

Online Appendix

Teo Firpo¹, Julia Baumann, Timm Teubner, Anastasia Danilov

Section I: Data and method

In this section, we describe in detail our method to extract and categorize digital and non-digital skills from our dataset of online job ads.

A. Cleaning the text data

The raw data includes the title of the job role being advertised, as well as the content of the job posting. Ads in our dataset may appear in different languages, primarily French, Arabic and English. We restrict our attention to French-language ads only because the textual analysis we employ below is not suitable for use in more than one language. French-language ads constitute the bulk of ads in our data (92% ads). Moreover, because ads can often be repeated, we deduplicate ads that share the entire text in the main content section.² This leaves us with 59,254 unique ads.

Next, we clean the text in our sample of unique ads. We lowercase all text, replace all accented letters with equivalents (e.g., ‘é’ with ‘e’), and remove all punctuation. We also remove non-essential words (also known as ‘stop words’), such as pronouns and prepositions, as they can interfere with the textual analysis.

B. Identifying digital skills

Our second step is to identify and extract digital skills from our processed text data. There are many definitions of digital skills, depending on context and purpose (Djumalieva & Sleeman, 2018b; van Laar et al., 2017). For instance, the European Commission’s ESCO classification defines digital competencies as the “ability to use information and communication technologies effectively to achieve work objectives” (Directorate-General for Employment, Social Affairs and Inclusion, 2017), while the OECD defines digital skills as “the capacity to use ICT devices and applications to access and manage information and solve problems” (OECD, 2015); relatedly, for UNESCO digital skills are “a range of abilities to use digital devices, communication applications, and networks to access and manage information” (UNESCO, 2018). In order to identify and extract a specific set of digital skills from our data, we employ an approach developed by Djumalieva and Sleeman (Djumalieva & Sleeman, 2018b), which involves three steps.

In a first step, we manually identify a small number of digital skills (either single words or bigrams). To do so, we select five ads from our dataset for jobs roles explicitly related to digital tasks (web development, web marketing, etc.) and extract 59 digital skills (the text of the ads and the skills extracted are presented in Figure A.1 below). These digital skills range from software or programming languages (‘Python’, ‘PHP’, ‘HTML’), digital marketing skills

¹ Corresponding author: teo.firpo@hu-berlin.de

² We previously deduplicate ads if they share the same unique URL; this is due to the fact that many ads remain online for extended periods of time and are thus crawled more than once. Removing them reduces our dataset by two thirds (75,1%). After this step, 9,831 duplicated ads (i.e., with different URLs) remain. There are different reasons why these ads are duplicated, e.g., employers might decide to take down the job offer and then repost it. Our final sample is therefore a conservative estimate of the absolute number of vacancies.

(‘SEO’, or Search Engine Optimization, the process of optimizing traffic to websites) to broader digital skills (‘e-commerce’). These 59 skills form our starting set from which we identify a other digital skills.

To do so, in a second step we employ a word embeddings model to identify synonymous skills. This is a machine learning model that turns words from a corpus of text into vectors; these vectors can be used to find other words that share a number of characteristics, such as appearing in similar contexts.³ We train our model on the text of all of our ads and use it to identify skills similar to our starting set of 59 skills; in each case, the resulting skills are manually checked against the ESCO, UNESCO and OECD definitions of skills given above, retaining the skills that fall under at least one definition. Through this last step, we identify an additional 544 skills, for a total of 603 digital skills.

