Documento de especificaciones funcionales del convertidor DOCX2XML

#### VERSIÓN 1.1

#### Borrador de trabajo para consulta

Índice

[1. Introducción 2](#_Toc212715399)

[1.1. Objetivo 2](#_Toc212715400)

[1.2. Alcance del proyecto 2](#_Toc212715401)

[1.3. Hipótesis, recomendaciones y limitaciones. 3](#_Toc212715402)

[2. Resumen de los requisitos 4](#_Toc212715403)

[2.1. Caso práctico 4](#_Toc212715404)

[3. Diseño detallado del programa 5](#_Toc212715405)

[3.1. Recursos XML 5](#_Toc212715406)

[3.2. Sistema de archivo y convertidor: ejemplo de diseño detallado 6](#_Toc212715407)

[3.3. Gestión de excepciones e informes de errores 6](#_Toc212715408)

[4. Referencias 7](#_Toc212715409)

[ANEXO I: Detalles de la conversión DOCX2XML de las oficinas de PI 8](#_Toc212715410)

[ANEXO II: Encabezados de sección para solicitudes de patente en formato DOCX 9](#_Toc212715411)

# Introducción

En su décima sesión, celebrada en noviembre de 2022, el Comité de Normas Técnicas de la OMPI (CWS) aprobó que el Equipo Técnico de la Transformación Digital (DTTF) explorara la conversión de DOCX a XML como opción para las oficinas (véanse los párrafos 105 y 106 del documento CWS/10/22). Según los resultados de la encuesta del DTTF comentados en el documento CWS/10/15, casi la mitad de los encuestados afirmaron que aceptaban archivos DOCX. En su undécima sesión, celebrada en diciembre de 2023, el CWS tomó nota de que el DTTF había llevado a cabo una evaluación y un análisis de la funcionalidad de los convertidores de DOCX a XML utilizados en la Oficina de Patentes y Marcas de Estados Unidos y en la Oficina Internacional del ePCT. En la misma sesión, el CWS aprobó que el DTTF elaborara una propuesta de recomendación sobre una especificación de requisitos comunes para un convertidor DOCX a XML (véanse los párrafos 56 a 58 del documento CWS/11/28). Sería muy beneficioso para las oficinas de PI utilizar una herramienta de conversión con una funcionalidad comúnmente acordada para apoyar los esfuerzos de aceptación y procesamiento de la información de las solicitudes de patentes en formato de texto estructurado, avanzando así hacia el objetivo común de utilizar la norma ST.36 o ST.96 de la OMPI como formato XML estándar. Este documento y su Anexo I consolidan y resumen los comentarios proporcionados por las oficinas de PI sobre la función del convertidor DOCX utilizado en sus oficinas. Cabe destacar que más de la mitad de los encuestados indicaron que facilitan información inicial de validación basada en el contenido a los solicitantes en el momento en que presentan la documentación, antes de su recepción. En el Anexo I se incluye una lista de los convertidores DOCX existentes, en la que se indican las coherencias y diferencias como referencia.

## Objetivo

El objetivo de este documento es proporcionar una descripción de los requisitos funcionales de un convertidor DOCX2XML que transformará las solicitudes de patente presentadas en formato DOCX a XML. El convertidor debe generar una salida conforme a las normas ST.36 o ST.96 de la OMPI. El formato DOCX se basa en el estándar de archivo de procesamiento de textos Office Open XML (OOXML) y está regulado por las normas internacionales ECMA‑376 e ISO/IEC 29500.

## Alcance del proyecto

Los usuarios deberían poder crear una nueva presentación que contenga uno o más documentos DOCX relacionados a través del sistema de archivo. Si un documento enviado contiene errores o avisos de validación DOCX, el sistema debería proporcionar al usuario la información necesaria para que pueda volver a enviarlo. Entonces, el sistema de archivo puede generar y enviar un archivo DOCX corregido al convertidor DOCX2XML para su procesamiento y conversión. Tras la conversión, el sistema debería enviar un mensaje al revisor de la oficina de PI indicando si la conversión ha sido correcta o no. El mensaje de error se refiere al incumplimiento de las normas operativas, que se produce antes del proceso de conversión propiamente dicho. Además, en este documento también se explica cómo deben convertirse a XML contenidos como cuadros y ecuaciones matemáticas, tal como se indica en la sección 3.1 siguiente.

