



Universidad Nacional de San Juan  
Facultad de Ciencias Exactas, Físicas y Naturales

## **Diferencias entre los formatos .doc, .docx, .djvu, .pdf y .epub**

### **Alumnos:**

- Jofré Santiago, Registro: 21260.
- Benemérito Ignacio, Registro: 21036.
- Sebastián García, Registro: 21434

**Asignatura:** Teoría de la Información.

### **Equipo de cátedra:**

- Mag. Ing. Raúl Oscar Klenzi.
- Lic. Manuel Oscar Ortega.
- Lic. Fabrizio Amaya.

# Introducción

En este informe se tratarán las características y las diferencias entre los formatos PDF, .doc, .docx, .djvu y .epub teniendo en cuenta: su concepto, su capacidad de edición, como se componen, su compatibilidad, y por último, analizando la forma en que comprimen la información cada uno de estos formatos, pudiendo observar como varía su tamaño.

Se detallarán algunas de las diferentes herramientas de software (tanto herramientas Online, como aplicaciones de escritorio) de conversión de los formatos presentados (por ejemplo: una aplicación que permita convertir archivos PDF a .docx y viceversa)

# Índice

## 1 Formatos.

### 1.1 El formato PDF

#### 1.1.1 Concepto

#### 1.1.2 Características

##### 1.1.2.1 Composición

##### 1.1.2.2 Compatibilidad

##### 1.1.2.3 Edición

##### 1.1.2.4 Tamaño y compresión

### 1.2 El formato .doc y .docx

#### 1.2.1 Concepto

#### 1.2.2 Características

##### 1.2.2.1 Composición

##### 1.2.2.2 Compatibilidad

##### 1.2.2.3 Edición

##### 1.2.2.4 Tamaño y compresión

### 1.3 El formato .djvu

#### 1.3.1 Concepto

#### 1.3.2 Características

##### 1.3.2.1 Compatibilidad

##### 1.3.2.2 Edición

##### 1.3.2.3 Tamaño y compresión

### 1.4 El formato .epub

#### 1.4.1 Concepto

#### 1.4.2 Características

##### 1.4.2.1 Composición

##### 1.4.2.2 Compatibilidad

##### 1.4.2.3 Edición

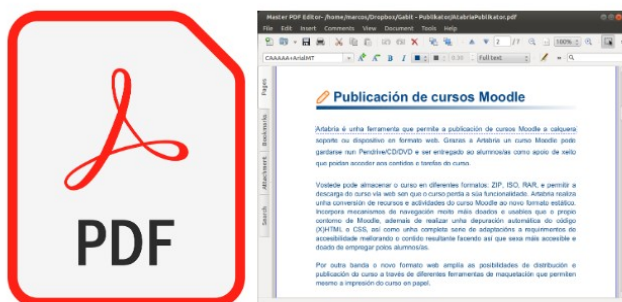
##### 1.4.2.4 Tamaño y compresión

## 2 Diferentes herramientas de software de conversión

## 3 Conclusiones

# 1. Formatos

## 1.1 El formato PDF



### 1.1.1 Concepto:

PDF (Portable Document Format) es el formato de archivo creado por Adobe. El formato PDF puede contener textos, imágenes y otros elementos, y se utiliza para mostrar contenido de texto formateado. Sin embargo, no es totalmente compatible con la edición de contenidos. Los PDF suelen crearse a partir de otros formatos de texto para transformar los archivos en elementos más compatibles universalmente. [1]

### 1.1.2 Características

#### 1.1.2.1 Composición

El contenido de PDF es una composición de gráficos vectoriales y de mapa de bits y de textos en una forma estable y portátil, lo que significa que puede abrirse con cualquier software compatible sin que se produzcan pérdidas de contenido.

#### 1.1.2.2 Compatibilidad

El formato PDF se caracteriza por su compatibilidad con muchas más aplicaciones, como navegadores web, aplicaciones de mensajería y, por supuesto, visores especiales de PDF como Adobe Acrobat, PDFelement, Foxit Reader y muchos otros. [2]

### **1.1.2.3 Edición**

Las opciones de edición en un archivo PDF son más limitadas si se comparase con un formato .docx, esto es debido a que el ciclo de uso del formato PDF es más corto. Está pensado para compartir archivos listos para ver y revisar, y a veces para archivar los acuerdos y contratos firmados.

