



Nota Técnica

Revisión excepcional del Índice de avisos laborales de internet del mes de agosto 2026^{1 2}

División Estadísticas
Banco Central de Chile

Resumen

Una revisión exhaustiva de los registros de agosto de 2026 (en particular, de la última semana del mes) permitió identificar un aumento inusual de avisos republicados desde otros portales en uno de los cuatro portales que convergen en el IALI (Tabla 1). Los casos examinados presentan coincidencias en el cargo, el empleador y la localización con avisos disponibles en otro de los portales de empleo incluidos en el indicador, aunque pueden contener cambios menores de redacción u omitir el nombre de la empresa³.

Tabla 1. Composición de avisos portal A según origen.

	abr-26	may-26	jun-26	jul-26	ago-26
Avisos directos en el portal	83,6%	86,0%	82,8%	85,2%	58,3%
Avisos republicados desde otros portales	16,4%	14,0%	17,2%	14,8%	41,7%

Fuente: Portal de empleo A

Si bien la republicación entre portales es un fenómeno normal y esperado, lo ocurrido en el mes de agosto, en un portal de empleo, elevó de forma abrupta el conteo general de avisos, sin una justificación ligada a la operatoria regular del mismo. Su efecto sobre las variaciones mensuales e interanuales dependerá de cómo cambie su incidencia respecto de los períodos de comparación.

Esta nota documenta la incorporación al cálculo del Índice de Avisos Laborales de Internet (IALI) de un procedimiento para detectar avisos duplicados publicados en más de un portal. El procedimiento combina similitud de texto y comparación de atributos que corresponden a un mismo aviso. Su aplicación se circunscribe solo a dos portales analizados y a publicaciones del mismo mes calendario. Los registros identificados como duplicados se agrupan y se contabilizan una sola vez.

¹ Contacto: contactocentral.bcentral.cl

² Esta nota técnica ha sido actualizada el 7 de octubre de 2026.

³ De acuerdo con información de carácter público difundida a través de sus sitios web y redes sociales, en agosto de 2026 dos portales firmaron un convenio técnico que permite la difusión cruzada de ofertas laborales entre ambas plataformas. Como resultado, las vacantes publicadas en un portal pueden replicarse en el otro, centralizando las postulaciones y ampliando el alcance de los procesos de selección.

1. Antecedentes

El IALI es una estadística experimental del Banco Central de Chile que aproxima la evolución de la demanda de trabajo a partir de avisos publicados en internet. Su desarrollo metodológico se presenta en Arraño y Jara (2019). El indicador se difunde mensualmente y está sujeto a revisiones asociadas a la actualización de su cobertura y metodología.

La metodología documentada en el citado estudio considera el promedio simple del conteo mensual de avisos únicos de cada portal, con una consolidación al interior de cada fuente, sin identificar duplicados entre portales.

La serie vigente corresponde a un índice con base 2018=100 y considera cuatro portales. El último de ellos fue incorporado en la publicación de mayo de 2026.

2. Procedimiento para detectar avisos duplicados

El procedimiento sigue la secuencia general de preparación, selección de candidatos, comparación y clasificación utilizada en el enlace de registros (Christen, 2012). Como antecedente, Eurostat documenta comparaciones de direcciones web, contenido y variables estructuradas en su procesamiento de avisos laborales (Eurostat, 2023b).

También se consideran las soluciones de la edición 2022–2023 del *Web Intelligence – Deduplication Challenge* (Eurostat, 2023a). Petridis *et al.* (2023) combinan representaciones vectoriales del texto, búsqueda de candidatos y reglas expertas; Palazzolo (2023) describe procedimientos de comparación y agrupación. Phelps (2024) resume las lecciones de estas experiencias. Las reglas aplicadas al IALI constituyen una adaptación local y requieren evaluación con sus propios datos.

2.1 Preparación de registros y selección de candidatos

Primero se normalizan los textos y los campos disponibles para reducir diferencias de formato, uso de mayúsculas, acentuación y puntuación. Se estandarizan las referencias geográficas y al empleador, conservando una distinción entre nombres informados y empresas no identificadas. Asimismo, se extraen atributos del puesto cuando están presentes, como jornada, horario, turno, modalidad de trabajo y requisitos de experiencia.