Figure A.1 – Job ads used to identify a starting set of digital skills

<p>First ad Job title: “Développeur .net Senior” Description: “Le groupe est leader des applications de la carte à puce (Pétrole, Collectivités, Loisirs et Monétique) en France. Fort d’une croissance importante, la filiale en Tunisie souhaite renforcer son effectif IT par un Développeur .NET senior Principales missions : Développement de toutes les fonctionnalités techniques. Être force de proposition et contribuer dans les choix techniques. Respect des bonnes pratiques de codage. Contribution dans l’encadrement et la formation des nouvelles recrues. Compétences : Maîtrise parfaites des environnements de développement .Net Une très bonne maîtrise des langages de programmation : C# Bonnes connaissances des technologies Web front end : HTML, Javascript et CSS (Bootstrap et JQuery) Maîtrise du langage SQL Maîtrise de l’orienté objets et des patterns les plus importants. Bon niveau de communication en français à l’oral et à l’écrit Capacité à encadrer des profils juniors Capacité à estimer et analyser les tâches Capacité à faire des POC et proposer des solutions techniques talents.”</p>
<p>Second ad Job title: “Responsable Marketing Digital / E-commerce” Description: “Société commerciale, sise à Dende-Manouba, désire recruter un chargé de marketing et de communication digital dont la mission: Valoriser l’image de la société, assurer sa présence sur le web et générer des visites de prospects qualifiés sur le site internet et les réseaux sociaux; Produire l’ensemble des contenus numériques et multimédias (rédactionnels, visuels, photos, vidéos...) et les diffuser avec une approche cross-canal (Réseaux sociaux, sites web, mini-sites, emailings, ...); Mise en place de stratégie marketing, développement et la gestion du site web; Animer les réseaux sociaux et faire grandir les communautés de la société; Exigences de l’emploi De formation multimédia ou infographie, avec une spécialisation en e-marketing/e-communication/marketing stratégique.</p>
<p>Third ad Job title: “Webmarketer” Description: “Connaissances : Passionné(e) d’internet, vous avez une forte culture du web et du digital Maîtrise des techniques de SEO / SEA : Google est votre ami ! Maîtrise des outils Google (Google Analytics...) Parfaite maîtrise du Français / bonnes connaissances en Anglais, Connaissances du langage HTML, Connaissances de base en PHP / MySQL. Expérience significative requise en webmarketing / référencement Formation exigée : Bac + 2/3 min et métiers du web (Webmarketing/Web/Communication...) Expérience exigée : 1 à 2 ans min” [...]</p>
<p>Fourth ad Job title: “Ingénieur développeur JS / MEAN Stack” Description: “Exigences: - Expérience dans le développement frontend avec Angular 2 => 9 -</p>

³ This approach is also used by Atalay et al. (2018) to expand the set of tasks they extract from newspaper job ads.

Connaissances **CSS / SCSS** - Expérience dans le **développement backend** avec **NodeJS / Loopback** ou frameworks **Express JS** équivalents / **Spring** - Connaissances basiques de **MongoDB / Mysql** - Connaissances basiques en **Data science** - Bonne communication en français et en anglais - Discipline et autonomie - Capacité à faire des recherches sur Google / Stack Overflow / lire la documentation / explorer le code source - 0-3 années d'expérience (Débutant / junior / confirmé) - Bac + 5 ou équivalent en **ingénierie informatique / logicielle / réseaux** ou équivalents”.

Fifth ad

Job title: “Ingénieur développement”

Description: “Compétences techniques exigées Maîtrise de la suite **BI SAP Business Objects**. Connaissances des **outils de visualisation de données (QLIK)** Maîtrise d'un ou de plusieurs des langages suivants : **Spark, Scala, R, Python, Java**. Connaissance des **outils ETL (Talend, ABINITIO)** et **entrepôts de données..** Connaissance des outils et technologies **Big Data: Hadoop, Hive, OOZIE, ELK**. Expérience avec les bases de données relationnelles (**Oracle, MySQL**) et **NoSQL (Mongo, Cassandra)**. Bonnes connaissances en administration **système Linux.**”

Note: This figure presents the five ads used to identify a ‘starting set’ of digital skills. The ads are presented in French as originally published. The starting set of digital skills is highlighted in bold.

B. Grouping digital skills

We next group our digital skills into categories. We start by categorizing our 603 digital skills into eight broader categories. We adapt a taxonomy developed by Djumalieva and Sleeman (2018a). The advantage of this taxonomy is that it is also derived from job postings data and thus the keywords in each category match closely our digital skills. The disadvantage is that it includes a large number of categories, which can reduce its analytical value; we therefore further consolidate the clusters that match our digital skills into eight broad categories.