Sin embargo, ofrece algunas opciones de edición como: anotaciones, comentarios, relleno de formularios, dibujos y adición de objetos visuales, con diferentes iteraciones y extensiones de los mismos.[2]

### **1.1.2.4 Tamaño y compresión**

El tamaño de un archivo PDF puede variar según:

- Si presenta imágenes, y el tamaño de las mismas
- Si posee contenido interactivo y multimedia, como vídeos, anotaciones, hipervínculos, entre otros.
- Si posee o no fuentes embebidas.

En general, el tamaño de un archivo PDF es menor si se lo compara con el archivo .docx, ya que los archivos PDF se comprimen para reducir su tamaño. Hay varias técnicas de compresión, como la compresión de imágenes (por ejemplo, JPEG para imágenes y JBIG2 para texto). De esta forma, al estar comprimida la información, la entropía de un fichero PDF será mayor (tendiendo al valor 8), reduciendo así la redundancia.

## **1.2 El formato .doc y .docx**



### **1.2.1 Concepto**

El formato .doc es un formato de documento de texto abierto inventado originalmente por Microsoft que se utiliza para crear archivos de texto editables, se introdujo en las primeras versiones de Microsoft Word y se utilizó hasta la versión Office 2003. El formato .docx fue introducido con Microsoft Word 2007 y se convirtió en el formato predeterminado para todas las versiones posteriores de Word. El "x" en .docx significa "XML", lo que refleja la estructura basada en XML

del archivo. Ambos ficheros son compatibles con imágenes, formas, gráficos, ecuaciones y muchos otros tipos de objetos.

## **1.2.2 Características**

### **1.2.2.1 Composición**

Los archivos .doc utilizan un formato binario propietario, lo que significa que los datos se almacenan en un formato que no es fácilmente legible por otros programas sin un conocimiento específico del formato.

En cambio, .docx es una colección de archivos XML que contienen diferentes tipos de contenidos comprimidos en un archivo. Requiere un software específico con capacidad para leer correctamente los contenidos XML y sus dependencias. DOCX puede abrirse con aplicaciones de procesamiento de documentos como Microsoft Office, ONLYOFFICE, LibreOffice y otras, que tienen un soporte variado para las características y contenidos del formato.

### **1.2.2.2 Compatibilidad**

Debido a la compatibilidad variable entre diferentes programas y a las especificaciones de los sistemas de los usuarios, el abrir archivos .doc y docx puede dar resultados diferentes con pérdidas de contenido. Por eso, la elección del software adecuado para trabajar con DOCX es una cuestión de gran importancia.

Los archivos .doc y .docx son ampliamente compatibles con muchas versiones de Microsoft Word (2007 y posteriores) y otros procesadores de texto que soportan el estándar Open XML. Además, debido a que se basa en XML, la estructura del archivo .docx es más fácilmente interpretable por otros programas y herramientas de software. El principal inconveniente es que si se abre un archivo .doc con versiones más recientes de Microsoft Word u otros, puede ocasionar problemas de compatibilidad, pudiendo producir pérdida de información

### **1.2.2.3 Edición**

A comparación del formato PDF, los archivos .doc y .docx tienen más posibilidad de edición: permiten aplicar formatos complejos a los textos, como fuentes y estilos, diseños de página y elementos. Admite la inserción de imágenes, formas, gráficos, ecuaciones y muchos otros tipos de objetos.

Son colaborativos por diseño. Otros usuarios pueden editar los documentos creados cuando los abren en un software compatible con DOCX, ya sea de forma independiente o en modo compartido en la nube. Los colaboradores tienen las mismas opciones para editar y dar formato al contenido del documento, a menos que el archivo se comparta con restricciones añadidas.

#### **1.2.2.4 Tamaño y compresión**

Los archivos .docx suelen ser más pequeños que los archivos .doc debido a la compresión incorporada que utiliza el formato ZIP. Esto resulta en un uso más eficiente del espacio y tiempos de carga más rápidos. Los archivos .doc carecen de técnicas avanzadas de compresión sobre el formato binario.

