lower	£20,198	£25,000	2	12	31%
	Mid	£35,000	£40,000	3	16	43%
	Upper	£60,000	£70,000	4	16	26%
Banking	Lower	£18,000	£21,000	2	12	36%
	Mid	£40,000	£47,559	2	16	38%
	Upper	£78,000	£90,000	3	16	26%
Insurance	Lower	£18,000	£20,000	1	11	43%
	Mid	£30,000	£35,000	2	12	39%
	Upper	£55,370	£65,000	4	16	18%
Real estate	Lower	£18,000	£20,175	1	12	25%
	Mid	£31,000	£35,000	2	14	54%
	Upper	£55,000	£65,000	3	16	22%
Nursing and patient care	Lower	£15,610	£17,978	1	11	24%
	Mid	£26,519	£30,000	1	14	54%
	Upper	£40,000	£47,559	1	16	21%

Specialist medicine and oncology	Lower	£20,800	£23,173	1	13	24%
	Mid	£30,302	£40,090	2	16	52%
	Upper	£75,249	£100,000	1	16	25%
Clinical research	Lower	£18,000	£20,000	2	12	34%
	Mid	£30,000	£35,000	2	16	46%
	Upper	£45,760	£55,000	3	16	19%
Therapy	Lower	£18,000	£20,030	1	12	34%
	Mid	£28,000	£34,530	2	16	47%
	Upper	£45,000	£53,367	2	16	19%
General medicine	Lower	£18,720	£22,016	1	12	31%
	Mid	£28,471	£34,530	1	16	50%
	Upper	£45,000	£55,000	1	16	19%
Social services	Lower	£15,453	£17,306	0.5	11	37%
	Mid	£28,000	£32,000	2	14	36%
	Upper	£52,000	£58,240	2	16	26%
Law	Lower	£18,000	£21,402	1	12	32%
	Mid	£30,000	£40,000	2	13	45%
	Upper	£60,000	£70,000	3	16	23%
Leisure and sport	Lower	£14,560	£16,000	2	11	34%
	Mid	£20,800	£24,000	2	13.5	38%
	Upper	£34,718	£41,600	2	12	29%
Enterprise Resource Planning management	Lower	£30,000	£35,000	2	16	27%
	Mid	£50,000	£55,000	4	16	47%
	Upper	£91,000	£104,000	4	16	25%
General management	Lower	£25,000	£29,000	2	16	34%
	Mid	£40,000	£50,000	3	16	42%
	Upper	£78,000	£90,000	5	16	24%
Human resource management	Lower	£18,000	£20,000	2	12	37%
	Mid	£28,180	£32,000	2	14	42%
	Upper	£45,518	£52,000	3	16	22%
Environmental protection	Lower	£19,333	£22,000	2	16	27%
	Mid	£30,000	£35,000	3	16	47%
	Upper	£46,800	£55,000	5	16	25%
Purchasing, procurement, logistics	Lower	£16,640	£18,000	2	11	37%
	Mid	£28,000	£31,342	3	13	40%
	Upper	£50,000	£55,000	5	16	23%
Manufacturing and processing	Lower	£18,000	£20,000	2	12	30%
	Mid	£30,000	£35,000	3	13	52%
	Upper	£52,000	£60,000	5	16	18%
Transport services	Lower	£15,600	£17,000	2	11	40%
	Mid	£21,500	£24,960	2	12	44%
	Upper	£30,160	£35,000	2	12	16%
Personal services	Lower	£13,208	£13,520	1	11	39%
	Mid	£15,205	£15,808	1	11	44%
	Upper	£19,500	£20,800	2	11	17%
Food preparation	Lower	£15,000	£16,000	2	11	37%
	Mid	£20,000	£22,000	2	11	39%
	Upper	£28,000	£30,000	2	12	24%
Telesales	Lower	£15,000	£17,000	1	11	41%
	Mid	£20,000	£23,000	1	16	43%
	Upper	£35,000	£40,000	2	16	15%
Business development	Lower	£18,000	£20,000	1	12	39%
	Mid	£30,000	£35,000	2	16	42%
	Upper	£55,000	£65,000	4	16	19%
Direct product sales	Lower	£15,000	£17,000	1	11	41%
	Mid	£25,000	£28,000	2	16	38%
	Upper	£40,000	£48,000	3	16	21%
General sales	Lower	£18,000	£20,000	2	11	43%
	Mid	£30,000	£36,000	3	16	39%

	Upper	£60,000	£70,000	3	16	17%
Strategic marketing	Lower	£18,000	£20,000	1	16	32%
	Mid	£30,000	£35,000	3	16	46%
	Upper	£52,000	£60,000	5	16	22%
Digital marketing	Lower	£17,213	£19,760	1	16	41%
	Mid	£27,000	£30,000	2	16	41%
	Upper	£45,000	£55,000	3	16	18%
General marketing	Lower	£18,000	£20,800	1	16	35%
	Mid	£29,000	£32,000	2	16	42%
	Upper	£46,132	£55,000	4	16	23%
Clerical, invoicing	Lower	£15,000	£16,000	1	11	40%
	Mid	£19,000	£21,000	2	12	39%
	Upper	£27,798	£31,200	2	12	20%
Secretarial	Lower	£15,000	£16,307	1	11	37%
	Mid	£18,000	£20,000	2	11	44%
	Upper	£25,000	£28,000	2	12	19%
General admin	Lower	£15,600	£17,000	1	11	48%
	Mid	£21,519	£24,500	2	12	35%
	Upper	£33,148	£37,700	3	15	17%
Teaching only	Lower	£15,600	£16,900	2	16	21%
	Mid	£23,400	£34,887	2	12	48%
	Upper	£35,360	£42,900	2	16	31%
Teaching and other responsibilities	Lower	£15,600	£17,372	1	12	29%
	Mid	£24,012	£30,568	2	12	47%
	Upper	£36,661	£42,900	3	16	25%
Life, physical and social sciences	Lower	£18,652	£21,000	2	16	28%
	Mid	£30,000	£36,298	2	16	53%
	Upper	£48,000	£56,160	2	16	19%
Humanities	Lower	£22,000	£29,247	1	13	33%
	Mid	£33,280	£42,000	2	16	53%
	Upper	£53,040	£60,000	5	16	15%
Security services	Lower	£15,288	£15,704	5	11	57%
	Mid	£26,007	£30,000	3	13	31%
	Upper	£49,920	£56,000	3	16	12%
Trade	Lower	£16,000	£17,000	1	11	40%
	Mid	£22,360	£25,000	2	12	40%
	Upper	£33,000	£36,000	2	16	21%
Travel and events	Lower	£14,976	£15,600	1	11	38%
	Mid	£21,000	£24,000	2	12	39%
	Upper	£33,000	£35,360	2	16	23%

Discussion

Validating an occupational taxonomy is a challenging task as there is no established definition of a 'true' taxonomy. In due course we will be publishing an applied analysis paper where we will compare and contrast the composition of the UK online labour market measured by the existing SOC and our occupational classification based on skills. Based on our results so far, the proposed methodology appears broadly reasonable. The skills and job titles in the largest skill categories are consistent with our understanding of these jobs (Tables 2, 3, 4). While we do observe instances of particular skills appearing in unexpected occupations (such as *Management* in *Humanities*), this is most likely due to the small number of job adverts in these occupations. Burning Glass assign a SOC code to each job advert. This allows us to examine the most common SOC codes in each of our skill categories. Doing this shows that the SOC codes are largely aligned with underlying *skill specialisation* (i.e. the most frequent SOC codes for marketing skill category correspond to *Marketing associate professionals* and *Marketing and sales directors*) (Table 5). One limitation is that the SOC codes automatically assigned by Burning Glass might be inaccurate in some cases.

