Unlocking skills dynamics: Harnessing big online data and NLP methods in emerging economies

Verónica Escudero (escudero@ilo.org) and Franziska Riepl, International Labour Organization, Switzerland

Abstract

In today's dynamic job market, creating and sustaining decent jobs requires a thorough understanding of current skill demand and its role in facilitating transitions to quality employment. This brief presents an innovative approach leveraging big data and natural language processing to extract skills information from online vacancy data, offering insights into skills dynamics in low- and middle-income economies. Findings reveal notable skill patterns, highlighting the importance of foundational socio-emotional and cognitive skills. This approach empowers policymakers and researchers to deepen their understanding of skills dynamics in previously understudied contexts and inform strategies for employment and skills development initiatives. Harnessing insights from online data opens up vast opportunities for economic development and sustainable growth.

In a rapidly evolving job landscape marked by technological advancements and the transition to a more sustainable economy, acquiring and adapting to the right skills has become imperative. Possessing these skills can increase the resilience to global transformations and shocks, whereas their absence increases the risk of unfavourable labour market outcomes, leaving individuals vulnerable to poverty and exclusion. Recognising this urgency, institutions and governments have elevated skills development and lifelong learning to policy priorities (ILO 2023; OECD 2023; UNESCO UIL 2020; European Commission 2020).

Scholarly research in Europe and the United States has extensively used skills taxonomies leveraging granular data from occupational classification systems, such as U.S. O-NET and ESCO, and vacancy data to study skills dynamics and their effect on wages and employment (e.g., Autor, Levy, and Murnane 2003; Acemoglu and Autor 2011; Deming and Kahn 2018; Atalay et al. 2020). However, knowledge on skills dynamics outside these regions remains limited due data constraints. Existing skills classifications and taxonomies are not easily transferable to diverse contexts. While efforts like the PIAAC and STEP skills measurement surveys (see OECD 2019 and World Bank 2014, respectively) provide some insight, they are limited in scope and coverage.

This brief presents an innovative solution leveraging big data to circumvent these challenges, developed in Escudero, Liepmann, and Podjanin (Forthcoming). They introduce a conceptual framework to categorize tasks performed on the job into skills categories and subcategories, coupled with a natural language processing (NLP) methodology to extract skills information from unstructured online vacancy data. The taxonomy, comprising 14 unique skills across cognitive, socioemotional, and manual categories, is tailored to low- and middle-income economies and adaptable to country-specific contexts. It captures all sectors and occupations, including manual labour, and can also be applied to applicant data to study skills supply and mismatch as well as the relationship between skills and job transitions.

This methodology aims to shed light on skills dynamics in previously understudied economies, as job board data is now available across numerous countries and years. These insights will empower governments, businesses, and individuals to target skills development efforts more effectively, fostering resilient economies and promoting decent work for all.

Methodology: Taxonomy and implementation

Taxonomy. The underlying building block of the approach is a skills taxonomy encompassing job-specific tasks and personal attributes, categorised into cognitive, socio-emotional and manual skills. Within each broad category, there are 14 more nuanced skills subcategories that are further categorised (Box 1). Drawing from labour economics (particularly Deming and Kahn 2018) and psychology literature, the taxonomy expands existing efforts to accommodate individual country-contexts, focusing on low- and middle-income countries, and is designed to be applicable to online data. Notably, the taxonomy includes three types of manual skills often overlooked in analyses focused on high-income countries.

For each skill, a set of keywords is selected based on the existing taxonomies and seminal studies not exclusively reliant on online data sources (Almlund et al. 2011; Autor, Levy, and Murnane 2003; Atalay et al. 2020; Deming and Noray 2020; Kureková et al. 2016; Heckman and Kautz 2012; Hershbein and Kahn 2018). This list of keywords is supplemented with country specific information from O-NET Uruguay, and 518 synonyms of the original words obtained from an online synonym generator (www.wordreference.com).