Sin embargo, si se lo compara con el formato PDF, en términos generales, el tamaño de los archivos .doc y .docx será mayor, debido principalmente a las técnicas de compresión utilizados en los distintos formatos, y también por la naturaleza de PDF, que está diseñado para ser un archivo portable y con pocas posibilidades de edición.

### **1.3 El formato Djvu**



#### **1.3.1 Concepto**

DjVu es un formato de archivo diseñado principalmente para almacenar imágenes escaneadas. Se caracteriza por incorporar avanzadas tecnologías tales como separación de capas de imágenes, carga progresiva, codificación aritmética y compresión con pérdida para imágenes bitonales (dos colores), permitiendo que imágenes de alta calidad se almacenen en un mínimo de espacio.[1]

#### **1.3.2 Características**

##### **1.3.2.1 Compatibilidad**

En términos de compatibilidad, DjVu no es tan común como PDF u otros formatos. Sin embargo, hay numerosos visores y herramientas compatibles en varias plataformas, incluidos Windows, macOS, Linux y dispositivos móviles.

##### **1.3.2.2 Edición**

Los ficheros Djvu son más limitados a la hora de edición, en comparación con otros formtos, ya que está diseñado principalmente para la visualización y la compresión eficiente de documentos escaneados, no para la edición directa. Sin embargo, existen algunas herramientas y métodos que permiten realizar ciertas modificaciones en archivos:

- Edición Básica de Texto e Imágenes
- Reorganización de Páginas

- Añadir o Eliminar Páginas

### **1.3.2.3 Tamaño y compresión**

DjVu utiliza técnicas avanzadas de compresión para reducir el tamaño de los archivos, especialmente para imágenes y texto escaneado. Esto lo hace ideal para la distribución de libros y documentos digitales.

A pesar de la alta compresión, DjVu conserva una calidad de imagen excelente, lo cual es útil para documentos que contienen tanto texto como gráficos o fotografías en color.

Algunas técnicas de compresión que utiliza este formato son:

- Compresión JB2 (Joint Bi-level Image Group 2).
- Compresión IW44 (Wavelet Compression for Continuous Tone Images)
- Compresión ZP (Zone Partitioning Compression).
- Compresión Textual con OCR (Optical Character Recognition).
- Compresión con Separación de Capas (Layered Compression).
- Compresión Delta

## **1.4 El formato .epub**



### **1.4.1 Concepto**

“EPUB” es una abreviatura o acrónimo y significa “Publicación Electrónica”. Los archivos .EPUB son la extensión de formato de archivo más popular utilizada para el almacenamiento de libros electrónicos y otros contenidos relacionados.



El formato .EPUB puede almacenar palabras, imágenes, fuentes, hojas de estilo, metadatos y tablas de contenido. Funcionan como contenedores zip en los que los distintos datos se agrupan en archivos específicos. Utiliza varios estándares web abiertos como XHTML, XML y CSS. [4] [5]

## **1.4.2 Características**

### **1.4.2.1 Composición**

Como se mencionó anteriormente, Un archivo EPUB es esencialmente un archivo comprimido en formato ZIP que contiene varios tipos de archivos, como:

- Archivos de contenido en XHTML o HTML5.
- Archivos de estilo CSS para el diseño y la presentación del texto.
- Metadatos (como información del autor, título, etc.).
- Imágenes y otros recursos multimedia.
- Un archivo de tabla de contenidos (TOC) que organiza la navegación dentro del libro.

### **1.4.2.2 Compatibilidad**

Los archivos EPUB están basados en XHTML o HTML5, lo que significa que se pueden ver en una variedad de dispositivos con la apariencia correcta. Es compatible con la mayoría de los lectores de libros electrónicos (eReaders), tablets, smartphones, y aplicaciones de lectura en computadoras.

### **1.4.2.3 Edición**

A diferencia de los formatos mencionados anteriormente, los archivos .epub presentan menos limitación a la hora de editar. Como se mencionó previamente, un archivo .epub contiene archivos XHTML, HTML5, hojas de estilo CSS, entre otros recursos. Por lo cual, si se desea editar un archivo .epub, se deberán editar estos archivos.

La principal desventaja en este aspecto, es que si se quiere tener un control completo para la edición de estos archivos, el usuario que desea editar este tipo de ficheros deberá tener conocimiento básico sobre como funciona HTML, CSS y XML. Sin embargo, existen herramientas que facilitan la edición, como Sigil, y Calibre.