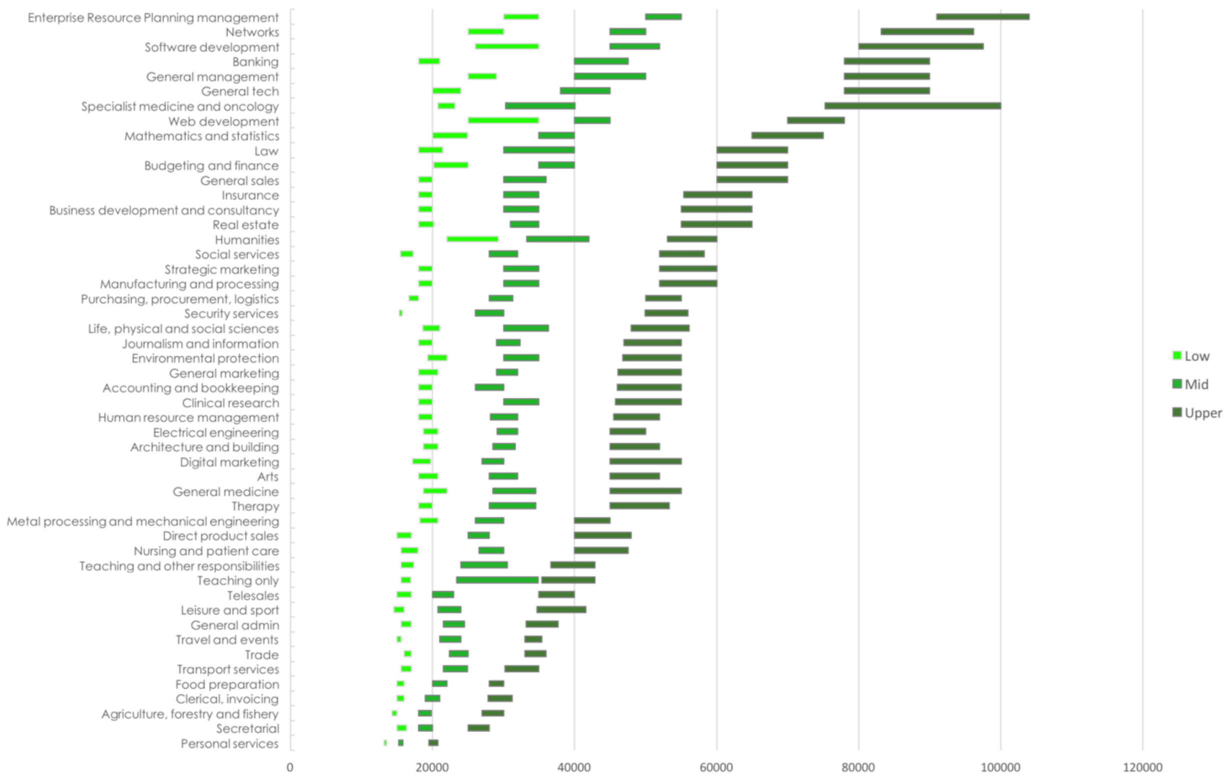

Figure 13: Median Minimum and Maximum salary for low, mid and upper skill level groups

While the resulting occupational classification of the UK online job adverts is informative in its own right, we believe that the major contribution of this work is the underlying methodology for grouping jobs based on skills. The methodology makes use of both semi-supervised and unsupervised learning methods. Although we use the DISCO skills taxonomy to inform the semi-supervised skill category layer, the methodology can be easily adapted to work with a different taxonomy. Regardless of the taxonomy, the selection of skill terms and phrases used to define reference categories will have an immediate impact on classification outcomes. This is demonstrated by the *Environmental protection* category, which, as currently described in DISCO, to a large extent focuses on consulting and management aspects of environmental protection. As a result, this category resides in the same broad group as *Management*. In a forthcoming paper, we intend to use Burning Glass data to develop a skills taxonomy based on the network analysis of skill co-occurrence in job adverts. We expect that if we were to use this 'organic' skills taxonomy to inform the skill category layer (in place of DISCO) it might help to further align the occupational classification with employer demands.

Apart from the skill category layer, the other layers are shaped by unsupervised learning techniques. However, we do impose certain thresholds to guide these techniques. These thresholds need to be validated by occupational classification experts and they will likely change to better meet the needs of practitioners. We currently split skill categories if they contain at least 5% of the job adverts in our dataset. There might be a more appropriate way to determine how to split or merge categories. For instance, we might take into account their share of UK employment, rather than their share of UK online adverts. Or, perhaps, increased granularity of skillsets should be preferred since it might allow practitioners to spot new emerging occupations. Similarly, there are alternative approaches to identifying appropriate skill level groups: using k-means algorithm allows us to partition the data on salary in such a way as to minimise the distances from observations to the centre of each cluster. An alternative approach that is based on identifying local peaks in salary probability density function might be more practical and intuitive.

A further strength of the proposed methodology is that it can be updated in response to new job adverts. The results of this paper are based on five years of data. However, the method could be re-run on an ongoing basis with the aim of identifying trends and changes over time. This real-time aspect of the approach could be of use to occupational classification practitioners. The proposed methodology, in particular using reference categories with word embeddings, could also be used on an ad-hoc basis to study a single occupation or skillset in more detail.

There are a number of ways to further refine the methodology in future work. One approach would be to train a word embeddings model that would be specific to the labour market. Word embeddings play an instrumental role in creating the methodology. A tailored word embeddings model would allow us to assign skills to skill categories with greater confidence. Currently, skills like *Scrum* are driven towards the *Leisure and sport* skill category, because in the broad corpora this term is used predominantly in relation to rugby. However, in an occupational context, the term is associated with agile software development techniques.

Conclusion

In this paper we propose a methodology to group occupations on the basis of skill requirements contained in 37 million UK job adverts. The resulting occupational classification captures both the *skill specialisations* and *skill levels* of occupations. In its current form, the methodology comprises four hierarchical layers. At the first three layers, we use skills from the adverts to place jobs into groups that require similar domain-specific skills. By identifying these distinct skillsets, we lay the groundwork for quantifying skill demands and analysing the composition of the UK workforce by skill type. The fourth layer of the hierarchy reflects a job's skill level, on the basis of the salary offered. Integrating a *skill level* dimension into the classification provides a pathway for the analysis of individuals' career progression within a given domain-specific skillset.