Cognitive skills	Examples of keywords and expressions
Cognitive (narrow sense)	Critical thinking, analyse, planning
Computer (general)	Computer, excel, internet skills
Software (specific) and technical support	Programming language, computer maintenance
Machine learning and artificial intelligence	Neural networks, supervised learning
Financial	Budgeting, accounting, SAP
Writing	Editing, reports, proposals
Project management	Project management
Socio-emotional skills	
Character	Organized, energetic, reliable
Social	Communication, teamwork
People management	Leadership, management
Customer service	Sales, client, advertise
Manual skills	
Finger dexterity	Sorting, control machine
Hand-foot-eye coordination	Attending animals, driving, repair
Physical	Resistance, walking, carrying

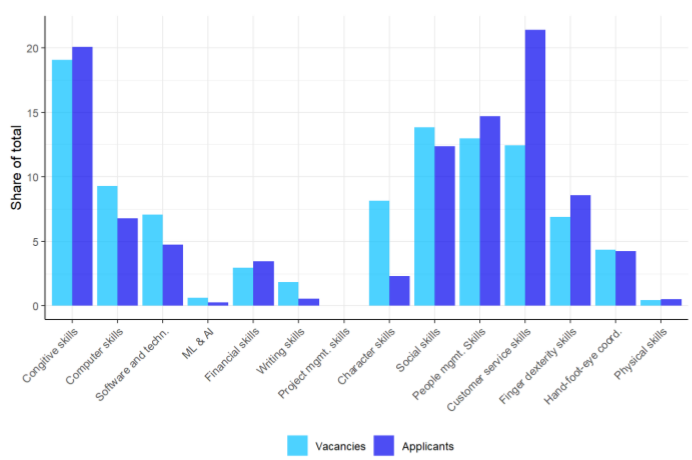
Implementation. The taxonomy is applied to online vacancy data, sourced by the online job board BuscoJobs (Uruguay) and the job aggregator Adzuna (South Africa), using natural language processing (NLP) techniques. First, open-text descriptions of job postings are pre-processed to fit the structured format of the skills dictionary. A skill is identified as present if at least one keyword or expression from the dictionary appears in the vacancy posting. For applicant data, available for Uruguay, the employment histories uploaded by applicants are analysed using the same NLP procedure applied to job postings, exploiting the descriptions of all job spells since their last application on the platform. The advancements in NLP have thus transformed the accessibility of unstructured data, enabling the exploration of previously unattainable labour and development questions.

Findings and possible applications

After providing a general overview of the prevalence of skills, we will then illustrate how the methodology can be used to study specific labour market-related questions.

The skill patterns observed in job postings and applicants' CVs for Uruguay reveal notable similarities and some key differences (Figure 1). While both vacancies and applicants prioritize social and general cognitive skills, applicants additionally highlight their customer service skills. Conversely, project management, physical skills, and machine learning and AI (ML&AI) skills are less prevalent for both applicants and vacancies. These patterns reflect labour market characteristics, but also the nature of the data: cognitive and social skills are widely applicable and crucial in white-collar clerical and service occupations, which dominate in online data. Meanwhile, specialized skills such as project management and ML&AI are specific to certain jobs, such as business managers and software engineers, that require extensive education and additional complementary skills. The small presence of physical skills is related to the typical underrepresentation of elementary occupations in the data, particularly within the agricultural sector, where jobs are primarily advertised via personal networks or offline postings. Therefore, when carrying out analyses using online job board data it is crucial to take into account data coverage: dynamics within well-represented sectors and occupations can be studied in great detail, but caution is needed regarding statements that are representative for the entire economy. Reassuringly, similar analyses for Brazil, Russia and South Africa, confirm that the overall ordering of skills frequencies is remarkably stable across countries.

Figure 1. Skills identified in vacancies and applicants' job histories in Uruguay, expressed as the share of all vacancies and job spells, respectively.

