

Uso de *big data* en el análisis de perfiles por competencias

Octubre 6, 2022

Verónica Escudero, Economista Principal, Departamento de Investigaciones de la OIT

Estudio realizado con:

Fidel Bennett (Universidad de Chile), Hannah Liepmann (OIT) & Ana Podjanin (OIT)

Motivación

- **Amplia bibliografía sobre la dinámica de las competencias en Europa y en EE.UU, gracias a la riqueza de los datos existentes (O-NET)**
 - La tecnología informática reemplaza el trabajo rutinario (por ej., Autor et al. 2003; Goos et al. 2014)
 - Algunos autores prevén que la IA y los robots reemplazarán también al trabajo analítico no rutinario (por ej., Frey and Osborne, 2017)
 - Sin embargo, se ha descubierto que las habilidades interactivas son menos replicables (Arntz et al., 2016)
- **Menos conocimiento fuera de estos países, a pesar de la importancia del tema**
 - La imputación de datos de EE.UU requiere suposiciones estrictas e irreales (por ej., Almeida et al. 2017; Bhorat et al. 2020)
 - Los datos de encuestas (PIACC y STEP) existen como cortes transversales únicos, principalmente para áreas urbanas (por ej., Lewandowski et al. 2019, 2020)

► Motivación

- Una alternativa consiste en utilizar los datos de plataformas laborales o bolsas de empleo en línea – **big data**
 - Se basa en la creciente bibliografía en los EE. UU. utilizando datos de vacantes (por ej., Deming and Kahn, 2018; Hershbein and Kahn, 2018)
 - Granularidad y estructura de panel de los datos
 - ¿Cómo cambian los perfiles ocupacionales por competencias a través del tiempo?
 - Análisis específico a un país: no se necesita imputaciones ni suposiciones
 - Datos actualmente disponibles en muchos países
 - **¿Cómo utilizarlos?**

► Nuestro estudio

Enfoque metodológico

- Desarrolla una taxonomía de competencias basada en la bibliografía de las ciencias sociales, adaptada para contextos de países de ingresos medios y bajos
- Implementa la taxonomía utilizando datos del portal uruguayo *BuscoJobs*
- Crea ocupaciones ISCO-08 a 2-dígitos para evaluar la variación de los perfiles ocupacionales por competencias
- Desarrolla una metodología para implementar la taxonomía que se puede aplicar en datos similares en cualquier otro país/contexto.

El objetivo final:

- Explorar si la dinámica de los perfiles ocupacionales por competencias se puede estudiar utilizando datos de vacantes en línea y datos de solicitantes de empleo fuera de Europa y EE.UU
- En el futuro, analizar el papel de las competencias para promover mejores transiciones en el mercado laboral y/o mediar shocks y transformaciones (por ej., tecnológicas, ambientales, etc.)

Taxonomía

- Objetivo: que la taxonomía sea completa pero sucinta, adecuada para estudiar problemáticas en países de ingresos medios y bajos, y adaptada para el análisis usando *big data*
- Basada en la bibliografía de las ciencias sociales (en particular, economía laboral y psicología)
 - ¿Cómo favorece el cambio tecnológico a ciertas competencias? (por ej., Acemoglu and Autor, 2011)
 - Estudios de psicología (Almlund et al. 2011) – por ej., “modelo de los cinco factores de la personalidad” (McCrae and Costa 2008)
- 3 amplias categorías – competencias cognitivas, socioemocionales y manuales:
 - Cada categoría engloba atributos productivos muy diferentes: patrones únicos de aprendizaje, y por lo tanto, rendimientos distintos (Lise and Postel-Vinay 2020).
 - Captan tanto las competencias que pertenecen a las tareas que los trabajadores realizan en el trabajo, como aquellas que se refieren a los atributos personales de los individuos.