Nota: DOCX corregido se refiere a la producción de una copia en formato DOCX (que se utilizará para la transformación a ST.36 o ST.96 XML) a partir de la copia original, aplicando todas las reglas pertinentes y eliminando cualquier metadato, como el autor, las observaciones y los marcadores, y haciendo explícita la numeración de las listas del documento original.

### Tipos de documentos de entrada admitidos

Los documentos producidos por editores que tienen la capacidad de guardar o exportar el contenido en formato DOCX son los siguientes:

* Microsoft Word 2007 o superior (para PC o Mac);
* Google Docs;
* Office Online;
* LibreOffice; y
* Pages para Mac.

## Hipótesis, recomendaciones y limitaciones.

* Es posible que el proceso de conversión no pueda transformar todas las características de DOCX al formato de la norma ST.36 o ST.96 de la OMPI y que solo admita una gama muy limitada de formatos básicos. Cabe señalar que esto no incluye los tipos de letra, el color o el tamaño de la letra, ni ningún posicionamiento específico de los elementos de texto en la página, como los tabuladores. No obstante, existe un soporte limitado para ello a través de cuadros, el uso de imágenes en lugar de texto y la etiqueta <pre>.
* Este documento no especifica ninguna solución concreta de sistema de gestión de contenidos.
* El archivo DOCX debe contener los encabezados de sección especificados en el Anexo II.

# Resumen de los requisitos

A diagram of a software process

AI-generated content may be incorrect.

El **diagrama anterior** muestra el flujo de trabajo de un sistema de conversión de archivos DOCX, que se describe con más detalle a continuación.

Cuando el usuario cargue el documento DOCX mediante un sistema de archivo, el convertidor DOCX2XML procesará la conversión. Los siguientes pasos ilustran el flujo de trabajo:

1. El sistema de archivo comprueba que el documento DOCX cumpla las normas de la oficina de PI y muestra los errores o avisos, si los hubiera.
2. El archivo DOCX corregido generado por el sistema se proporcionará al convertidor DOCX2XML.
3. Si no se detectan errores ni avisos durante el proceso de validación, el servicio convertidor transforma el documento DOCX corregido en XML y este debe validarse con los esquemas de las normas ST.36 o ST.96 de la OMPI.

## Caso práctico

El siguiente caso práctico describe cómo el sistema de archivo procesa el documento DOCX. El sistema puede dividir el documento DOCX en varios documentos de sección basándose en el encabezado de cada una (reivindicaciones, descripción, resumen y dibujos). Véanse los títulos de las secciones en el Anexo II del presente documento.

### Flujo de procesamiento de documentos DOCX

1. El usuario carga el archivo DOCX como contenido de la solicitud en un sistema de archivo.
2. El sistema realiza validaciones (nombre del archivo, extensión, etc.).
3. El sistema puede convertir el archivo DOCX original en un archivo DOCX corregido (siempre que se cumplan las normas operativas).
4. Se envía al solicitante una respuesta sobre si la corrección del archivo DOCX es correcta o incorrecta.
5. El sistema almacena el archivo DOCX original (antes de la corrección).
6. El DOCX corregido se convierte en un XML conforme a la norma ST.36 o ST.96 mediante el convertidor DOCX2XML.
7. Se envía al revisor de la OPI la información sobre si la conversión XML ST.36 o ST.96 ha sido correcta o incorrecta.