The first category (‘programming’) includes skills related to web, app and software development; it comprises programming languages, frameworks and software. ‘Technical support and IT’ brings together a series of skills relating to system administration, IT support and security, and networks. The category ‘office software’ covers typical IT tools, such as the Microsoft Office Suite, as well as references to common office IT skills (‘emailing’). ‘Graphic design and digital content creation’ combines related tools (such as ‘Photoshop’) with skills associated with the creation of content for the web (such as blogs or newsletters). Finally, we create four additional categories for ‘data science and data engineering’, ‘digital marketing’, ‘data and business analysis’ and ‘machine learning and AI’. A small number of skills that do not fall under these eight categories are left unclassified as ‘other’.⁴

As in the original (Djumalieva & Sleeman, 2018a) taxonomy, the categories are not mutually exclusive. The reason is that there are skills that span a variety of uses; for instance, ‘SQL’ – a computer language used to manage databases – is related to both data engineering and programming, as well as technical support and IT.

Additionally, we identify nine non-digital skill categories in our data. We adopt the taxonomy used by Deming and Noray (2020), focusing on nine non-digital skills.⁵ The first skill is

⁴ These are relatively broad digital skills, such as ‘digital skills’, or ‘IoT’ (Internet of Things).

⁵ We do not use their five remaining categories which cover digital skills. These five categories map well onto our eight digital skill categories: ‘office software’, ‘technical support’, ‘data analysis’, ‘specialized software’ (which directly relates to our

‘social’, which includes references to ‘communication’, ‘teamwork’ and ‘collaboration’. The next category is ‘cognitive’ skills, which is modelled after ‘nonroutine analytical’ tasks described in Autor et al (2003), and includes keywords such as ‘thinking’, ‘researching’ and ‘analyzing’. A third category, ‘character’, includes a number of ‘soft skills’ and personality traits, such as ‘time management’ and ‘attention to details’. The ‘management’ category includes skills related to project and people management (e.g., ‘mentoring’, ‘staff development’). Other categories include skills related to ‘creativity’, ‘writing’, and ‘finance’. The ‘business systems’ category includes techniques used in business processes, such as ‘Six Sigma’ or ‘KPIs’. Finally, ‘customer service’ includes keywords related to interacting with clients (such as ‘customer’ or ‘sales’). All keywords used to identify each non-digital category are presented in Table A.1 below. We translate the keywords in each category into French and identify alternative spellings in our data. Following (Atalay et al., 2018), we also use our word embeddings model to uncover synonymous skills in our data. In total, we identify 104 keywords and phrases across the nine non-digital skill categories.

‘programming’ category, as well as software skills in our ‘data science and data engineering’ and ‘graphic design and digital content creation’ categories) and ‘machine learning and AI’.