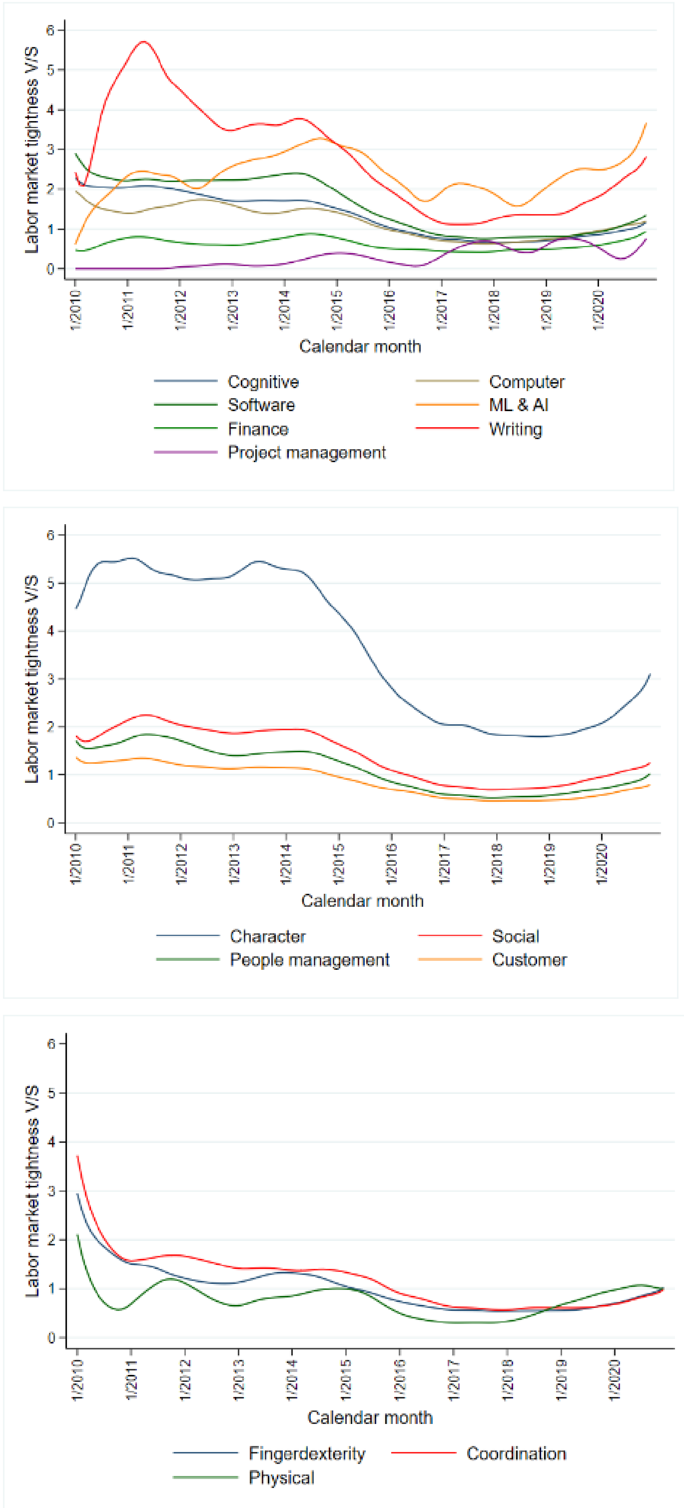


Data source: BuscoJobs Uruguay, 2010-2020.

These findings provide a robust foundation for further analyses on skills dynamics, mismatches between skills required and those available in the workforce, skills complementarities, and the effect of skills on labour market outcomes. Figure 2 provides an illustration of how the data can be used to analyse labour market tightness, which can help identify priorities for skills development policies by pinpointing areas with skill shortages. The data shows stable trends over time, with tightness generally declining, in concurrence with the economic slowdown and rise in unemployment in Uruguay, from 2015 onwards. Tightness is largest for character and writing skills, whereas it is generally low for manual skills. Notably, the measures derived from online data indicate higher tightness for more general skills, which are highly valued by employers but challenging to quantify. While many workers may have character skills like being trustworthy, punctual, and well-organized, they do not necessarily showcase their abilities when applying to a job. This underscores the importance of understanding job search and recruiting behaviours to enhance labour market matching procedures.