Operacionalización de la taxonomía

- 14 subcategorías – por ej., informática, financiera; carácter, social; destreza de los dedos, etc. → [snapshot](#)
- Las subcategorías se definen en función de palabras clave y expresiones, o “competencias únicas”
 - Compilación de palabras clave/expresiones de varios artículos, lo más exhaustiva posible
 - Cuando aparece una palabra clave en los datos, asumimos que la subcategoría respectiva es relevante para la vacante o para la historia laboral del solicitante
 - Contamos la cantidad de veces que se usa una palabra clave como indicador de la intensidad de la competencia.
- Incluso si algunas subcategorías están estrechamente relacionadas, las palabras utilizadas para categorizarlas son mutuamente excluyentes
 - Esto permite la identificación única de categorías de competencias
 - Por ejemplo: “habilidades de carácter” $\leftrightarrow$ “competencias cognitivas”

► *BuscoJobs*

- Portal privado de empleo, líder en Uruguay, con una cobertura efectiva estimada del 60% de las vacantes en línea
- Serie de larga duración y de alta frecuencia de 2010 a 2020:
 - Amplia información: vacantes de empleo (N=87.000), biografías laborales de los solicitantes (N=1,23 millones), y patrones de búsqueda (es decir, postulaciones realizadas)
 - Característica única, ya que la mayoría de las bolsas y plataformas de empleo solo contienen datos de vacantes
 - Datos detallados: alrededor de 40 posibles campos/características para cada vacante y alrededor de 60 para cada solicitante
 - Información de texto

► **BuscoJobs**

- Los datos de las plataformas de empleo no son totalmente representativos:
 - Depende de la tasa de penetración de Internet (83% en Uruguay)
 - Porcentaje de personas viviendo en áreas urbanas (95% en Uruguay)
 - Incidencia de la informalidad (24% en Uruguay)
 - El tamaño de la empresa y el sector (el uso de Internet es menor en la construcción, la agricultura y la hostelería)
- En nuestros datos:
 - Los solicitantes de empleo tienden a ser más jóvenes, más educados y con más probabilidades de vivir en la capital, en comparación con la población activa
 - Podría requerir ponderación y/o enfoque en segmentos seleccionados de la fuerza laboral (Fabo & Kurokova, 2022)
 - Sin embargo, los datos capturan de manera significativa a postulantes con nivel educativo medio e incluso bajo, además de mano de obra altamente calificada.

Ocupaciones (CIUO-08): Trabajos de los postulantes de *BuscoJobs* vs trabajador promedio en Uruguay, 2020