We believe that this work contributes to the occupational classification field in a number of ways. First, we offer a data-driven approach for dynamically capturing skills, competencies and knowledge required by employers. A vast collection of job adverts is used to develop the methodology, which means that we can gauge the needs of employers across the UK with high resolution and accuracy. The approach is cost effective, because it requires little manual input. The methodology can also be easily extended to work with any skills taxonomy and thus offers policymakers, educators and researchers the flexibility to choose a taxonomy that is most closely aligned with their objectives. Finally, the proposed approach can be applied to analyse skill requirements across all occupations on an on-going basis or to focus on a skillset/occupation of interest. Apart from the choice of the skills taxonomy, the methodology is algorithmic in nature, which means that the methodology can be used to automatically code large volumes of job adverts to occupations.

Further research will help to validate the methodology and increase its relevance to occupational classification practitioners. There is also scope to refine the analytical methods used to develop the methodology by training an occupation-specific word embeddings model and to improve the accuracy of job assignment to reference categories. The results of our work will be released publicly and shared with labour market researchers, with the aim of showing how online job advert data can be used to improve our understanding of labour markets.

Appendices

Table 2: Overview of skill categories

Broad group	Skill category	Number of adverts	Proportion of adverts	Median minimum salary

Agriculture	Agriculture, forestry and fishery	15,023	0.1%	£16,328
Arts and journalism	Arts	358,080	1.2%	£25,000
Arts and journalism	Journalism and information	110,799	0.4%	£26,000
Computing and maths	Computing	4,590,369	15.3%	£36,080
Computing and maths	Mathematics and statistics	116,826	0.4%	£30,680
Engineering and architecture	Architecture and building	974,322	3.2%	£28,000
Engineering and architecture	Electrical engineering	751,565	2.5%	£29,120
Engineering and architecture	Metal processing and mechanical engineering	964,675	3.2%	£25,400
Financial services	Banking	454,425	1.5%	£35,000
Financial services	Finance, accountancy	2,443,013	8.1%	£30,000
Financial services	Insurance	112,756	0.4%	£26,000
Financial services	Real estate	86,377	0.3%	£30,000
Health and care	Health	1,822,726	6.1%	£27,300
Health and care	Social services	765,404	2.6%	£26,000
Law	Law	172,079	0.6%	£30,000
Leisure and sport	Leisure and sport	22,192	0.1%	£20,800
Management	Environmental protection	132,044	0.4%	£30,000
Management	Human resource management	487,271	1.6%	£26,000
Management	Management	2,375,986	7.9%	£40,000
Manufacturing and transport	Manufacturing and processing	136,904	0.5%	£30,000
Manufacturing and transport	Purchasing, procurement, logistics	946,332	3.2%	£25,000
Manufacturing and transport	Transport services	233,389	0.8%	£20,000
Personal services	Food preparation	646,811	2.2%	£18,720
Personal services	Personal services	593,340	2.0%	£14,643
Sales, marketing and admin	Marketing, advertising, PR	1,761,227	5.9%	£26,000
Sales, marketing and admin	Office and administration	1,467,823	4.9%	£18,000
Sales, marketing and admin	Sales and distribution	4,974,908	16.6%	£24,000
Sciences and education	Education	1,857,984	6.2%	£23,400
Sciences and education	Humanities	24,674	0.1%	£31,894
Sciences and education	Life, physical and social sciences	348,554	1.2%	£29,249
Security services	Security services	80,538	0.3%	£18,720
Trade	Trade	117,146	0.4%	£20,000
Travel and events	Travel and events	40,419	0.1%	£20,000

Table 3: Top twenty most important skills in each skill category (measured by tf-idf)

Broad group	Skill category	Top 20 skills with highest tf-idf
Agriculture	Agriculture, forestry and fishery	grass cutting, animal care, agricultural industry experience, farm management, lotus domino, garden industry experience, animal husbandry, herbicides, agricultural tractors, lawn mowing, irrigation, fertilizers, agronomy, machinery, farm machinery, wildlife conservation, lawnmowers, solar farm, land planning, tree felling
Engineering and architecture	Architecture and building	repair, construction industry knowledge, plumbing, carpentry, civil engineering, commercial construction, inspection, construction management, revit, project management, building industry experience, home building, team building, computer aided draughting design cad , electrical work, contract management, demolition, roofing, hvac, painting
Arts and journalism	Arts	painting, graphic design, music, adobe photoshop, editing, adobe indesign, photography, digital design, adobe acrobat, video production, image processing, computer aided draughting design cad , technical drawings, hand tools, adobe illustrator, art direction, brand design, website production, typesetting, video editing
Financial services	Banking	financial industry experience, cash handling, portfolio management, asset management, mergers and acquisitions, financial services industy experience, derivatives, corporate finance, capital markets, business management, investment management, acquisitions, investment banking, equities, credit risk, contract management, account closing, financial management, mortgage advice, securities trading
Computing and maths	Computing	sql, microsoft c#, java, .net programming, sql server, asp, linux, technical support, software engineering, web site development, hypertext preprocessor php , software development, oracle, troubleshooting, c++, information technology industry experience, jquery, project management, extensible markup language xml , unix
Sciences and education	Education	teaching, teaching english, tutoring, teaching mathematics, lesson planning, teaching science, management, lecturer, graduate teaching, teaching geography, teaching information and communication technology, condition learning disabilities, teaching pe, teaching history, psychology, research, workshops, condition autism, music, teaching art
Engineering and architecture	Electrical engineering	electrical engineering, electrical work, computer numerical control cnc , computer aided draughting design cad , wiring, telecommunications, repair, systems engineering, electrical design, electronic design, scanners, inspection, cabling, engineering industry background, siemens nixdorf hardware, calibration, electrical systems, printers, analogue design, test equipment

Management	Environmental protection	environmental remediation, environmental management, sustainability, renewable energy, environmental consultancy, environmental engineering, environmental health and safety, environmental protection, environmental policy, project management, environmental science, energy conservation, workplace health and safety, civil engineering, carbon reduction, iso 14001 standards, quality assurance and control, pollution control, energy efficiency, waste reduction
Financial services	Finance, accountancy	accountancy, budgeting, invoicing, financial accountancy, contract accountancy, budget management, account reconciliation, budget forecasting, account auditing, contract management, forecasting, payroll processing, balance sheet, bank reconciliation, financial reporting, bookkeeping, accounts payable and receivable, sap, account analysis, financial analysis
Personal services	Food preparation	cooking, food safety, food service industry background, restaurant management, restaurant industry experience, dining experience, meal preparation, stock control, beverage industry knowledge, bartending, hospitality industry experience, meal serving, management, restaurant experience, caregiving, cleaning, food service, cash handling, budgeting, planning menus
Health and care	Health	mental health, patient care, surgery, condition dementia, occupational health and safety, occupational therapy, nursing home, dentistry, therapy, pediatrics, medical industry background, healthcare industry experience, care planning, primary care, research, immunisations, oncology, pharmacist, physiotherapy, medication administration
Management	Human resource management	it recruiting, staff coordination, contract administration, facility supervision, employee training, faculty training, employee relations, training programmes, engineering consultation, contract preparation, administration management, facility management, staff management, training materials, itil, staff development, team management, administrative support, technical training, technical recruiting
Sciences and education	Humanities	sociology, teaching, psychology, lecturer, archaeology, teaching history, research, music, european history, poetry, art history, teaching speakers of other languages, management, prose, architectural history, journalism, anthropology, teaching english, teaching geography, fine art
Financial services	Insurance	insurance underwriting, insurance industry experience, claims adjustments, mortgage advice, home health, risk management, claims service, claims knowledge, benefits management, auto repair, cemap, insurance sales, insurance knowledge, contract management, repair, property claims, home care, home management, customer contact, commercial insurance sales

