S

OMPI
ORGANIZACIÓN MUNDIAL
DE LA PROPIEDAD
INTELECTUAL

CWS/12/16 REV.
ORIGINAL: INGLÉS
FECHA: 20 DE SEPTIEMBRE DE 2024

Comité de Normas Técnicas de la OMPI (CWS)

Duodécima sesión
Ginebra, 16 a 19 de septiembre de 2024

PROPUESTA DE NUEVA NORMA TÉCNICA DE LA OMPI RELATIVA A LA DEPURACIÓN DE DATOS DE LOS NOMBRES

preparada por los corresponsables del Equipo Técnico de Normalización de Nombres

RESUMEN

1. El Equipo Técnico de Normalización de Nombres presenta un borrador final de nueva norma técnica de la OMPI relativa a la depuración de datos de los nombres, para su examen y aprobación en la duodécima sesión del Comité de Normas Técnicas de la OMPI (CWS).

ANTECEDENTES

2. En su undécima sesión de 2023, el CWS aprobó la descripción revisada de la tarea N.° 55:

> *"Preparar una propuesta de acciones futuras destinadas a lograr la normalización de los nombres en los documentos de propiedad intelectual (PI), con vistas a desarrollar una norma de la OMPI que ayude a las oficinas de PI a proporcionar una mejor "calidad en origen" en relación con los nombres."*
>
> (Párrafos 75 y 78 del documento CWS/11/28)

3. En el documento CWS/12/8 figura más información sobre la historia del Equipo Técnico y el avance realizado desde la última sesión del CWS.

4. En su undécima sesión de 2023, el CWS estudió nuevas directrices en materia de depuración de nombres de solicitantes presentadas por el Equipo Técnico de Normalización de Nombres. El CWS acordó utilizar el término "recomendaciones" en lugar de "directrices" en el nombre de la nueva norma propuesta, por considerarlo más claro en cuanto a su alcance. El

CWS también tomó nota de la propuesta de nombre realizada por la Secretaría: “Norma ST.93 de la OMPI” (párrafo 135 del documento CWS/11/28).

5. No obstante, el CWS no aprobó la norma propuesta y la devolvió al Equipo Técnico para que siguiera examinándola y mejorándola. Además, el CWS tomó nota de que la Secretaría estudiaría la posibilidad de publicar varios cuadros de transliteración en el sitio web de la OMPI (párrafos 136 y 137 del documento CWS/11/28).

PROPUESTA DE NUEVA NORMA TÉCNICA

6. Las Oficinas de PI experimentan problemas a la hora de identificar a las personas en una familia de patentes, ya que pueden utilizarse diferentes nombres de solicitantes dentro de la misma familia de patentes. Además, pueden cometerse errores ortográficos o tipográficos al introducir los nombres de los solicitantes. El deseo de disponer de datos depurados sobre los nombres de los solicitantes para fines estadísticos es bien conocido.

7. El Equipo Técnico de Normalización de Nombres, en el marco de la tarea N.° 55, ha preparado una propuesta final de nueva norma técnica de la OMPI relativa a la depuración de los nombres de los solicitantes. Esta propuesta figura en el Anexo del presente documento.

Objetivos

8. Estas recomendaciones tienen por objeto ofrecer pautas de carácter general. La variedad de factores, como los requisitos legales, las prácticas en materia de datos, la finalidad de la depuración, el uso previsto de los datos, las necesidades de recursos y las consideraciones técnicas, hacen que no exista un método único para todas las Oficinas de PI. Estas recomendaciones reflejan prácticas generales que pueden aplicarse en cualquier Oficina de PI a fin de depurar los datos de los nombres de los clientes, con lo que mejoran la normalización y la correspondencia de los nombres en las fases posteriores.

Alcance

9. La norma propuesta proporciona recomendaciones generales sobre la obtención, el procesamiento, la depuración y la publicación de datos de nombres. En esta norma no se ofrecen recomendaciones sobre los detalles en relación con los métodos de depuración de datos, adaptación o transformación de nombres, como la transliteración, la transcripción o la traducción, ni con los métodos de normalización de nombres, como la selección de algoritmos, dónde y cuándo se aplican las transformaciones, la frecuencia o las estrategias de fusión.