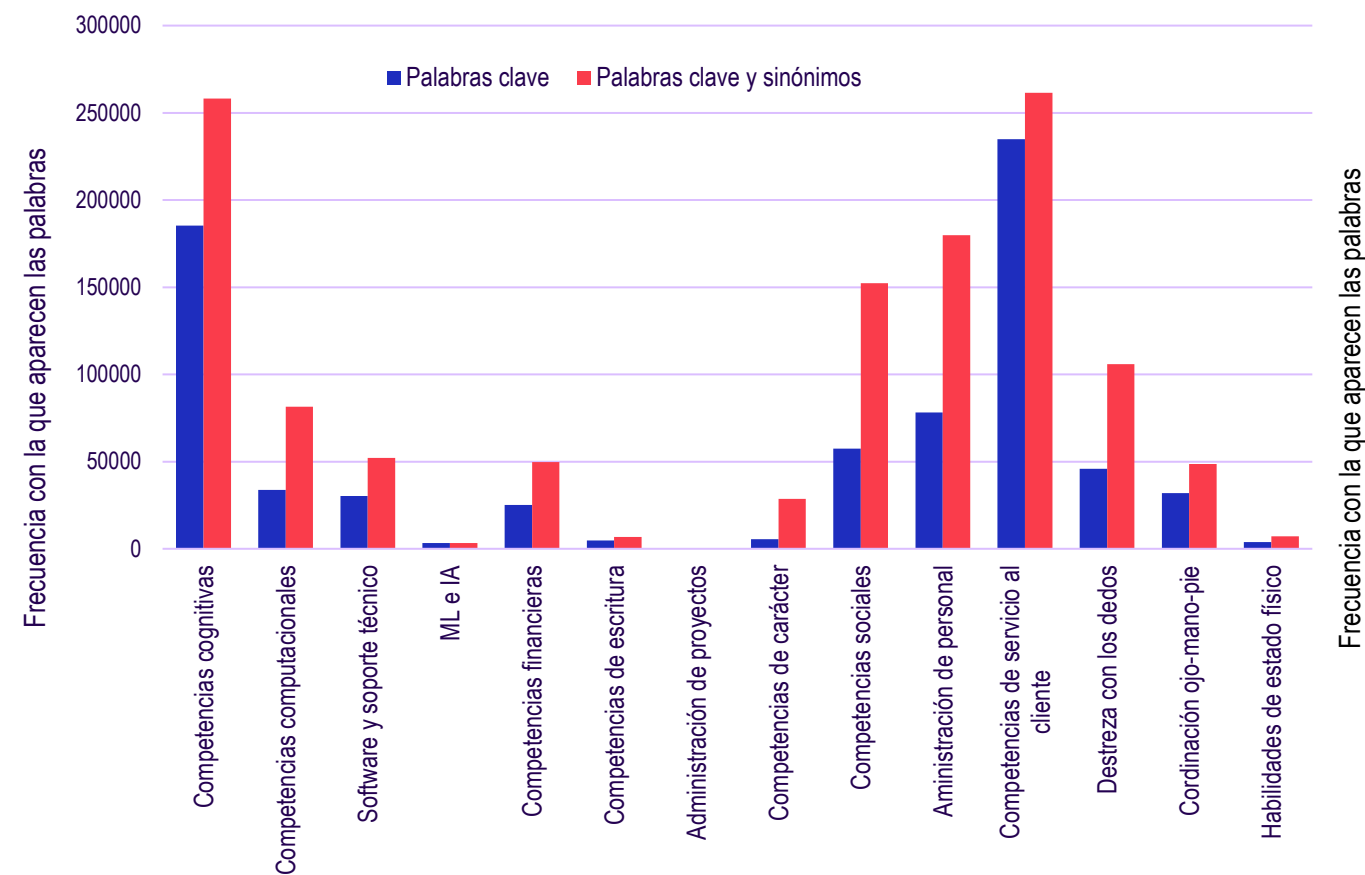
► Implementación: Machine learning y técnicas de análisis de texto

- Usamos palabras clave y expresiones: teniendo en cuenta los sinónimos (webscrapping)
 - 800 palabras clave/expresiones
- Mapeamos las palabras clave/expresiones que caracterizan cada una de las 14 subcategorías y las descripciones en texto abierto de las vacantes e historias laborales de los postulantes de BuscoJobs
- Aplicamos la técnica PNL / PLN para almacenar las palabras clave y sus sinónimos y crear las subcategorías de competencias como variables en los datos
- Creamos variables proxy de la intensidad de la competencia (es decir, la cantidad de veces que aparece una habilidad única)