La comparación se realiza entre avisos publicados en el mismo mes calendario de los dos portales en cuestión. Se utilizan representaciones vectoriales del texto para identificar descripciones de contenido similar. El modelo pertenece a la familia *Qwen3 Embedding* (Zhang et al., 2025). Estas representaciones permiten identificar posibles coincidencias, pero no determinan por sí solas que dos avisos correspondan al mismo registro.

Luego, se seleccionan como candidatos los pares que superan el umbral de similitud establecido y no presentan incompatibilidades verificables mediante los filtros iniciales. Esta selección reduce el número de comparaciones que deben evaluarse en detalle.

2.2 Clasificación de pares

Para cada par candidato se combinan la similitud de las descripciones y los títulos con la concordancia del empleador, la ubicación y los atributos del puesto. En textos de distinta extensión, se considera si la descripción más breve está contenida en la más extensa, sin omitir las diferencias que puedan distinguir ofertas.

La información del empleador se trata de manera diferenciada según los nombres coincidan, difieran o no estén disponibles. Cuando esa información no permite establecer concordancia, se exige mayor evidencia en las restantes señales. La ausencia de un dato no se interpreta como coincidencia.

Se descartan las coincidencias que presenten incompatibilidades relevantes de ubicación, sucursal, jornada, turno o modalidad de trabajo. Los pares restantes se clasifican mediante reglas y umbrales ajustados con una muestra revisada manualmente.

2.3 Formación de grupos y conteo

Los pares clasificados como duplicados se utilizan para proponer grupos de avisos. La existencia de una cadena de coincidencias no garantiza que todos sus integrantes correspondan al mismo aviso: si A coincide con B y B con C, la unión de los tres requiere revisar la consistencia del conjunto. El uso de transitividad para agrupar registros puede propagar errores de clasificación (Hassanzadeh et al., 2009). Por ello, la formación de grupos incorpora controles sobre la compatibilidad geográfica, las sucursales y los códigos internos, además de un límite de tamaño. Cuando la incorporación de un registro vulnera esos controles, la unión no se acepta automáticamente.

Cada grupo aceptado se contabiliza una sola vez. Se conserva como representante el registro con la fecha de publicación más reciente y se le asigna el portal de ese registro, manteniendo la procedencia de todos los integrantes para fines de trazabilidad. Los avisos no agrupados conservan su conteo individual.

El alcance de este procedimiento comprende agosto 2026 y los meses siguientes y está limitado a los dos portales comparados. No identifica necesariamente todos los duplicados existentes en las fuentes del IALI, ni elimina republicaciones que ocurren en meses diferentes.

2.4 Calibración, evaluación y trazabilidad

Los umbrales de las reglas se calibraron mediante muestras de pares revisadas manualmente. A partir de una muestra aleatoria independiente utilizada para evaluación, el procedimiento alcanzó una precisión de 98,6%, es decir, que la gran mayoría de los pares clasificados como duplicados lo eran. Asimismo, se obtuvo una cobertura de 89,6%, lo que indica que se logró identificar aproximadamente nueve de cada diez avisos duplicados presentes en la muestra. La combinación de ambas métricas se resume en un estadístico $F1^4$ de 93,9%.

El procedimiento es determinista y reproducible. Los mismos datos y parámetros producen siempre la misma serie y, para cualquier par de avisos, puede reconstruirse qué señales se calcularon y qué regla determinó la clasificación final.

3. Revisión del indicador

El impacto de la eliminación de avisos duplicados entre los dos portales considerados se evalúa comparando el IALI publicado con la serie revisada. Cabe destacar que ambas series utilizan la misma cobertura de portales, período de comparación y tratamiento de la base. Por lo tanto, las diferencias observadas se explican exclusivamente por el procedimiento detallado en esta nota.

⁴ La puntuación $F1$ es una métrica de rendimiento fundamental en aprendizaje automático que combina la precisión y la exhaustividad en una única media armónica.