Arts and journalism	Journalism and information	report writing, research, journalism, editing, copy writing, proofreading, research reports, technical writing editing, newspaper, project management, microsoft publisher, grant writing, mailing, questionnaires, social media, online research, data collection, broadcast, blogging, content management
Law	Law	litigation, commercial litigation, case management, civil litigation, legal support, arbitration, legal compliance, criminal justice, claims knowledge, employment rights, tupe, regulatory affairs, legal documentation, intellectual property, territory management, prosecution, legal research, law enforcement or criminal justice experience, business development, claims adjustments
Leisure and sport	Leisure and sport	pilates, yoga, zumba, air travel industry background, music, travel arrangements, bartending, drills, business consultancy, spa industry knowledge, hospitality industry experience, football, soccer, exercise programmes, sports massage, instruction, aerobics, tennis, teaching, gymnastics
Sciences and education	Life, physical and social sciences	research, biology, chemistry, physics, psychology, teaching, teaching biology, lecturer, teaching science, molecular biology, teaching physics, biochemistry, physiology, clinical psychology, psychiatry, economics, geology, hematology, experiments, pathology
Management	Management	project management, business development, business management, business analysis, project planning and development skills, contract management, operations management, research, procurement, business consultancy, organisational development, business process, management, strategic management, budgeting, change management, quality assurance and control, budget management, prince2, business planning
Manufacturing and transport	Manufacturing and processing	sap, packaging, lean methods, lean manufacturing, manufacturing processes, good manufacturing practises gmp , manufacturing industry experience, quality assurance and control, machinery, manufacturing resource planning mrp , purchasing, food service industry background, procurement, food safety, grinders, inspection, production management, product sales, lean processes, supply chain management
Sales, marketing and admin	Marketing, advertising, PR	marketing, social media, marketing sales, advertising copywriting, campaign management, fundraising, marketing management, marketing communications, brand management, market strategy, strategic marketing, research, brand marketing, product marketing, merchandising, market research, online marketing, digital marketing, e-commerce, brand experience
Computing and maths	Mathematics and statistics	data analysis, spreadsheets, sas, statistics, research, physics, economics, forecasting, spss, mathematical modelling, matlab, simulation, calculation, surveys, trend analysis, c++, econometrics, geographic information system gis , sql, r

Engineering and architecture	Metal processing and mechanical engineering	mechanical engineering, repair, welding, machinery, automotive repair, machining, mechanical design, engineering industry background, computer numerical control cnc, automotive industry experience, computer aided draughting design cad , materials design, inspection, hydraulics, mig and tig welding, electrical engineering, engineering management, lathes, machine operation, injection moulding
Sales, marketing and admin	Office and administration	office administration, typing, office management, mailing, administrative support, secretarial skills, administrative functions, file management, administration management, calendar management, telephone skills, general office duties, data entry, contract administration, order and invoice processing, invoicing, spreadsheets, travel arrangements, note taking, office skills
Personal services	Personal services	cleaning, cooking, laundry, housekeeping, caregiving, ironing, toileting, equipment cleaning, food safety, meal preparation, cash handling, home management, work area maintenance, bed making and linen changes, facility supervision, home care, stock control, inspection, babysitting, care planning
Manufacturing and transport	Purchasing, procurement, logistics	forklift operation, procurement, warehouse management, logistics, purchasing, stock control, contract management, supply chain management, inspection, transportation logistics, machinery, operations management, repair, packaging, supplier management, supply chain, quality assurance and control, sorting, supply chain knowledge, facility supervision
Financial services	Real estate	property management, real estate experience, property management systems, portfolio management, estate planning, contract management, acquisitions, real estate planning, business development, general practise, land planning, land management, asset management, home building, tax planning, repair, business management, management, mortgage advice, home management
Sales, marketing and admin	Sales and distribution	sales, customer contact, business management, product sales, product sale and delivery, sales recruiting, sales management, business development, telesales, marketing sales, contract management, sales goals, retail setting, account management, store management, prospective clients, inside sales, product knowledge, sales engineering, retail sales
Security services	Security services	security industry knowledge, surveillance, cctv monitoring, report writing, inspection, emergency services, security experience, asset protection, access and or egress control, report maintenance, security patrol, loss prevention, security industry authority, workplace health and safety, surveillance system monitoring, repair, quality assurance and control, systems monitoring, prevention of criminal activity, traffic management

Health and care	Social services	social work, caregiving, care planning, child protection, mental health, condition learning disabilities, social services, home management, nursing home, learning disability, elder care, senior care, condition physical disability, condition dementia, community development, home care, condition autism, supportive care, companionship, record keeping
Trade	Trade	store management, retail management, retail setting, shipping through ups, cross sell, stock control, management, brand management, shipping, retail industry background, cash handling, buying experience, trade shows, trading floor, market trend, food safety, retail channel, trade marketing, merchandise labelling, purchasing
Manufacturing and transport	Transport services	transportation logistics, heavy large goods vehicle driving, haulage, forklift operation, lift trucks, delivery driving, transportation planning, traffic management, vehicle maintenance, freight forwarding, transporting, bus driving, crane operation, commercial driving, delivery unload and breakdown, dump truck driving, transport planning, transportation industry knowledge, repair, motor vehicle operation
Travel and events	Travel and events	event management, event planning, hospitality industry experience, hotel industry experience, restaurant management, dining experience, fundraising, budget management, travel arrangements, calendar management, management, contract management, cash handling, restaurant industry experience, team building, secretarial skills, staff management, work area maintenance, staff coordination, guest services

Table 4: Top 20 most frequent job titles for each skill category

Broad group	Skill category	Top 20 job titles
Agriculture	Agriculture, forestry and fishery	farm manager, assistant farm manager, gardener, animal technician, dog walker pet carer, agronomist, grounds maintenance operative, horticulture apprentice, lawn care operative, grounds maintenance operator, agriculture apprentice, landscape operative, relief farm manager, poultry production apprentice, trainee animal technician, apprentice horticulture, farm worker, grower, countryside ranger, animal care technician
Engineering and architecture	Architecture and building	project manager, structural engineer, electrician, site manager, carpenter, quantity surveyor, plumber, engineer, estimator, cad technician, construction manager, civil engineer, project engineer, mechanical engineer, electrical engineer, site engineer, mechanical design engineer, contract manager, design engineer, structural design engineer