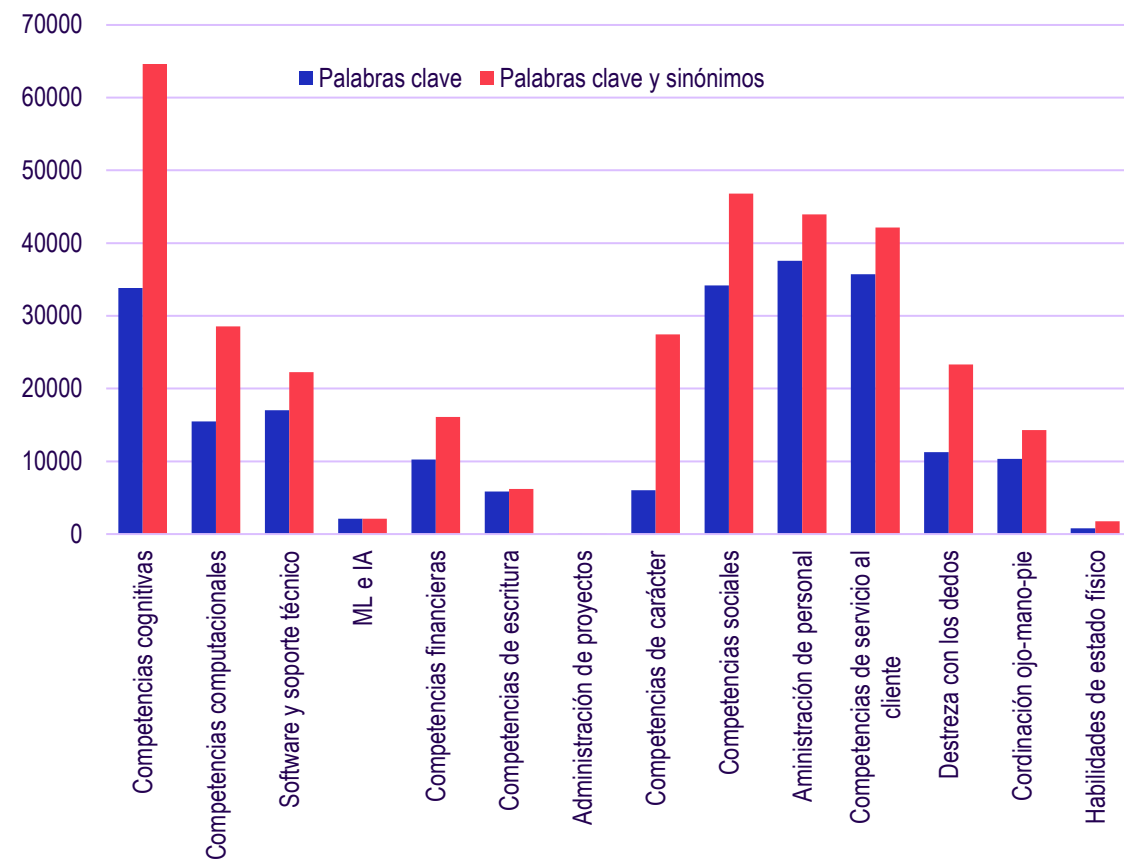
Resultados

- La implementación de la taxonomía es exitosa → la metodología funciona:
 - Clasificamos más del 64% de las observaciones de los postulantes y el 94% de las vacantes
 - O: 3,9 (sd 2,5) subcategorías en promedio por vacante y 2,5 (sd 2,5) por trabajador
 - Esto es sorprendente, dada la heterogeneidad en la calidad de la información proporcionada
- El uso de varias fuentes de información para extraer palabras y expresiones es central para la eficacia de la metodología → recuadro
- Es imprescindible utilizar datos específicos del país para estos análisis
 - La comparación de nuestro enfoque (datos de BuscoJobs) con el enfoque alternativo de utilizar los datos de O-NET (asumiendo perfiles ocupaciones por competencias similares en EE.UU y Uruguay), muestra grandes diferencias → gráficos

