

Comité de Normas Técnicas de la OMPI (CWS)

Undécima sesión
Ginebra, 4 a 8 de diciembre de 2023

PROPUESTA DE DIRECTRICES PARA LA DEPURACIÓN DE NOMBRES

Documento preparado por los corresponsables del Equipo Técnico de Normalización de Nombres

ANTECEDENTES

1. En la quinta sesión del Comité de Normas Técnicas de la OMPI (CWS), celebrada en 2017, se añadió una tarea al programa de trabajo del CWS: la tarea n.º 55. En la misma sesión se creó el Equipo Técnico de Normalización de Nombres, al que se asignó la tarea n.º 55, cuya descripción era:

"Estudiar la posibilidad de crear una norma técnica de la OMPI para ayudar a las Oficinas de PI a brindar una mejor 'calidad en el origen' en relación con los nombres de los solicitantes, preparar una propuesta para la adopción de medidas destinadas a la normalización de los nombres de los solicitantes en documentos de PI y someterla al examen del CWS."

(Párrafos 85 a 88 del documento CWS/5/22)

2. En su séptima sesión de 2019, el CWS tomó nota de que la Secretaría había distribuido en noviembre de 2018 una encuesta sobre el uso de identificadores para los solicitantes por parte de las Oficinas de PI, tal y como se acordó en la sexta sesión del CWS. Puesto que la encuesta sobre el uso de identificadores estaba terminada, el Equipo Técnico propuso revisar la descripción de la tarea n.º 55 haciendo referencia a esa encuesta. El CWS aprobó la descripción revisada de la tarea n.º 55, que ahora es:

"Estudiar la posibilidad de crear una norma técnica de la OMPI para ayudar a las Oficinas de PI a brindar una mejor 'calidad en el origen' en relación con el nombre de los solicitantes, preparar una propuesta para la adopción de medidas destinadas a la normalización de los nombres de los solicitantes en documentos de PI y someterla al examen del CWS."

(Párrafos 72 a 79 del documento CWS/7/29)

3. La Oficina Internacional organizó un taller en 2019 para tratar cuestiones relacionadas con la normalización de nombres. Se invitó a oradores tanto de las Oficinas como del sector privado para que compartieran prácticas y posibles soluciones. En ese taller, la Oficina Internacional observó que las Oficinas de PI aplicaban métodos muy diferentes a la depuración de datos para la normalización de nombres.

4. En el documento CWS/11/22 se exponen más detalles sobre el avance realizado por el Equipo Técnico en su plan de trabajo desde la última sesión del CWS.

DIRECTRICES PROPUESTAS

5. El Equipo Técnico de Normalización de Nombres, en el marco de la tarea n.º 55, preparó una propuesta final de directrices para ayudar a las Oficinas de PI a depurar los datos de los nombres. Esas directrices figuran en el Anexo del presente documento.

6. Las directrices se prepararon para ofrecer recomendaciones generales y de alto nivel. 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 pretenden plasmar las prácticas generales que pueden aplicarse en cualquier Oficina de PI para depurar los datos de los nombres de los clientes, lo que a su vez favorece la normalización y las técnicas de correspondencia de nombres en las fases posteriores.

7. El Equipo Técnico señala 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. Por consiguiente, las directrices finalizadas incluirán una referencia a esa norma y recomendarán su uso.

8. Si el CWS las aprueba en la presente sesión, se propone que pida a la Secretaría que publique las directrices en la parte 3 del Manual de la OMPI:

https://www.wipo.int/standards/es/part_03_standards.html.

9. *Se invita al CWS a:*
- a) *tomar nota del contenido del presente documento y del Anexo del presente documento;*
 - b) *examinar y aprobar las directrices mencionadas en los párrafos 5 a 7 anteriores y reproducidas en el Anexo del presente documento; y*
 - c) *pedir a la Secretaría que publique las directrices en la parte 3 del Manual de la OMPI, como se menciona en el párrafo 8 anterior.*

[Sigue el Anexo]

DIRECTRICES PARA LA DEPURACIÓN DE NOMBRES

*Propuesta presentada para su aprobación por el Comité de Normas Técnicas de la OMPI (CWS)
en su undécima 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 | 1 |
| Ejemplos de transliteración:..... | 1 |
| Ejemplos de transcripción:..... | 2 |
| Ejemplos de traducción: | 2 |

DIRECTRICES PARA LA DEPURACIÓN DE NOMBRES

*Propuesta presentada para su aprobación por el Comité de Normas Técnicas de la OMPI (CWS)
en su undécima sesión*

INTRODUCCIÓN

1. Estas directrices se refieren a las consideraciones generales para la obtención, el procesamiento, la depuración y la publicación de datos de nombres. No se tratan las numerosas y complejas cuestiones que plantean la depuración de datos, la adaptación o transformación de nombres, como la transliteración, la transcripción o la traducción, ni la normalización de nombres, como la selección de algoritmos, dónde y cuándo se aplican las transformaciones, la frecuencia o las estrategias de fusión. Estas 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 conexas, 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, libres de errores y duplicaciones. 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 conexas, 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 conexas. Las Oficinas de PI también podrán permitir que los clientes introduzcan y actualicen su nombre o la información conexas 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-8¹. Por ejemplo, una Oficina que trabaje en inglés podría admitir 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 muchos de 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 estas directrices.

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 estas directrices se mencionan tres formas de adaptación o conversión 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 "Φωνι 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 estas directrices 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.

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

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 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 las cuestiones relacionadas con los datos depurados, 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 estas directrices:

Norma ST.20 de la OMPI Preparación de índices de nombres propios de los documentos de patente

[Sigue el Anexo]

ANEXO

DIFERENTES MÉTODOS DE TRANSFORMACIÓN DE NOMBRES

Propuesta presentada para su aprobación por el Comité de Normas Técnicas de la OMPI (CWS) en su undécima sesión

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.

Siendo el inglés la lengua franca de la economía mundial(izada), se suele pasar por alto que la transcripción rara vez está estandarizada entre dos lenguas distintas. 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 | <i>h</i> is a silent letter (no sound is associated to the letter) and is not transliterated |
| جان /dʒɒn/ | ج ا ن | |
| Arabic to English | | |
| نجيب /nædʒiːb/ | ن ج ي ب | short vowel /æ/ on N is normally not written in Arabic script |
| Najib /nædʒiːb/ | Na j i b | |
| English to Japanese | | |
| Bill /bi:l/ | B i l l | each syllable in Japanese is a consonant-vowel sequence |
| ビル [bi-ru] | ビ ル | |
| English to Hindi | | |
| Adam /ædəm/ | A d a m | the second "a" is not transliterated in Hindi |
| अदम /ædəm/ | 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]: sch³

[ru]: Ш → [en]: sh

[ko]: ㅏ → [de]: ja⁴

[ko]: ㅏ → [en]: ya

[gr] : Ω → latín: O⁵

[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 Силовые машины → [en] : OJSC "Power Machines" - O bien - [en]: Open Joint-stock Company "Power Machines"

La figura 2 muestra un típico caso límite 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 y de las directrices]

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