Arts and journalism	Arts	graphic designer, designer, digital designer, user experience designer, artworker, creative artworker, interior designer, web designer, motion graphic designer, design engineer, creative designer, visual designer, landscape architect, mechanical design engineer, cad technician, d designer, packaging designer, art director, technical author, editor
Financial services	Banking	business analyst, mortgage adviser, project manager, corporate solicitor, analyst, credit controller, account manager, accounts assistant, financial adviser, manager, investment analyst, business development manager, credit risk analyst, finance manager, credit analyst, corporate lawyer, property manager, paraplanner, independent financial adviser, risk manager
Computing and maths	Computing	developer, web developer, java developer, software engineer, php developer, software developer, .net developer, c# developer, front end developer, engineer, network engineer, project manager, test analyst, systems engineer, data analyst, business analyst, consultant, solution architect, embedded software engineer, infrastructure engineer
Sciences and education	Education	teacher, english teacher, science teacher, teaching assistant, year teacher, music teacher, lecturer, tutor, geography teacher, school teacher, primary teacher, sen teacher, chemistry teacher, teacher of english, history teacher, teacher of, pe teacher, sen teaching assistant, teacher of science, teacher of music
Engineering and architecture	Electrical engineering	electrical engineer, electrical design engineer, electronics engineer, engineer, electrician, design engineer, electronics design engineer, maintenance engineer, mechanical design engineer, field service engineer, systems engineer, control systems engineer, electrical maintenance engineer, hardware engineer, electronic design engineer, control engineer, maintenance electrician, service engineer, quality inspector, quality engineer
Management	Environmental protection	engineer, environmental consultant, environmental engineer, project manager, sustainability consultant, geotechnical engineer, consultant, mechanical engineer, manager, environmental adviser, process engineer, ecologist, energy manager, acoustic consultant, energy consultant, electrical engineer, project engineer, adviser, quality engineer, environmental manager
Financial services	Finance, accountancy	management accountant, accounts assistant, finance manager, accountant, financial accountant, financial controller, quantity surveyor, assistant accountant, payroll administrator, purchase ledger clerk, finance assistant, finance analyst, bookkeeper, credit controller, project manager, financial analyst, assistant management accountant, administrator, account manager, business analyst

Personal services	Food preparation	chef, head chef, chef de partie, commis chef, apprentice chef, restaurant manager, chef manager, cook, catering assistant, cook chef, waiting staff, bar staff, support worker, assistant restaurant manager, chef cook, kitchen assistant, food service assistant, cleaner, kitchen porter, care assistant
Health and care	Health	staff nurse, registered nurse, nurse, occupational therapist, registered general nurse, care assistant, support worker, physiotherapist, healthcare assistant, dental nurse, consultant, practice nurse, pharmacy technician, occupational health adviser, associate dentist, dental associate, radiographer, theatre practitioner, pharmacist, clinical psychologist
Management	Human resource management	human resource adviser, human resource manager, human resource administrator, human resource officer, administrator, human resource assistant, chef, recruitment consultant, manager, assistant manager, trainer, project manager, deputy manager, team leader, training manager, it trainer, engineer, building surveyor, general manager, quantity surveyor
Sciences and education	Humanities	lecturer, psychology teacher, teacher of psychology, teacher, psychology and teacher, lecturer history, teacher of and psychology, teacher of, lecturer psychology, history teacher, level lecturer, lecturer modern european history, lecturer creative, history and teacher, lecturer ancient history, head of psychology, lecturer modern history, lecturer lecturer, psychology, lecturer early modern history
Financial services	Insurance	claims handler, mortgage adviser, commercial account handler, account handler, underwriter, motor claims handler, mortgage broker, claims adjuster, customer service adviser, project manager, insurance sales executive, commercial account executive, business analyst, commercial claims handler, commercial underwriter, risk manager, claims manager, home manager, commercial insurance broker, personal injury claims handler
Arts and journalism	Journalism and information	paraplanner, editor, copywriter, technical author, researcher, editorial assistant, research assistant, medical writer, project manager, bid writer, technical writer, administrator, content editor, research associate, research fellow, paralegal, reporter, communications officer, broadcast journalist, research executive
Law	Law	commercial litigation solicitor, litigation solicitor, paralegal, legal secretary, commercial litigation, solicitor, property litigation solicitor, litigation lawyer, civil litigation solicitor, litigation paralegal, lawyer, litigation, legal counsel, property litigation, commercial litigation lawyer, employment solicitor, litigation associate, personal injury solicitor, commercial litigation associate, construction solicitor

Leisure and sport	Leisure and sport	business travel consultant, bar staff cruise ship, group exercise instructor, group exercise manager, fitness instructor, aerobics instructor, football coach, cruise staff, fitness professional additional, trainer, group exercise lead, travel consultant, personal trainer, instructor, store floor manager, corporate travel executive, class instructor, centre assistant manager, assistant manager centre, sport massage lecturer
Sciences and education	Life, physical and social sciences	science teacher, biology teacher, teacher, research associate, clinical psychologist, scientist, research assistant, research fellow, teacher of biology, teacher of science, teacher of, lecturer, analytical chemist, laboratory technician, research technician, geotechnical engineer, research scientist, analyst, technician, biomedical scientist
Management	Management	project manager, business analyst, programme manager, business development manager, operations manager, manager, human resource manager, it project manager, project engineer, human resource adviser, consultant, engineer, analyst, quantity surveyor, recruitment consultant, account manager, planner, engineering manager, digital project manager, project planner
Manufacturing and transport	Manufacturing and processing	production manager, manufacturing engineer, buyer, quality engineer, operations manager, production planner, quality manager, production engineer, production supervisor, supply chain manager, maintenance engineer, engineer, manufacturing manager, quality assurance manager, project engineer, material planner, supplier quality engineer, process engineer, technical manager, production team leader
Sales, marketing and admin	Marketing, advertising, PR	marketing manager, marketing executive, marketing assistant, account manager, brand manager, digital marketing executive, digital marketing manager, business development manager, administrator, marketing coordinator, recruitment consultant, manager, product manager, business development executive, account executive, account director, graphic designer, head of marketing, designer, campaign manager
Computing and maths	Mathematics and statistics	analyst, data analyst, statistician, stress engineer, engineer, data scientist, quantitative analyst, business analyst, research associate, research assistant, risk analyst, research analyst, credit risk analyst, biostatistician, model analyst, research fellow, consultant, economist, manager, statistical analyst
Engineering and architecture	Metal processing and mechanical engineering	mechanical design engineer, mechanical engineer, design engineer, maintenance engineer, engineer, cnc machinist, process engineer, manufacturing engineer, field service engineer, technician, project engineer, vehicle technician, quality engineer, electrical maintenance engineer, mechanical fitter, service engineer, production engineer, hgv technician, toolmaker, cnc miller