10. La propuesta de norma se estructura de la siguiente manera:

- Cuerpo principal: define las recomendaciones generales para el tratamiento de los nombres de los solicitantes con el fin de depurar los datos.

- Anexo: ofrece ejemplos de transliteración, transcripción y traducción en relación con las recomendaciones formuladas en el cuerpo principal.

11. Para la nueva norma de la OMPI se propone el nombre siguiente:

> “Norma ST.93 de la OMPI: Recomendaciones sobre la depuración de datos de los nombres”

Cambios realizados desde el último borrador

12. A la luz del debate sobre la propuesta de depuración de los datos de los nombres y de las aportaciones de varias delegaciones en la undécima sesión del CWS, el Equipo Técnico revisó

el borrador original de las directrices propuestas (Anexo del CWS/11/23). Se han introducido los siguientes cambios:

- El Equipo Técnico señala que la anterior definición de “datos depurados” como “exentos de errores y duplicaciones” era problemática porque no es realista garantizar que los datos estén al 100 % exentos de errores y duplicaciones. Por consiguiente, el Equipo Técnico acordó modificar la definición de “datos depurados” de la siguiente manera: “*datos precisos, coherentes y fiables. Puesto que el grado de depuración en un gran conjunto de datos complejos es difícil de medir, se pueden utilizar diversos parámetros como indicadores de la depuración o propiedades afines, como la idoneidad para un fin determinado.*”
- En el apartado “Transformación de nombres”, el Equipo Técnico convino en cambiar el término “conversión” por “transformación” para ajustarse mejor al título del apartado y permitir una interpretación más flexible.
- En el apartado “Referencias”, el Equipo Técnico analizó la posibilidad de hacer referencia a las normas ISO para la romanización de varias lenguas, tal como sugirió la Oficina Internacional. El Equipo Técnico llegó a la conclusión de que la norma propuesta debería incluir solo las normas pertinentes de la OMPI como planteamiento general, en lugar de incorporar las normas ISO pertinentes, ya que es posible que las Oficinas de PI no se atengan sistemáticamente a las normas ISO y que cambien sus prácticas con el tiempo.

13. En cuanto a los cuadros de transliteración utilizados por las Oficinas de PI, el Equipo Técnico señala que el objetivo principal es ofrecer un punto de referencia para mantener un debate razonable con los solicitantes, y no cambiar toda la base de datos en función de los cuadros de transliteración. Se pidió a las Oficinas del Equipo Técnico que facilitaran sus cuadros de transliteración, si estaban disponibles, para que los solicitantes, los mandatarios o las Oficinas de PI puedan remitirse a los cuadros de otras Oficinas de PI que utilicen lenguas diferentes a la hora de presentar nombres o de depurar datos de nombres. Se propone que el CWS pida a sus miembros que faciliten sus cuadros de transliteración. También se propone publicar los cuadros de transliteración proporcionados por las Oficinas de PI en la Parte 7 del Manual de la OMPI.

14. Si el CWS aprueba la nueva norma en la sesión en curso, se propone que el CWS pida a la Secretaría que publique esas recomendaciones en la Parte 3 del Manual de la OMPI.

15. Se invita al CWS a:

a) tomar nota del contenido del presente documento y del Anexo del presente documento;

b) pedir al Equipo Técnico de Normalización de Nombres que vuelva a examinar y mejore el proyecto de Norma ST.93 de la OMPI;

c) instar a las Oficinas y al sector de la PI a nominar especialistas para que formen parte del Equipo Técnico de Normalización de Nombres;

d) pedir a la Oficina Internacional que organice un taller

sobre la depuración de datos de los nombres al que puedan asistir todas las partes interesadas; y

e) pedir a sus miembros y observadores que presten apoyo a la Oficina Internacional promoviendo ese taller.