Subcategorías: palabras clave/expresiones más sinónimos



(a) Historias laborales de los solicitantes



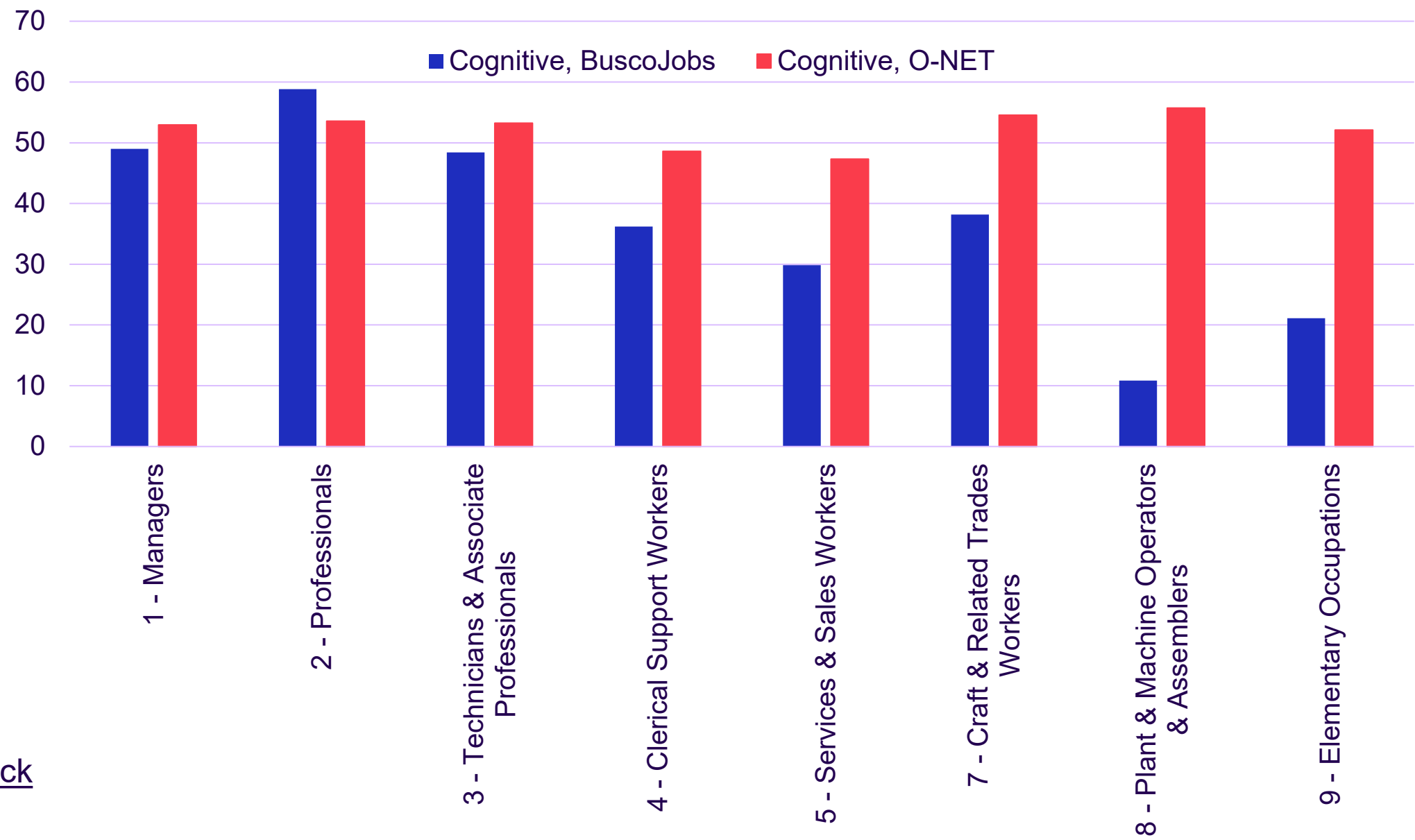
(b) Vacantes

Palabras clave/expresiones identificadas por tipo de fuente, todos los años

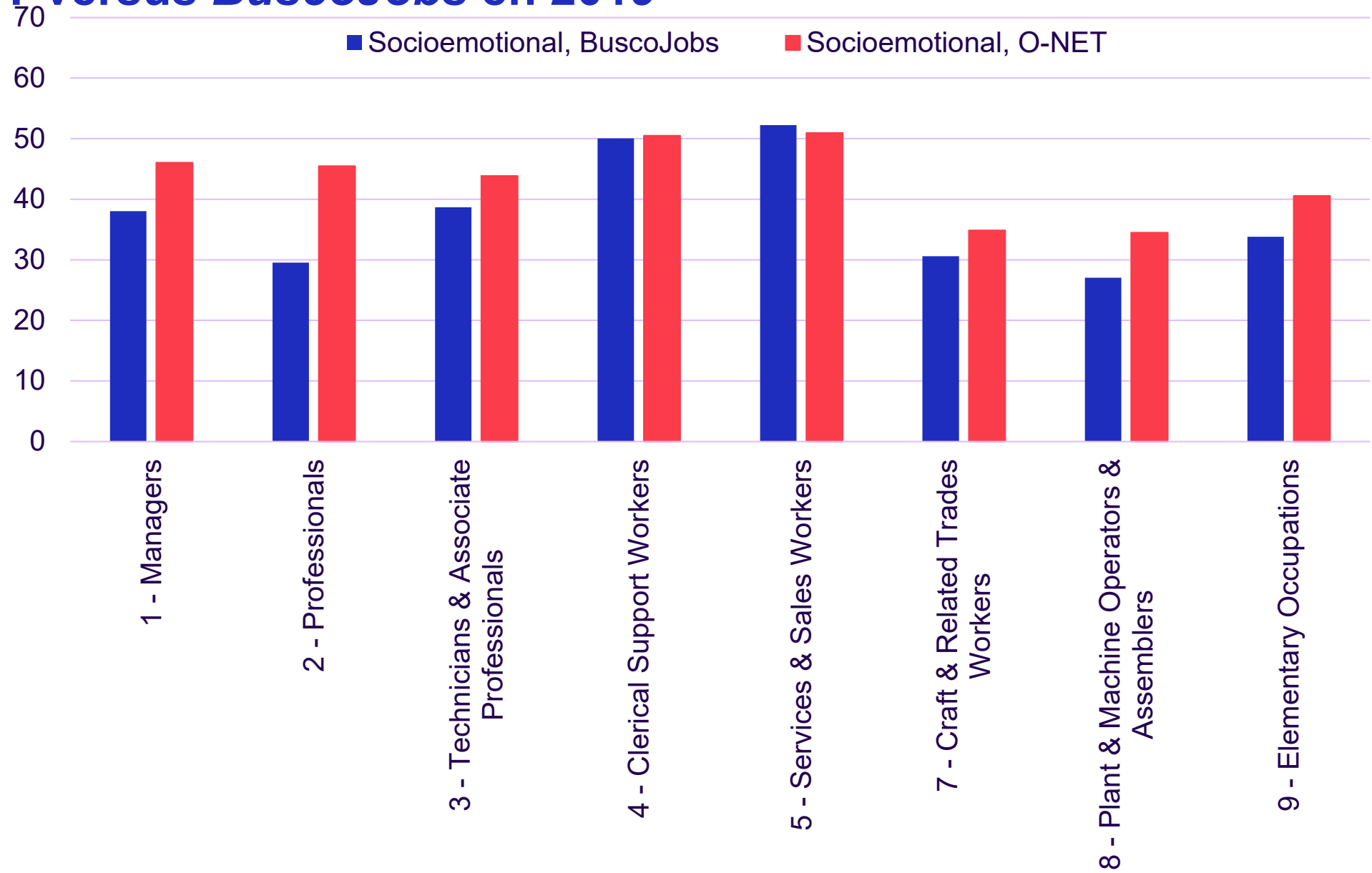
	Unique skills (keywords/expressions) captured (1)	Source type 1: Online data (2)	Source type 2: Non-online data (3)	Source type 3: O-NET Uruguay (3)
Vacancies	372,879	269,603 (72.31%)	65,853 (17.66%)	37,411 (10.03%)
Applicants	1,065,305	645,809 (60.62%)	325,416 (30.55%)	94,080 (8.83%)

[Back](#)