Sales, marketing and admin	Office and administration	administrator, legal secretary, office administrator, administrative assistant, receptionist, office manager, administration assistant, personal assistant, secretary, human resource administrator, executive assistant, pa, medical secretary, apprentice administrator, receptionist administrator, office assistant, customer service administrator, team secretary, customer service adviser, administration apprentice
Personal services	Personal services	support worker, housekeeper, care assistant, care worker, cleaner, catering assistant, chef, cleaning operative, domestic assistant, housekeeping assistant, apprentice chef, healthcare assistant, nanny, care and support worker, head housekeeper, kitchen assistant, cook, home care worker, nanny housekeeper, care support worker
Manufacturing and transport	Purchasing, procurement, logistics	buyer, supply chain manager, project manager, operations manager, warehouse operative, procurement manager, quantity surveyor, warehouse manager, project engineer, logistics manager, contract manager, production planner, maintenance engineer, purchasing manager, engineer, store manager, quality engineer, logistics coordinator, manager, warehouse supervisor
Financial services	Real estate	property manager, commercial property solicitor, private client solicitor, estate surveyor, mortgage adviser, land manager, real estate solicitor, apprentice lettings negotiator, real estate, planning solicitor, property management surveyor, commercial property lawyer, estate manager, home manager, project manager, front office manager, lettings negotiator, private client lawyer, concierge, receptionist
Sales, marketing and admin	Sales and distribution	sales executive, business development manager, sales manager, account manager, sales administrator, store manager, sales assistant, sales adviser, area sales manager, sales consultant, business development executive, field sales executive, recruitment consultant, telesales executive, sales representative, assistant manager, product manager, customer service adviser, project manager, sales engineer
Security services	Security services	security officer, retail security officer, security guard, security officer relief, commis chef, relief retail security officer, relief security officer, site engineer, security officer retail, store detective, static security officer, chef de partie, pcv driver, mobile security officer, corporate security officer, security support officer, loss prevention officer, security, skilled delivery catering, security area relief officer
Health and care	Social services	support worker, care assistant, home manager, social worker, nursing home manager, qualified social worker, care worker, care home manager, home care worker, deputy manager, healthcare assistant, staff nurse, registered nurse, home care assistant, deputy home manager, relief support worker, carer, registered manager, service manager, registered general nurse

Trade	Trade	store manager, assistant manager, retail store manager, assistant store manager, deputy manager, shop manager, assistant retail manager, store manager designate, buyer, store manager store, branch manager, store manager area, retail manager, store manager beauty store store, retail assistant, assistant shop manager, deputy store manager, supervisor, pharmacist store manager, concession manager
Manufacturing and transport	Transport services	transport planner, forklift truck driver, driver, bus driver, warehouse operative, hgv driver, transport manager, class driver, air import the area, hgv class driver, air import operator, recovery driver, flt driver, highway maintenance operative, ocean freight import operator, logistics coordinator, field service engineer, forklift driver, transport coordinator, transport supervisor
Travel and events	Travel and events	event manager, restaurant manager, special event manager, restaurant general manager, conference and banqueting operations supervisor, assistant restaurant manager, general manager, assistant manager, conference and event manager, housekeeper, event coordinator, bar staff, waiting staff, guest services manager, receptionist, personal assistant, food and beverage supervisor, head housekeeper, pa, community and event manager

Table 5: Top SOC codes in each skill category (shown are SOC codes that in total account for 90 percent of jobs)

Skill category	Top SOC codes
Agriculture, forestry and fishery	1211, 5113, 6139, 2112, 5449, 5111, 9111, 9139, 3119, 3550, 5114, 5112, 9119, 6145, 1121, 2211, 3416, 3113, 2141, 1122, 2434, 8113, 1259, 3539, 8133, 2142, 8129, 2426, 2319, 7125, 9120, 2312, 8223, 8114, 7111, 7130
Architecture and building	2121, 5314, 2126, 1122, 1259, 5241, 3113, 2122, 5315, 3531, 2123, 2433, 5231, 9120, 2434, 3122, 5223, 2431, 3114, 5249, 3119, 3545, 5323, 5319, 5245, 2129, 9139, 8149, 1251, 2461, 2135, 1121, 3422, 2136, 3121, 3567, 8129, 4159, 5313, 2435, 2150, 3562, 8222, 2432, 9235, 7129
Arts	3421, 3422, 2137, 3411, 2471, 2126, 3417, 2431, 3122, 2136, 3416, 4215, 3412, 7111, 5323, 3119, 1259, 3413, 2139, 5245, 5422, 4159, 3113, 2314, 3121, 2319, 3543, 4133, 2129, 2135, 5449, 8134, 1121, 3415, 2121
Banking	3534, 2423, 1131, 2413, 2419, 2424, 1259, 2462, 2136, 3538, 3542, 4159, 4129, 2421, 3539, 1115, 3532, 3545, 4122, 2139, 3543, 1251, 2135, 2134, 3562, 1132, 3544, 4215, 3520, 4123, 3132, 4161, 7129, 3533, 4121, 7219, 7211, 2434, 7111, 2429, 3311, 3111, 3535, 4162, 7130, 1190
Computing	2136, 2137, 2135, 3132, 3131, 2139, 2423, 3539, 2126, 2134, 1259, 5242, 2461, 8133, 4159, 2133, 2462, 3119, 2429, 5249, 2129
Education	2314, 2315, 3562, 6125, 2312, 2319, 2311, 2231, 2316, 3563, 4159, 2211, 6121, 2317, 6126, 6145, 2136, 3119