[Sigue el Anexo]

NORMA ST.93 DE LA OMPI

RECOMENDACIONES SOBRE LA DEPURACIÓN DE DATOS DE LOS NOMBRES

Propuesta presentada para ser aprobada por el Comité de Normas Técnicas de la OMPI (CWS) en su duodécima sesión

INTRODUCCIÓN 2
DEFINICIONES 2
OBTENCIÓN DE DATOS 2
TRANSFORMACIÓN DE NOMBRES 3
VALIDACIÓN Y DESAMBIGUACIÓN 3
MANTENIMIENTO 3
PUBLICACIÓN E INTERCAMBIO DE DATOS 4
FINES ESTADÍSTICOS 4
Referencias 4
ANEXO 5
Ejemplos de transliteración: 5
Ejemplos de transcripción: 6
Ejemplos de traducción: 6

NORMA ST.93 DE LA OMPI

RECOMENDACIONES SOBRE LA DEPURACIÓN DE DATOS DE LOS NOMBRES

Propuesta presentada para ser aprobada por el Comité de Normas Técnicas de la OMPI (CWS) en su duodécima sesión

INTRODUCCIÓN

1. Esta norma proporciona recomendaciones generales sobre la obtención, el procesamiento, la depuración y la publicación de datos de nombres. En esta norma no se ofrecen recomendaciones sobre los detalles en relación con los métodos de depuración de datos, adaptación o transformación de nombres, como la transliteración, la transcripción o la traducción, ni con los métodos de normalización de nombres, como la selección de algoritmos, dónde y cuándo se aplican las transformaciones, la frecuencia o las estrategias de fusión. Esas decisiones varían enormemente en función de la parte que las aplique, del objetivo de las transformaciones y de la rápida evolución de los algoritmos de correspondencia.

2. Cabe señalar que la Norma ST.20 de la OMPI ofrece recomendaciones para elaborar índices de documentos de patente en los que figuren nombres de solicitantes y otros clientes, y para fomentar la presentación uniforme de los nombres que aparecen en esos índices, así como un método uniforme de ordenación de los nombres en el propio índice.

DEFINICIONES

3. En el contexto de este documento:

a) Por “Oficina de PI” se entiende una Oficina de propiedad intelectual encargada de tramitar las solicitudes y registrar los derechos de PI.

b) Por “datos de clientes” se entienden los datos de solicitantes, titulares de registros, representantes legales u otras partes que obran en poder de una Oficina de PI en relación con un derecho de PI, una solicitud, un registro u otro instrumento. Esta norma se ocupa principalmente de los datos de los nombres de clientes, a saber: nombres de personas, nombres de empresas e información conexa, como la ciudad, la dirección o el correo electrónico, que se pueden utilizar para desambiguar posibles coincidencias de nombres.

c) Por “datos depurados” se entienden datos precisos, coherentes y fiables. Puesto que el grado de depuración en un gran conjunto de datos complejos es difícil de medir, se pueden utilizar diversos parámetros como indicadores de la depuración o propiedades afines, como la idoneidad para un fin determinado.

d) Por “transliteración” se entiende la correspondencia de los caracteres de la lengua de origen con los caracteres (fonéticos) de la lengua de destino.

e) Por “transcripción” se entiende la correspondencia de un carácter, logograma, sílaba o fonema de la lengua de partida con un sonido en el sistema de la lengua de llegada.

f) Por “traducción” se entiende la representación del significado de una palabra o concepto en la lengua de partida con algo que se corresponde con el significado en la lengua de llegada.

OBTENCIÓN DE DATOS

4. Las Oficinas de PI pueden ofrecer la posibilidad de que los clientes creen y gestionen registros electrónicos que contengan información publicada sobre los nombres, a saber: nombres de personas, nombres de empresas, nombres de representantes legales e información conexa, como la ciudad, la dirección o el correo electrónico.

5. Las Oficinas de PI deberán permitir que un registro de cliente se asocie a varias solicitudes o registros de derechos de PI, de modo que los clientes puedan reutilizar la misma información sobre los nombres para varias solicitudes o registros y actualizar esa información en un solo lugar.