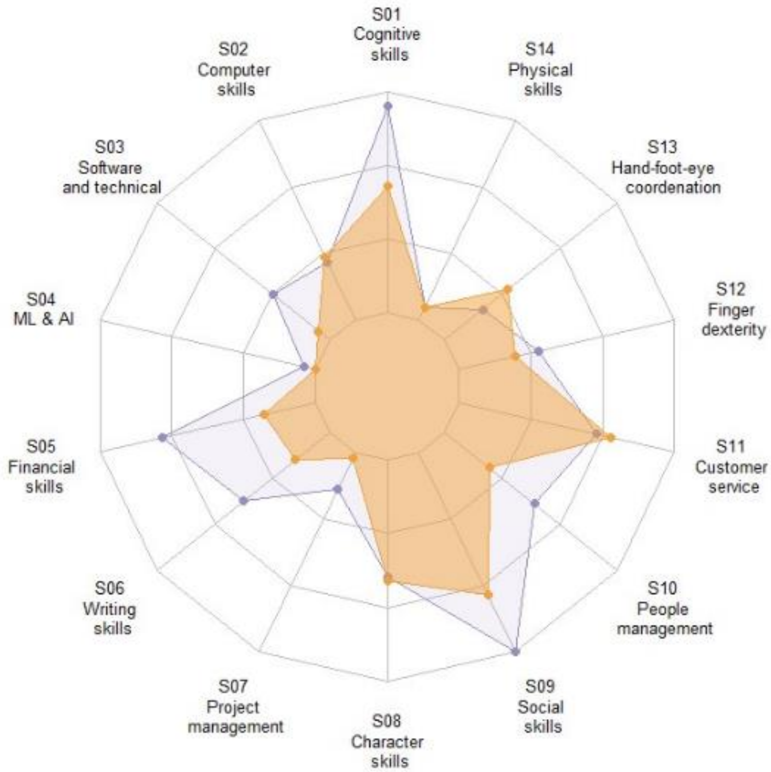
Lastly, Figure 3 (see below) presents an example of how exploring the relationship between skills and wages offers insights into salary distributions and the value of different skill sets in the job market. The spider web chart shows the average skills profile of jobs in the top and bottom 20% of the salary distribution for South Africa. High-paying jobs require a broader range of skills, especially social, cognitive and financial skills, while low-paying jobs prioritize coordination and customer service skills. The prevalence of general abilities, such as cognitive and social skills, also among well-paid jobs highlights their importance as foundational competencies that complement more specialized skills like software or ML&AI. These analyses contribute to a deeper understanding of skills dynamics and inform strategies for skills development and workforce planning.

Figure 2. Evolution of labour market tightness, measured as the number of vacancies requiring a given skill over the number of applicants furnishing said skill, for the period 2010-2020 (BuscoJobs, Uruguay). Top panel: cognitive skills, middle panel: social skills, bottom panel: manual skills.



Data source: BuscoJobs Uruguay, 2010-2020.

Figure 3. Radar chart of skills for the top (purple) and bottom (yellow) quintile of posted salaries. The largest value is rescaled to one, i.e. the outermost ring of the chart. The further away from the center a marker is, the more important is a given skill.



Data source: Adzuna, South Africa, 2016-2021.

Policy recommendations

The exemplary findings from applying the taxonomy and NLP methodology by Escudero, Liepmann, and Podjanin (Forthcoming) offer valuable insights with significant policy implications. The analysis highlights the persistent importance of social and character skills and core cognitive skills, which are often undervalued despite their critical role in fostering collaboration and integration into modern workplaces. To enhance workforce readiness and promote career advancement, policymakers could prioritize the development of these skills, recognizing their relevance across various job levels and occupations. By establishing a strong foundation in broadly applicable skills, individuals can better adapt to evolving job requirements, with additional technical skills building upon these core competencies. Further analyses regarding complementarities between different skills as well as the skill content of occupations, combined with insights on individuals' labour market trajectories, allow to identify pathways for skill development and career progression that can then be promoted in the context of reskilling and upskilling initiatives.