Electrical engineering	2126, 2123, 5241, 3113, 2124, 2135, 2136, 5249, 5221, 2461, 5231, 3115, 5242, 8133, 3119, 5245, 3122, 2139, 5223, 2121, 8131, 3131, 2122, 1121, 8125, 2129, 3112, 2127, 3132, 1259, 8129
Environmental protection	2129, 3567, 2142, 2121, 1259, 5449, 1121, 2462, 2126, 2122, 3113, 2123, 2139, 2461, 3119, 2112, 2127, 3562, 2136, 2135, 1122, 3539, 2113, 2424, 2111, 1123, 3531, 3543, 2141, 2431, 2423, 2231, 2426, 8133, 4159, 5249, 1251, 3131, 2434, 7121, 2133, 2429, 3115, 5223, 3111, 1190, 3550, 5314, 4215, 5319
Finance, accountancy	2421, 4122, 1131, 3534, 3538, 4159, 2423, 2433, 1259, 4129, 2424, 3539, 3562, 2136, 2135, 4121, 3531, 3542, 2462, 1190, 3545, 3535, 4162, 2429, 1251, 1121, 4161, 3541, 4215, 3537, 4132, 3131, 1132, 2139, 3543, 2134, 3132, 2434
Food preparation	5434, 5435, 9272, 1223, 9273, 6145, 5436, 9274, 3219, 3546, 9233, 7111, 8212, 6122, 1259, 9279, 2136, 6121, 4159
Health	2231, 2211, 6145, 3219, 2221, 2222, 2112, 6141, 2219, 1181, 2213, 2217, 3218, 6143, 4159, 2212, 3217, 1242, 2215, 2223, 3111, 2462, 3235, 3562, 3239, 2136, 3119, 2426, 1259, 4216, 4131
Human resource management	3562, 3563, 4159, 1135, 2231, 3567, 1259, 4138, 1121, 3132, 5434, 2121, 2462, 1190, 2136, 1242, 2135, 3131, 7130, 1251, 3539, 2139, 4162, 2434, 2424, 6145, 2319, 1181, 1223, 1131, 2133, 4161, 4215, 2134, 3113, 2423, 4216, 2433, 2413, 3239, 3520, 9273, 1132, 1122, 3543, 7220, 3538, 3119, 4131, 4214, 2461
Humanities	2312, 2314, 2311, 2114, 3412, 2212, 2426, 2452, 2211, 3411, 2136, 2231, 2319, 2315, 2135, 6145, 3219
Insurance	4132, 3533, 3531, 3543, 3534, 3542, 7129, 2423, 2424, 1242, 3532, 3538, 5231, 2462, 7219, 1259, 7211, 2419, 4159, 3544, 3562, 2231, 1131, 2136, 2425, 2434, 2413, 4112, 3119, 3520, 3545, 3539, 2135, 4129, 2139, 1181, 4162, 4123, 3132, 2433, 1190, 2134, 1132
Journalism and information	2471, 3412, 2426, 3534, 4159, 3543, 2472, 1259, 2136, 2137, 2112, 3539, 3520, 4215, 3416, 2135, 2121, 3542, 2129, 4214, 3562, 2150, 3131, 1132, 2139, 3119, 1134, 4131, 2429, 3132, 2311, 7214, 4129, 2312, 3421, 2451, 2119
Law	2413, 3520, 2419, 4212, 2462, 3562, 4132, 3531, 3544, 4159, 4131, 4215, 2443, 3534, 2231, 9241, 1135, 3567
Leisure and sport	3443, 6212, 3442, 9274, 2319, 1173, 3219, 1259, 2136, 3563, 3441, 1225, 6123, 3520, 3414, 3413, 2312, 3546, 3542, 4215, 7130, 7129, 4214, 2221, 6122, 7219, 3311, 2429, 2135, 2139
Life, physical and social sciences	2314, 2112, 2426, 3119, 2111, 2312, 2311, 2119, 2212, 2113, 3111, 2211, 2136, 2315, 2129, 2121, 2429, 6125, 3235, 2425, 2462, 3562, 2150, 2231, 3218, 1259, 2126, 3539, 3543, 2139, 4215, 3443, 3219, 2122
Management	2423, 1259, 2134, 2135, 3562, 2136, 3539, 3545, 2424, 1121, 1190, 2121, 2139, 1132, 1135, 2462, 1131, 4215, 3543, 3534, 4161, 3541, 1133, 3538, 2133, 2129, 3131, 2461, 2429, 2413, 4159, 2126, 3542, 3132, 7129, 1122, 2122, 4162, 1136, 7111, 2432, 2127, 3563, 2433, 3567, 1139, 2419, 2150, 2426, 1115, 7130
Manufacturing and processing	1121, 3113, 3541, 2461, 2462, 2127, 3115, 1133, 4133, 1190, 2122, 3116, 4134, 3538, 1259, 2126, 2129, 3119, 2133, 2136, 8129, 1162, 8133, 9273, 3543, 2135, 4131, 7130, 8114, 3531, 5241, 3131, 5223, 2121, 3111, 9260, 5221, 5449, 2429, 2423

Marketing, advertising, PR	3543, 3545, 1132, 3562, 4151, 3542, 2472, 2137, 4159, 7129, 3421, 7111, 2135, 1134, 1259, 7130, 3539, 2136, 3131, 2423, 4215, 3412, 3538, 3546, 7219, 2471, 7113, 3541, 2473, 3416, 2426, 7125, 1190, 3534, 2139, 7211, 7220
Mathematics and statistics	2136, 2425, 3539, 2423, 3534, 2426, 2135, 5449, 3543, 2119, 2429, 3111, 2112, 3122, 2113, 2129, 2121, 2126, 2122, 2424, 1115, 2139, 3119, 3131, 3413, 3132, 3562, 2111, 2461, 4215, 3531, 2133, 4159, 7111, 3542
Metal processing and mechanical engineering	2126, 2122, 3113, 5231, 5221, 5223, 5249, 2127, 8125, 5215, 2129, 5241, 1121, 2123, 3122, 2461, 3119, 2121, 2136, 5314, 5222, 3115, 8129, 2135, 5449, 3531, 1259, 8133, 5232, 3116, 8211, 2462, 5242, 3567
Office and administration	4159, 4215, 4214, 4216, 4212, 4161, 3562, 3132, 7219, 4131, 3539, 4138, 7211, 4122, 4162, 3520, 4211, 4129, 4217, 3131, 1259, 4132, 2136, 3541, 3543, 1251, 4112, 4151
Personal services	6145, 6231, 9272, 9233, 5434, 5435, 6122, 6232, 6240, 6121, 4159, 3219, 6141, 9234, 9273, 2231, 9279, 6221, 3239, 3132, 4214, 1242, 9274, 7111, 6211, 6146, 6222, 3119, 2129, 9249, 1251, 9132
Purchasing, procurement, logistics	3541, 1133, 4134, 1190, 1259, 8129, 1162, 1121, 4133, 3113, 9260, 3543, 3538, 3545, 2461, 4159, 7130, 2135, 2462, 5231, 2433, 7111, 1251, 2136, 2122, 8211, 5249, 3531, 4131, 2123, 1122, 3119, 2129, 3539, 3116, 5223, 3115, 3131, 2126, 2121, 1161, 7219, 2134, 2423, 8222, 3542, 5241, 5434, 2133, 8125, 2429, 8212, 3132, 8133
Real estate	1251, 2413, 3544, 2434, 2419, 3534, 3520, 4159, 1242, 4216, 1259, 2421, 4161, 6232, 1131, 4215, 9279, 3539, 7111, 2432, 7219, 2462, 7129, 3545, 3538, 4212, 3542, 1132, 3541, 2423, 2424, 3546
Sales and distribution	3542, 7129, 3545, 7130, 7111, 1132, 7113, 1190, 3543, 3562, 7219, 4151, 1259, 2423, 7211, 4159, 3538, 2136, 3541, 3534, 2135, 4161, 1121, 3132, 3539, 5434, 1131, 4162, 5231, 2139, 4215, 2133, 3563
Security services	9241, 3567, 5434, 2121, 7111, 2462, 1173, 2139, 8211, 3319, 8149, 4159, 3539, 5249, 3113, 1259, 1190, 2424, 2461, 6232, 3119, 9249, 2136, 3563, 8212, 8213, 2129, 5436, 3132, 2231, 5231, 1122, 2429, 1121, 3565
Social services	6145, 2442, 2231, 1242, 3239, 4159, 6121, 6146, 1181, 3562, 6141, 3219, 3132, 1121, 1190, 3231, 4162, 2211, 1259, 1251, 2413, 3539, 3520, 3543, 4214, 2219, 3235
Trade	1190, 7130, 7111, 3541, 4159, 1254, 7219, 1131, 9272, 8129, 3520, 7129, 3542, 2136, 3545, 5231, 9273, 3538, 4133, 1132, 6212, 5232
Transport services	8211, 2436, 4134, 8212, 5231, 8222, 8213, 8129, 9260, 1161, 2121, 3536, 5249, 3538, 8233, 3119, 5223, 8142, 8239, 4159, 3113, 3539, 1259, 9211, 1190, 7211, 4133, 1122, 9120, 5330, 2136, 7219, 3565, 8214, 5449, 2126
Travel and events	3546, 1223, 9273, 3131, 4215, 1259, 4216, 9272, 6231, 9274, 1221, 7219, 5436, 1190, 4159, 4214, 3562, 9279, 3543, 1121, 6212, 2136, 1225, 3563, 3542, 7130, 6240, 3239, 7211, 1135, 1122

References

Nikolaos Askitas and Klaus F. Zimmermann. The internet as a data source for advancement in social sciences. *International Journal of Manpower*, 36(1):2–12, apr 2015. ISSN 0143-7720. doi: 10.1108/IJM-02-2015-0029. URL http://www.emeraldinsight.com/doi/10.1108/IJM-02-2015-0029.