6. Las Oficinas de PI pueden facilitar formularios para que los clientes pidan a las Oficinas de PI que creen o cambien su nombre o información conexa. Las Oficinas de PI también podrán permitir que los clientes introduzcan y actualicen su nombre o la información conexa por sí mismos o podrán exigir que sea una parte designada, por ejemplo empleados, contratistas o un servicio externo, la que introduzca y actualice los registros de los clientes a petición de estos.

7. Los distintos registros de un mismo cliente podrán ser creados y gestionados por entidades diferentes, por ejemplo por representantes legales diferentes. Las Oficinas de PI deben tener esto en cuenta a la hora de diseñar los sistemas de registro, ya que varios registros de un mismo cliente pueden tener ligeras variaciones de los mismos datos o ser actualizados en distintos momentos por distintos representantes.

8. Las Oficinas de PI pueden admitir la introducción del nombre del cliente en caracteres propios del idioma del cliente, además de hacerlo en los idiomas de trabajo de la Oficina; para ello se utilizará la codificación UTF-8[1]. Por ejemplo, una Oficina que trabaje en inglés podría disponer de distintos campos para el nombre del solicitante en inglés y el nombre original en coreano.

9. Las Oficinas de PI podrán utilizar opcionalmente números para identificar a los clientes. Esos números los podrán crear las Oficinas o podrán obtenerlos de una fuente externa, como un número de registro de una empresa o un número de pasaporte. Los números de identificación por sí solos no resuelven los problemas que se plantean con los datos depurados de los clientes, como las entradas duplicadas, los cambios de nombre y la información obsoleta o incorrecta. Las Oficinas que utilicen números de identificación deberán seguir prestando atención a las consideraciones contenidas en otras partes de esta norma.

TRANSFORMACIÓN DE NOMBRES

10. Para el intercambio y procesamiento de datos, incluida la recepción de solicitudes o registros internacionales, las Oficinas de PI pueden considerar la posibilidad de transformar los nombres (Anexo de este documento). Se recomienda que las Oficinas de PI envíen y reciban los datos de los nombres utilizando la codificación UTF-8.

11. Debe tenerse en cuenta que la adaptación o conversión de los nombres de los clientes se presta a numerosos errores, ya que no existen normas generalmente aceptadas o uniformes. En esta norma se mencionan tres formas de adaptación o transformación de nombres: transliteración, transcripción y traducción. Si las Oficinas de PI transliteran, transcriben o traducen caracteres de una lengua (como el griego) a otra (como el inglés), deberán publicar su sistema de transliteración, transcripción o traducción. El documento transliterado, transcrito o traducido, o partes del mismo, deben ponerse a disposición del cliente, quien deberá disponer de un medio para presentar correcciones si la transliteración, transcripción o traducción presenta fallos.

12. Si es posible, debe evitarse la transliteración inversa; en su lugar, se recomienda utilizar el nombre original. Por ejemplo, una solicitud presentada por “Phony Corp” se podría transliterar en caracteres griegos como “Φονι Κορπ” en el sistema de una Oficina y, al publicarla, la transliteración inversa del griego a los caracteres latinos podría ser “Foni Corp”, lo que daría lugar a discrepancias. En el Anexo de esta norma figuran ejemplos de los problemas más comunes que surgen en la transliteración, transcripción o traducción inversa o doble.

VALIDACIÓN Y DESAMBIGUACIÓN

13. Los métodos de validación y desambiguación deben diseñarse para cumplir objetivos específicos, ya sean administrativos o estadísticos, y deben aplicarse procedimientos adecuados en función de esos objetivos. Los métodos de concordancia y desambiguación de nombres deben tener el alcance adecuado y someterse a una evaluación de riesgos en función del objetivo previsto a fin de lograr un nivel adecuado de desambiguación.

14. Las Oficinas de PI podrán optar por validar la información presentada por los clientes, incluso mediante comprobaciones automatizadas. Los resultados de la validación se pondrán a disposición de los clientes y estos aceptarán las correcciones de ser necesario, lo que incluye fórmulas para eludir un mecanismo de validación automatizado en caso de que surjan resultados incorrectos o incompletos.