Table A1: Keywords used to identify non-digital skill categories	
<i>Category</i>	<i>Description</i>
Social	Communication ('communication', 'comuniquer'), Collaboration ('collaboratif', 'collaboration', 'collaborer'), Negotiation ('negotiation', 'negotier'), "Team" ('sens equipe', 'travailler equipe', 'esprit equipe'), Persuasion ('persuader'), Listening ('ecoute'), Presentation ('presenter', 'presentation'), English* ('anglais'), French* ('francais'), Sharing Spirit* ('esprit de partage'), Interpersonal Skills* ('relationnel')
Cognitive	Solving ('resoudre'), Research ('recherche'), "Analy" ('analyse', 'analyser'), Decision ('choisir', 'decider', 'decideurs'), Thinking ('penser', 'pensez', 'reflechir', 'reflexion'), Math ('mathematique'), "Statistic" ('statistique'), Calculation ('calcul', 'calculer')
Character	Organization Skills ('organisation', 'organise'), Time Management ('gestion temps'), Detail-Oriented ('attention detail', 'attention details', 'sens detail', 'souci detail'), Meeting Deadlines ('honorer deadlines', 'respecter deadlines', 'respectez deadlines'), Multi-Tasking, Energetic ('energetique'), Self-Starter, Initiative ('esprit initiative', 'prendre initiative', 'prise initiative', 'prise initiatives'), Self-Motivation ('motivation personnelle'), Rigorous* ('rigoureux', 'rigueur'), autonomy* ('autonome', 'autonomie'), Self-Management* ('self-management')
Creativity	"Creativ" ('creatif', 'creative', 'creativite')
Writing	Writing ('ecrire', 'redactionnelle', 'redactionnelles'), Editing ('editer', 'editeur', 'edition'), Preparing Reports ('elaborer rapport', 'redaction documentations', 'redaction rapports', 'rediger rapports'), Preparing Proposals ('redaction proposition')
Management	Supervisory ('superviser'), Leadership ('leadership', 'team leader'), Mentoring ('mentoring', 'mentorat', 'coaching'), Staff Supervision/Development ('supervision personnel', 'developpement personnel'), Performance/Personnel Management ('organisation personnel')
Finance	"Financ" ('finance', 'financiere'), Budgeting ('budgetisation', 'elaborer budget', 'gere budget', 'gestion budget'), Accounting ('comptabilite'), Cost ('cout')
Business systems	Systems Development/Integration/Architecture ('architecture systeme', 'architecture systemes', 'developpement systeme', 'developpement systemes', 'integration systeme', 'integration systemes'), Business Intelligence/Systems/Planning/Strategy, Six Sigma, KPIs ('kpi', 'kpis')
Customer service	Customer, Sales ('ventes', 'vente'), Patient ('patients', 'patient', 'patience'), Client ('client', 'clients')
<p>Note: This table presents the categorization of non-digital skills following Deming and Noray (2020). The first column present the name of the non-digital skills category. The second column presents the keywords and phrases used to identify each category. The English-language keywords and phrases are taken from Deming and Noray (2020); in parentheses we present the French-language translations (where not identical to the original). * indicates the keywords are not present in the original category; rather, they are were identified by the authors using a word embeddings model.</p>	

References

- Atalay, E., Phongthientham, P., Sotelo, S., & Tannenbaum, D. (2018). New technologies and the labor market. *Journal of Monetary Economics*, *97*, 48–67. <https://doi.org/10.1016/j.jmoneco.2018.05.008>
- Autor, D., Levy, F., & Murnane, R. (2003). The Skill Content of Recent Technological Change: An Empirical Exploration. *Quarterly Journal of Economics*, *118*(4), 1279–1333.
- Deming, D. J., & Noray, K. (2020). Earnings dynamics, changing job skills, and stem careers. *Quarterly Journal of Economics*, *135*(4), 1965–2005. <https://doi.org/10.1093/qje/qjaa021>
- Djumaieva, J., & Sleeman, C. (2018a). An Open and Data-driven Taxonomy of Skills Extracted from Online Job Adverts. *Developing Skills in a Changing World of Work*, August, 425–454. <https://doi.org/10.5771/9783957103154-425>
- Djumaieva, J., & Sleeman, C. (2018b). Which digital skills do you really need? Exploring employer demand for digital skills and occupation growth prospects. *Nesta.Org.Uk*, 1–10. https://media.nesta.org.uk/documents/Which_digital_skills_do_you_really_need.pdf
- Inclusion, D.-G. for E. S. A. and. (2017). *ESCO Handbook* (Issue September). <https://ec.europa.eu/esco/portal/document/en/0a89839c-098d-4e34-846c-54cbd5684d24>
- OECD. (2015). *Does having digital skills really pay off?* <https://doi.org/https://doi.org/10.1787/24121401>
- UNESCO. (2018). *Digital skills critical for jobs and social inclusion*.
- van Laar, E., van Deursen, A. J. A. M., van Dijk, J. A. G. M., & de Haan, J. (2017). The relation between 21st-century skills and digital skills: A systematic literature review. *Computers in Human Behavior*, *72*, 577–588. <https://doi.org/10.1016/j.chb.2017.03.010>