H. Bakhshi, J. Downing, M. Osborne, and P. Schneider. *The Future of Skills: Employment in 2030*. London: Pearson and Nesta, 2017.

Michele Belloni, Agar Brugiavini, Elena Meschi, and K. Tijdens. Measurement Error in Occupational Coding: An Analysis on Share Data. SSRN Scholarly Paper ID 2539080, Social Science Research Network, Rochester, NY, dec 2014. URL https://papers.ssrn.com/abstract=2539080.

Roberto Boselli, Mirko Cesarini, Stefania Marrara, Fabio Mercorio, Mario Mezzanzanica, Gabriella Pasi, and Marco Viviani. WoLMIS: a labor market intelligence system for classifying web job vacancies. *Journal of Intelligent Information Systems*, pages 1–26, sep 2017. ISSN 0925-9902, 1573-7675. doi: 10.1007/s10844-017-0488-x. URL https://link.springer.com/article/10.1007/s10844-017-0488-x.

Ian Brinkley and Elizabeth Crowley. From 'inadequate' to 'outstanding': making the UK's skills system world class, apr 2017. URL https://www.cipd.co.uk/knowledge/work/skills/uk-skills-system-report.

Burning Glass Technologies. Markets, Technology, Solutions, 2017. URL http://burning-glass.com/uk/.

Burning Glass Technologies. JobPulse™ Analytics Dashboard, 2018. URL http://burning-glass.com/jobpulse/.

Anthony Carnevale, Tamara Jayasundera, and Dmitri Repnikov. Understanding online job ads data: A technical report. Technical report, 2014. URL https://georgetown.app.box.com/s/nre5ybcw97e8gpyq502w.

Theresa Cosca and Alissa Emmel. Revising the Standard Occupational Classification system for 2010. *Monthly Labor Review*, pages 32–41, 2010. ISSN 0098-1818. URL http://www.jstor.org/stable/monthlylaborrev.2010.08.032.

Roxana Danger. A methodology for taxonomy generation and maintenance from large collections of textual data, 2016. URL https://conferences.oreilly.com/strata/strata-eu-2016/public/schedule/detail/53307.

David J. Deming and Lisa B. Kahn. Skill Requirements across Firms and Labor Markets: Evidence from Job Postings for Professionals. *National Bureau of Economic Research.*, 2017.

DISCO II Portal. European Dictionary of Skills and Competences. URL http://disco-tools.eu/disco2_portal/index.php.

Peter Elias and Margaret Birch. SOC2010: revision of the Standard Occupational Classification. *Economic & Labour Market Review*, 4(7):48–55, 2010. URL https://ideas.repec.org/a/pal/ecolmr/v4y2010i7p48-55.html.

Emsi. Emsi Developer - In-Depth Workforce Analytics, 2018. URL http://www.economicmodeling.com/developer-workforce/.

ESDC. National Occupational Classification, 2017. URL http://noc.esdc.gc.ca/English/noc/Introduction.aspx?ver=16.

Inna Grinis. The STEM Requirements of 'Non-STEM' Jobs: Evidence from UK Online Vacancy Postings and Implications for Skills & Knowledge Shortages. SSRN Scholarly Paper ID 2864225, Social Science Research Network, Rochester, NY, may 2017. URL https://papers.ssrn.com/abstract=2864225.

Hyukjun Gweon, Matthias Schonlau, Lars Kaczmirek, Michael Blohm, and Stefan Steiner. Three Methods for Occupation Coding Based on Statistical Learning. *Journal of Official Statistics*, 33(1):101–122, 2017. doi: 10.1515/jos-2017-0006. URL https://www.degruyter.com/view/j/jos.2017.33.issue-1/jos-2017-0006/jos-2017-0006.xml.

Ray Harper. The collection and analysis of job advertisements: A review of research methodology. *Library and Information Research*, 36(112):29–54, sep 2012. ISSN 1756-1086. URL http://www.lirgjournal.org.uk/lir/ojs/index.php/lir/article/view/499.

Christian Hennig. Cluster-wise assessment of cluster stability. *Computational Statistics and Data Analysis*, pages 258–271, 2007.

International Labour Organization. ISCO-08: Introductory and Methodological Notes, jun 2016. URL http://www.ilo.org/public/english/bureau/stat/isco/isco08/index.htm.

Faizan Javed and Ferosh Jacob. Data Science and Big Data Analytics at Career Builder. In *Big-Data Analytics and Cloud Computing*, pages 83–96. Springer, Cham, 2015. ISBN 978-3-319-25311-4 978-3-319-25313-8. URL https://link.springer.com/chapter/10.1007/978-3-319-25313-8_6. DOI: 10.1007/978-3-319-25313-8_6.

Daniel Jurafsky and James H. Martin. Question Answering and Summarization. In *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*. Pearson, 2nd. edition, may 2008. ISBN 978-0-13-187321-6.

Lucia Mýtna Kureková, Miroslav Beblavý, and Anna Thum-Thysen. Using online vacancies and web surveys to analyse the labour market: a methodological inquiry. *IZA Journal of Labor Economics*, 4:18, sep 2015. ISSN 2193-8997. doi: 10.1186/s40172-015-0034-4. URL https://doi.org/10.1186/s40172-015-0034-4.

Jey Han Lau and Timothy Baldwin. An Empirical Evaluation of doc2vec with Practical Insights into Document Embedding Generation. *arXiv:1607.05368 [cs]*, jul 2016. URL http://arxiv.org/abs/1607.05368. arXiv: 1607.05368.

Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. Distributed Representations of Words and Phrases and their Compositionality. In C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems 26*, pages 3111–3119. Curran Associates, Inc., 2013. URL http://papers.nips.cc/paper/5021-distributed-representations-of-words-and-phrases-and-their-compositionality.pdf.

Office for National Statistics. SOC2010 volume 2: the structure and coding index - Office for National Statistics. URL https://www.ons.gov.uk/methodology/classificationsandstandards/standardoccupationalclassificationsoc/soc2010/soc2010volume2thestructureandcodingindex.

Jeffrey Pennington, Richard Socher, and Christoper Manning. Glove: Global Vectors for Word Representation. In *EMNLP*, volume 14, pages 1532–1543, jan 2014. doi: 10.3115/v1/D14-1162.

Christian Posse. Cloud Jobs API: machine learning goes to work on job search and discovery, 2016. URL https://cloud.google.com/blog/big-data/2016/11/cloud-jobs-api-machine-learning-goes-to-work-on-job-search-and-discovery.

Roger Thomas and Peter Elias. Development of the Standard Occupational Classification. *Population Trends*, (55):16–21, 1989.

Arthur Turrell, Bradley Speigner, James Thurgood, Jyldyz Djumalieva, and David Copple. Understanding the demand for labour from the bottom-up. forthcoming.

U.S. Bureau of Labor Statistics. SOC Major Groups, mar 2010. URL https://www.bls.gov/soc/major_groups.htm.