15. Para desambiguar los registros de nombres (es decir, encontrar entradas duplicadas), las Oficinas podrán examinar aspectos distintos de los nombres de los clientes. Los nombres no son intrínsecamente únicos. Por ejemplo, puede haber varias personas llamadas “John Smith” o varias empresas llamadas “Data Corp”. La comparación de otros datos como la ciudad, el código postal o la fecha de nacimiento, cuando estén disponibles, puede aumentar la calidad de las correspondencias.

16. Cualquier proceso de validación o desambiguación iniciado por la Oficina de PI que pueda tener efectos legales, como la corrección o la normalización del nombre del titular registrado de un derecho de PI, deberá ser confirmado por el cliente antes de que se formalice el cambio en el sistema de la Oficina.

MANTENIMIENTO

17. Las Oficinas de PI deben elaborar una estrategia para depurar periódicamente las bases de datos de nombres de clientes que incluya la detección y resolución de los registros duplicados, es decir, varios registros de una misma entidad. En algunos casos, los registros duplicados se pueden fusionar o combinar, por ejemplo, los registros con pequeñas diferencias ortográficas involuntarias, como “ABC Corp.” y “ABC Corp”. En otros casos puede ser preferible mantener registros independientes. Cada Oficina deberá decidir qué enfoque se ajusta mejor a su propio sistema de gestión de los registros de nombres. La estrategia puede incluir la participación de los clientes interesados en la depuración de los registros y la responsabilidad de los datos depurados.

18. Las Oficinas de PI deberán proporcionar un mecanismo que permita a los clientes actualizar la información sobre su nombre en varias solicitudes o derechos de PI introduciendo la información una sola vez. Esto se puede conseguir, por ejemplo, asociando cada solicitud o derecho de PI a un único registro de cliente que contenga información sobre el

[1] UTF-8 es un sistema de codificación de Unicode.

nombre, o permitiendo a los clientes seleccionar varias solicitudes o derechos de PI y presentar un solo conjunto de datos actualizados para que se apliquen a todos ellos.

19. Las Oficinas de PI podrán designar a una persona responsable de la depuración de datos, lo que incluye el desarrollo de parámetros para medir los datos depurados, el seguimiento y la presentación de informes periódicos sobre dichos parámetros y la adopción de medidas para mejorar los datos de los clientes cuando sea necesario.

PUBLICACIÓN E INTERCAMBIO DE DATOS

20. Las Oficinas de PI deberán dar a conocer las actualizaciones de la información sobre los nombres que se realicen después de la publicación de un derecho de PI. Por ejemplo, si “ABC Corp” cambia su nombre por “XYZ Corp” en el registro de clientes, en las publicaciones en línea se deberá asociar el nombre “XYZ Corp” al derecho de PI. El nombre original también podrá aparecer en el derecho de PI publicado, de acuerdo con los requisitos legales de la Oficina.

21. Si una Oficina tiene varias formas del nombre de un cliente, como el nombre original expresado con caracteres nativos, esas variaciones deben incluirse en los datos publicados y en los datos intercambiados con otras Oficinas.

22. Si una Oficina utiliza números de identificación para distinguir las entidades, estos se deberán incluir en los datos que publique y en los que intercambie con otras Oficinas de PI. Si los números de identificación son confidenciales y no pueden compartirse, la Oficina deberá indicar qué datos de los clientes utilizan esos números de identificación, por ejemplo, sustituyéndolos por números únicos generados para su publicación.

FINES ESTADÍSTICOS

23. Con fines estadísticos, las Oficinas de PI podrán intentar cotejar los datos de los clientes con variaciones en los nombres u otros campos para obtener recuentos más precisos. En esos casos, las Oficinas deberán publicar su estrategia o algoritmo de correspondencia junto con los resultados estadísticos para que otras personas puedan entender la metodología utilizada.

REFERENCIAS

24. Las referencias a la siguiente norma son pertinentes para la siguiente norma:
Norma ST.20 de la OMPI Preparación de índices de nombres propios de los documentos de patente

[Sigue el Anexo a la norma propuesta].