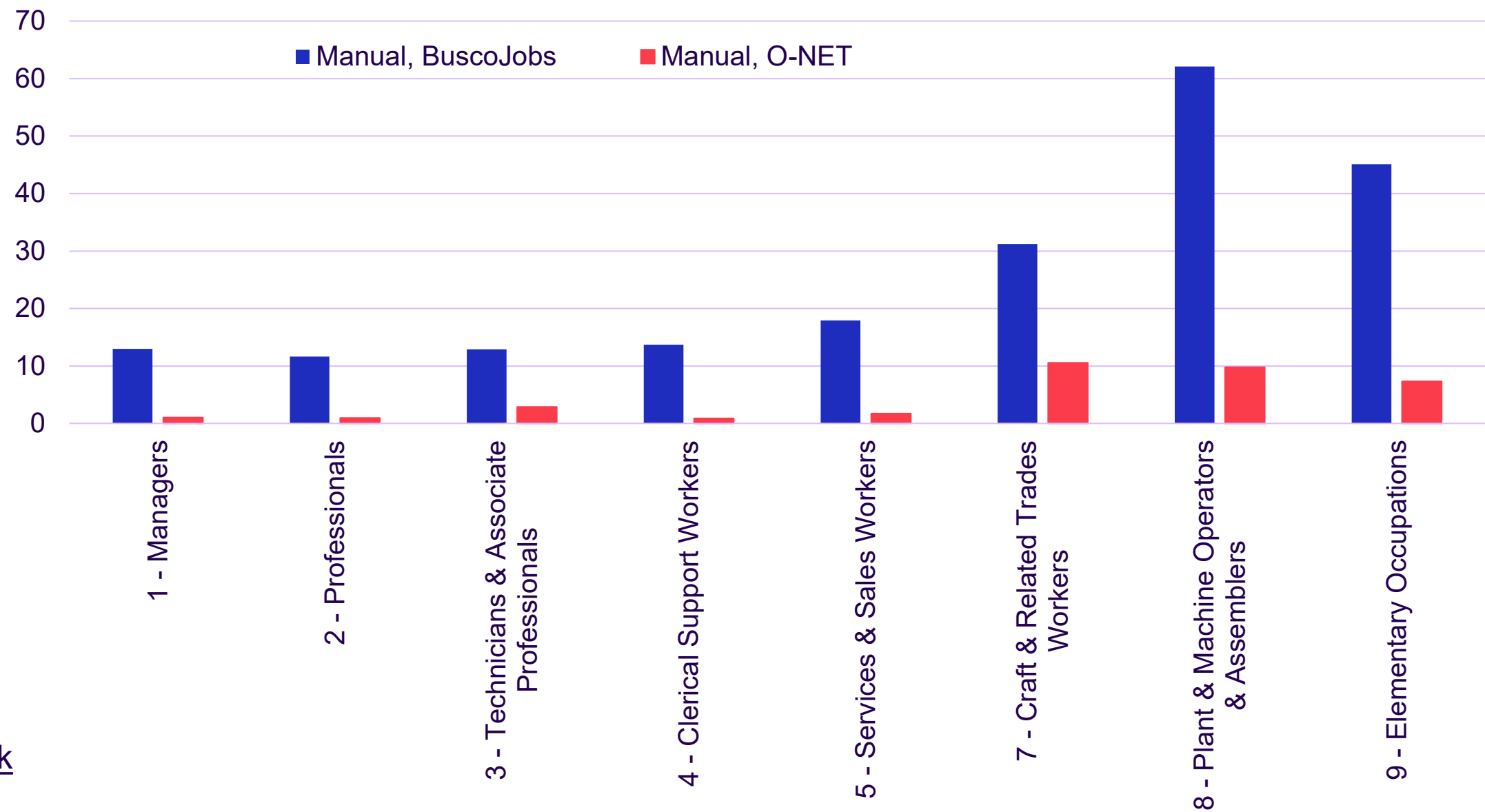
Distribución de los perfiles ocupacionales por competencias: O-NET versus *BuscoJobs* en 2019



Distribución de los perfiles ocupacionales por competencias: O-NET versus *BuscoJobs* en 2019



Distribución de los perfiles ocupacionales por competencias: O-NET versus *BuscoJobs* en 2019



Conclusión

- Los datos en línea sobre vacantes y postulantes en plataformas laborales son una fuente adecuada para estudiar la dinámica de las competencias y los perfiles ocupacionales por competencias en países de ingresos medios.
- Nuestro enfoque se basa en datos que actualmente están disponibles en muchos países del mundo.
 - Permite análisis específicos de cada país que no necesitan asumir que los perfiles ocupacionales por competencias son iguales en todos los países.
- Capacidad de estudiar dinámicas de perfiles ocupacionales por competencias en detalle y a lo largo del tiempo, tomando en cuenta tanto la demanda como la oferta laboral.
 - Datos de panel y detalle de los datos
- Los problemas de representatividad de los datos deben tenerse en cuenta al elegir las preguntas de investigación.
- Hasta donde sabemos, somos los primeros en explorar este enfoque fuera de EE.UU. y Europa.