En agosto de 2026, el procedimiento identificó registros duplicados entre los portales comparados, equivalentes al 13,1% del total de avisos publicados por ambos portales ya depurados. De esta forma, el IALI revisado alcanza un índice (2018=100) equivalente a 68,3 puntos, frente a los 71,6 puntos de la serie publicada. La variación mensual pasa de +3,3% a -1,4%, mientras que la variación interanual cambia de -1,0% a -5,6% (Tabla 2).

Tabla 2. IALI agosto 2026, publicado y revisado.

	Índice (2018=100)	Var. Anual (%)	Var. Mensual (%)
Publicado	71,6	-1,0%	3,3%
Revisado	68,3	-5,6%	-1,4%

Fuente: Banco Central de Chile en base a avisos de los portales web incluidos en el IALI.

4. Consideraciones finales

El procedimiento detallado en esta nota busca reducir el efecto de las republicaciones entre portales sobre el IALI y mejorar la interpretación de sus movimientos. Su contribución debe evaluarse considerando los resultados de clasificación, el comportamiento de los grupos identificados y las diferencias entre las series publicadas y las recalculadas.

El procedimiento puede omitir duplicados o agrupar avisos distintos. Estas limitaciones requieren seguimiento, especialmente ante cambios en la forma de publicación de los portales o en la disponibilidad de atributos. La revisión periódica de muestras permitirá detectar deterioros del desempeño y modificar las reglas de cálculo.

La eliminación de avisos duplicados entre portales pasa a formar parte del cálculo regular del IALI. Una ampliación a otros portales requerirá de una nueva evaluación de desempeño y de sus efectos sobre el indicador.

5. Bibliografía

Arraño, E. y Jara, K. (2019). Índice de Avisos Laborales de Internet. Estudios Económicos Estadísticos, n.º 129. Banco Central de Chile. [Documento](#)

Banco Central de Chile (2024). Actualización del Índice de Avisos Laborales de Internet (IALI). Nota Técnica 2, División Estadísticas. Versión actualizada al 7 de octubre de 2024. [Nota](#)

Banco Central de Chile (s. f.). Índice de Avisos Laborales de Internet, base 2018 = 100. Base de Datos Estadísticos. Consulta: 2 de octubre de 2026. [BDE IALI](#)

Christen, P. (2012). *Data Matching: Concepts and Techniques for Record Linkage, Entity Resolution, and Duplicate Detection*. Springer. [DOI 10.1007/978-3-642-31164-2](https://doi.org/10.1007/978-3-642-31164-2)

Eurostat (2023a). *Web Intelligence – Deduplication Challenge. European Statistics Awards*, edición 2022–2023. [Reglas y descripción del concurso](#)

Eurostat (2023b). *Reference Metadata in Euro SDMX Metadata Structure (ESMS)*. WIHOJA_OJAES_Q_4D_2023_0004. Actualización del 11 de diciembre de 2023. [Metadatos WIH OJA](#)

Hassanzadeh, O., Chiang, F., Lee, H. C. y Miller, R. J. (2009). *Framework for Evaluating Clustering Algorithms in Duplicate Detection. Proceedings of the VLDB Endowment*, 2(1), 1282–1293. [DOI 10.14778/1687627.1687771](https://doi.org/10.14778/1687627.1687771)

Palazzolo, A. (2023). *Reproducibility Award Methodology Description – Team Nins. European Statistics Awards, Deduplication Challenge*. [Documento metodológico original](#)

Petridis, D., Pasch, S., Cutura, J. y Lagonidis, C. (2023). *Reproducibility Award Methodology Description. Equipo TheDeDuplicators, European Statistics Awards, Deduplication Challenge*. [Documento metodológico original](#)

Phelps, S. (2024). *Lessons learned from Eurostat's Deduplication Challenge*. WIN Blog, issue 23. Eurostat CROS. [Artículo institucional](#)

Zhang, Y., Li, M., Long, D., Zhang, X., Lin, H., Yang, B., Xie, P., Yang, A., Liu, D., Lin, J., Huang, F. y Zhou, J. (2025). *Qwen3 Embedding: Advancing Text Embedding and Reranking Through Foundation Models*. *arXiv:2506.05176*. [Versión 3 del artículo](#)