# Diseño detallado del programa

## Recursos XML

### Correspondencia de los principales elementos del formato DOCX con las normas ST.36 o ST.96

| **Elemento DOCX** | **Equivalente XML** |
| --- | --- |
| Cuadros | Cuadros (OASIS CALS) |
| Ecuaciones matemáticas | Los objetos de ecuaciones matemáticas integradas se convertirán a formato MathML y los demás se convertirán a imagen. |
| Objetos ChemDraw (es posible que algunas oficinas de PI no admitan ChemDraw). | Convertido a imagen |
| Dibujos, SmartArt, gráficos y formas (es posible que algunas oficinas de PI no lo admitan). | Convertido a imagen |
| Imágenes JPEG, TIFF, PNG, GIF, EMF, WMF y SVG (es posible que algunas oficinas de PI no admitan el formato SVG) | Convertido a imagen |

## Sistema de archivo y convertidor: ejemplo de diseño detallado

Diagrama del proceso de conversión y sistema de archivo DOCX2XML

A diagram of a process flow

AI-generated content may be incorrect.

## Gestión de excepciones e informes de errores

A continuación, se muestra un cuadro con los errores y avisos que pueden producirse durante el procesamiento de archivos DOCX.

**Errores y avisos del sistema de presentación de documentos DOCX y del servicio de conversión:**

|  |  |  |  |
| --- | --- | --- | --- |
| **Categoría del mensaje** | **Ejemplos** | **DOCX corregido** | **XML generado** |
| Error: nivel de sistema | Virus o protegido por contraseña | No | No |
| Error: nivel de contenido | Faltan partes, tipo de letra no válido | No | No |
| Aviso: sin corregir | Resumen demasiado largo | Sí | Sí |
| Aviso: corregido | Eliminar comentarios, marcar cambios | Sí | Sí |
| Error: sin conversión a XML de ST.36 o ST.96 | Formato de imagen no compatible, errores de validación | Sí | No |
| Sin problemas | La mayoría de los casos | Sí | Sí |

# Referencias

ISO/IEC 29500 <https://www.iso.org/standard/71691.html>

ECMA-376 <https://ecma-international.org/publications-and-standards/standards/ecma-376/>

Norma ST.36 de la OMPI [https://www.wipo.int/documents/d/standards/docs-es-03-36-01.pdf](https://www.wipo.int/documents/d/standards/docs-en-03-36-01.pdf)

Norma ST.96 de la OMPI [https://www.wipo.int/documents/d/standards/docs-es-03-96-01.pdf](https://www.wipo.int/documents/d/standards/docs-en-03-96-01.pdf)

# ANEXO I: Detalles de la conversión DOCX2XML de las oficinas de PI

Los detalles de la conversión DOCX2XML (XSLX) de las oficinas de PI están disponibles en: [https://www.wipo.int/edocs/mdocs/cws/es/cws\_13/cws\_13\_12-annexii.xlsx](https://www.wipo.int/edocs/mdocs/cws/en/cws_13/cws_13_12-annexii.xlsx)

# ANEXO II: Encabezados de sección para solicitudes de patente en formato DOCX

El Anexo II contiene ejemplos de encabezados de sección que sirven de guía para la presentación de solicitudes basadas en las normas ST.36 o ST.96 de la OMPI. Estos encabezados ayudarán al convertidor a detectar automáticamente las distintas secciones de la solicitud, como la descripción, las reivindicaciones, el resumen y los dibujos. Se puede utilizar un encabezado genérico para los que no figuren en la siguiente lista.

|  |  |
| --- | --- |
| **Elemento** | **Título de sección** |
| 1 | Campo técnico |
| 2 | Estado de la técnica |
| 3 | Divulgación |
| 4 | Resumen de la invención |
| 5 | Descripción de los dibujos |
| 6 | Descripción de la realización |
| 7 | Mejor realización |
| 8 | Realización de la invención |
| 9 | Aplicabilidad industrial |
| 10 | Referencias |
| 11 | Material biológico depositado |
| 12 | Lista de secuencias |
| 13 | Citas |
| 14 | Reivindicaciones |
| 15 | Resumen |
| 16 | Dibujos |

[Fin del Anexo y del documento]