Gracias.

escudero@ilo.org

Categorización de las competencias, palabras clave y fuentes de información:

19

Cognitive skills	Cognitive skills (narrow sense)	DK (2018)	Problem solving, research, analytical, critical thinking, math, statistics Mathematics, adaptability, direction, control, planning Data analysis, data engineering, data modelling, data visualization, data mining, data science, predictive analytics, predictive models Analyse, design, devising rule, evaluate, interpreting rule, sketch Calculation Bookkeeping, correcting, measurement Information processing, decision making, generation of ideas, memory
	Computer (general) skills	DK (2018)	Computer, spreadsheets, common software, Excel, PowerPoint Computer literacy, Internet skills, Word, Outlook, Office, Windows
	Software (specific) skills and technical support	DK (2018) & DN (2020)	Programming language or specialized software, Java, SQL, Python Computer installation, computer repair, computer maintenance, computer troubleshooting, web development, site design
	Machine Learning and Artificial Intelligence	DK (2018) DN (2020)	Artificial intelligence, machine learning, decision trees, apache hadoop, Bayesian Networks, Automation Tools, Neural Networks, Support Vector Machines (SVM), Supervised learning, TensorFlow, MapReduce, Splunk, Convolutional Neural Network (CNN), Cluster Analysis
	Financial skills	DK (2018)	Budgeting, accounting, finance, cost
	Writing skills	DK (2018)	Writing Editing, reports, proposals
	Project management skills	DK (2018)	Project management

Categorization of skills, keywords and sources:

Socio-emotional skills	Character skills (conscientiousness, emotional stability and openness to experience)	DK (2018)	Organized, detail oriented, multitasking, time management, meeting deadlines, energetic Self-starter, initiative, self-motivated Competent, achieving, hardworking, reliable, punctual, resistant to stress, creative, independent
	Social skills (including agreeableness and extraversion)	DK (2018)	Communication, teamwork, collaboration, negotiation, presentation Team, persuasion, listening Flexibility, empathy, assertiveness Advice, entertain, lobby, teaching Interact with others, verbal abilities
	People management skills	DK (2018)	Supervisory, leadership, management (not project), mentoring, staff Staff supervision, staff development, performance management, personnel management
	Customer service skills	DK (2018)	Customer, sales, client, patient Persuading, selling Advertise, sell, buy, purchase Repetitive customer service

Categorization of skills, keywords and sources:

Manual skills	Finger dexterity skills	ALM (2003), under routine manual tasks	Picking, sorting, repetitive assembly, mixing ingredients, baking ingredients, sewing and decorative trimming, operating tabulating machines, packing agricultural produce Control, equip, operate Repetitive movements
	Hand-foot-eye coordination skills	ALM (2003), under nonroutine manual tasks	Attending cattle, attending other animals, driving to transport passengers, <i>driving to transport charge</i> , piloting airplanes, pruning and treating ornamental and shade trees, performing gymnastic feats, <i>performing other sports requiring skill and balance</i> Accommodate, renovate, repair, restore, serving, <i>cleaning</i> Reaction on time, fine manipulations
	Physical skills	O-NET Uruguay	Resistance, time dedicated to walking and running, <i>carrying heavy loads</i>

[Back](#)

► Creación de variables ocupacionales: Machine learning and text mining approach

- Creación de las ocupaciones a 2 dígitos de la CIUO-08 para los datos de las vacantes y de los solicitantes de empleo
 - Permitir evaluar cómo cambian las competencias dentro de las ocupaciones
 - Utilizando un enfoque similar de PNL, más un modelo predictivo de aprendizaje automático para mejorar la precisión

[Back](#)