The freely available taxonomy presented here provides a valuable resource for policymakers and researchers worldwide, with the flexibility to be adapted to local contexts through translation and the inclusion of country-specific keywords and expressions. This is possible thanks to the advancements in NLP that have revolutionized the accessibility of unstructured information, allowing for the exploration of previously unattainable developmental questions. Harnessing the insights derived from online data together with a comprehensive skill taxonomy geared towards low- and middle-income economies offers vast opportunities to deepen our understanding of labour market dynamics. Through the analysis of online data, policymakers, researchers, and workers' and enterprises' organizations can now gain deeper insights into skills dynamics and their implications for labour market outcomes and, more broadly, for economic development and achieving the sustainable development goals.

References

Acemoglu, Daron, and David Autor. 2011. 'Skills, Tasks and Technologies: Implications for Employment and Earnings'. In The Handbook of Labor Economics, O. Ashenfelter and D. Card (eds), 4:1043–1171. Amsterdam: Elsevier.

Almlund, Mathilde, Angela Lee Duckworth, James Heckman, and Tim Kautz. 2011. 'Chapter 1 - Personality Psychology and Economics'. In Handbook of the Economics of Education, edited by Eric A. Hanushek, Stephen Machin, and Ludger Woessmann, 4:1–181. Handbook of The Economics of Education. Elsevier.

Atalay, Enghin, Phai Phongthiengtham, Sebastian Sotelo, and Daniel Tannenbaum. 2020. 'The Evolution of Work in the United States'. American Economic Journal: Applied Economics 12 (2): 1–34.

Autor, David, Frank Levy, and Richard Murnane. 2003. 'The Skill Content of Recent Technological Change: An Empirical Exploration'. The Quarterly Journal of Economics 118 (4): 1279–1333.

Deming, David, and Lisa Kahn. 2018. 'Skill Requirements across Firms and Labor Markets: Evidence from Job Postings for Professionals'. Journal of Labor Economics 36 (S1): S337–69.

Deming, David, and Kadeem Noray. 2020. 'Earnings Dynamics, Changing Job Skills, and STEM Careers'. Quarterly Journal of Economics 135 (8): 1965–2005.

Escudero, Verónica, Hannah Liepmann, and Ana Podjanin. Forthcoming. 'Using Online Vacancy and Job Applicants' Data to Study Skills Dynamics'. Research in Labor Economics.

European Commission. 2020. 'European Skills Agenda for Sustainable Competitiveness, Social Fairness and Resilience'. https://ec.europa.eu/social/main.jsp?catId=1223&langId=en.

Heckman, James, and Tim Kautz. 2012. 'Hard Evidence on Soft Skills'. Labour Economics 19 (4): 451–64.

Hershbein, Brad, and Lisa B. Kahn. 2018. 'Do Recessions Accelerate Routine-Biased Technological Change? Evidence from Vacancy Postings'. American Economic Review 108 (7): 1737–72.

International Labour Organization (ILO). 2023. 'The ILO Strategy on Skills and Lifelong Learning 2030'. ISBN 978-92-2-038965-2.

Kureková, Lucia Mýtna, Miroslav Beblavý, Corina Haita, and Anna-Elisabeth Thum. 2016. 'Employers' Skill Preferences across Europe: Between Cognitive and Non-Cognitive Skills'. Journal of Education and Work 29 (6): 662–87.

OECD. 2019. Skills Matter: Additional Results from the Survey of Adult Skills. Paris: OECD Publishing. https://doi.org/10.1787/1f029d8f-en.

———. 2023. OECD Skills Outlook 2023: Skills for a Resilient Green and Digital Transition. Paris: OECD Publishing. https://doi.org/10.1787/27452f29-en.

UNESCO Institute for Lifelong Learning (UIL). 2020. Embracing a Culture of Lifelong Learning - Contribution to the Futures of Education Initiative. Hamburg: UNESCO UIL.

World Bank. 2014. 'STEP Skills Measurement Snapshot 2014'. https://www.worldbank.org/content/dam/Worldbank/Feature%20Story/Education/STEP%20Snapshot%202014_Revised_June%2020%202014%20%28final%29.pdf.