ANEXO

DIFERENTES MÉTODOS DE TRANSFORMACIÓN DE NOMBRES

Aunque la transliteración y la transcripción son conceptos diferentes desde una perspectiva lingüística, el resultado suele ser muy similar para los sistemas de escritura basados en caracteres. Sin embargo, la transcripción proporciona un resultado más práctico, ya que para la conversión solo se necesitan los caracteres estándar de la lengua de destino.

El inglés es un idioma que se adopta como lengua común entre hablantes cuyas lenguas maternas son diferentes, si bien suele pasarse por alto que la transcripción entre dos lenguas rara vez está estandarizada. En el mejor de los casos existen definiciones oficiales para [xx] -> [en] que llevan a suponer que [xx] -> [en] -> [yy] es igual a [xx] -> [yy], lo que no suele ser correcto.

EJEMPLOS DE TRANSLITERACIÓN:[2]

En la figura 1 se muestra un ejemplo de correspondencia de las letras y observaciones relativas a la transliteración.

Source and Target words	Letter Correspondence				Description
English to Persian					
John /dʒɒn/	J	o	h	n	*h* is a silent letter (no sound
	|	|		|	is associated to the letter)
جان /dʒɒn/	ج	ا		ن	and is not transliterated
Arabic to English					
نجيب /nædʒiːb/	ن	ج	ی	ب	short vowel /æ/ on N is
	|	|	|	|	normally not written in
Najib /nædʒiːb/	Na	j	i	b	Arabic script
English to Japanese					
Bill /biːl/	B	i	l	l	each syllable in Japanese is a
	\	/	\	/	consonant-vowel sequence
ビル [bi-ru]	ビ		ル		
English to Hindi					
Adam /ˈædəm/	A	d	a	m	the second "a" is not
	|	|		|	transliterated in Hindi
अदम /ˈædəm/	अ	द		म	

Figura 1: Ejemplo de transliteración

[2] Encuesta sobre transliteración automática https://www.researchgate.net/figure/Transliteration-examples-in-four-language-pairs-Letter-correspondence-shows-how-the_fig1_220566444

EJEMPLOS DE TRANSCRIPCIÓN:

A continuación se muestran ejemplos en los que la transcripción puede dar lugar a imprecisiones:
[ru]: Ш → [de]: sch[3]

[ru]: Ш → [en]: sh

[ko]: ㅑ → [de]: ja[4]

[ko]: ㅑ → [en]: ya

[gr] : Ω → latín: O[5]

[da]: Æ → [de]: Ä o AE, [en]: AE [6]

EJEMPLOS DE TRADUCCIÓN:

En el primer ejemplo, está claro que la traducción directa puede dar lugar a problemas:
[de]: Aktiengesellschaft → [en]: corporation, stock co, …

[ru]: ОАО Силовы́е маши́ны → [en] : OJSC “Power Machines” - O bien - [en]: Open Joint-stock Company “Power Machines”

La figura 2 muestra casos límite típicos de romanización del nombre de una empresa china:

- [zh]: 北京东土科技股份有限公司 → [en] transliteración (pinyin): běi jīng dōng tǔ kē jì gǔ fèn yǒu xiàn gōng sī ;
- [zh]: 北京东土科技股份有限公司 → [en] transcripción (pinyin): beijing dongtu keji gufen youxian gongsi
- [zh]: 北京东土科技股份有限公司 → [en] traducción (inglés): Beijing, China Science and Technology Joint-stock Limited Company
- [zh]: 北京东土科技股份有限公司 → en realidad: Kyland Technology Co., Ltd.

(71) 申请人：北京东土科技股份有限公司(**KYLAND TECHNOLOGY CO., LTD**) [CN/CN]；中国北京市石景山区实兴大街30号院2号楼8层901, Beijing 100041 (CN)。

Figura 2: Romanización del nombre chino de una empresa

[Fin del Anexo a la norma propuesta y de la norma]

[Fin del Anexo y del documento]

[3] https://de.wikipedia.org/wiki/Kyrillisches_Alphabet#Russisch
[4] https://de.wikipedia.org/wiki/Koreanisches_Alphabet
[5] https://en.wikipedia.org/wiki/Romanization_of_Greek
[6] https://en.wikipedia.org/wiki/Dania_transcription