### **1.4.2.4 Tamaño y compresión**

Un archivo EPUB es esencialmente un archivo ZIP que contiene todos los componentes necesarios para mostrar un libro electrónico. La compresión se utiliza para reducir el tamaño del archivo final, facilitando su distribución y descarga. La compresión en EPUB afecta principalmente a los siguientes elementos:

- **Archivos de contenido** (XHTML/HTML): El texto del libro, que es el componente más voluminoso en la mayoría de los casos, se comprime para reducir el tamaño total.
- **Archivos CSS:** Hojas de estilo que definen la apariencia del contenido.  
Imágenes y otros recursos multimedia: Archivos JPEG, PNG, GIF, videos, y audio también están comprimidos para optimizar el tamaño del EPUB.
- **Archivos XML:** Metadatos y tablas de contenido que también se benefician de la compresión.

La compresión en los archivos EPUB sigue el estándar ZIP. Esto significa que utiliza la compresión ZIP Deflate: el método de compresión más común utilizado en los archivos ZIP y, por extensión, en los archivos EPUB. El algoritmo Deflate combina el enfoque de compresión sin pérdida de LZ77 con la codificación de Huffman. Esto reduce el tamaño del archivo sin perder información, lo cual es crucial para mantener la integridad del contenido del libro electrónico.

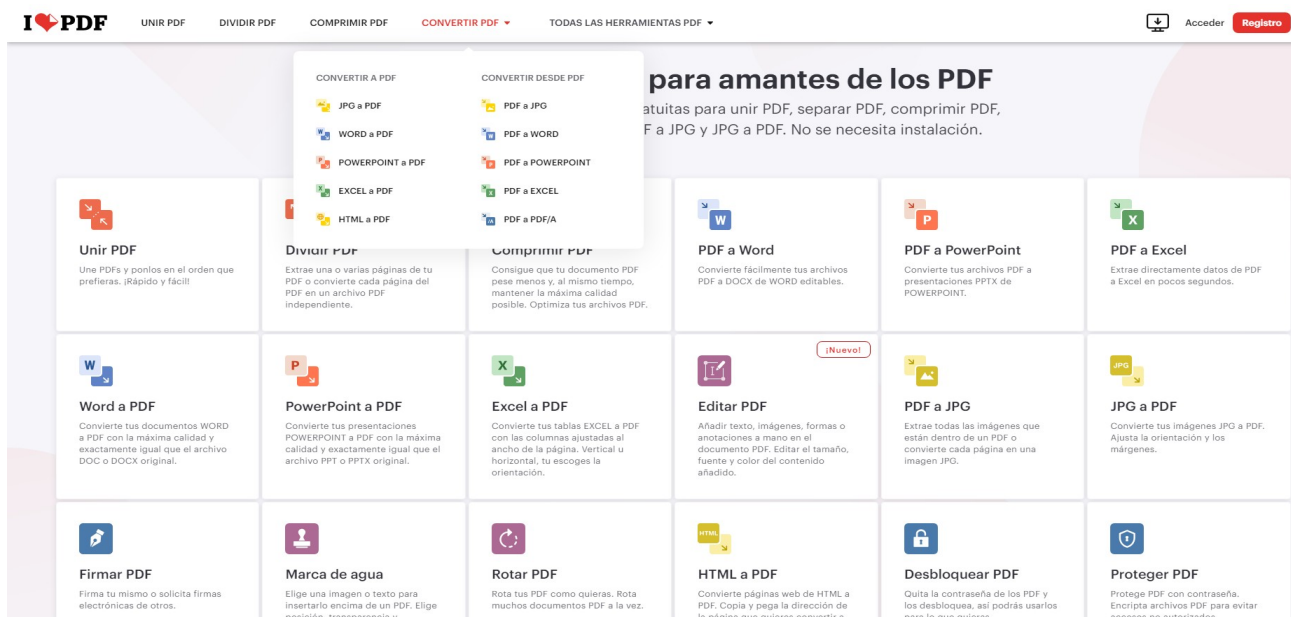
## 2. Diferentes herramientas de software de conversión

A continuación se detallan algunas aplicaciones web y de escritorio que permiten la conversión entre diferentes ficheros o formatos:

### iLovePDF

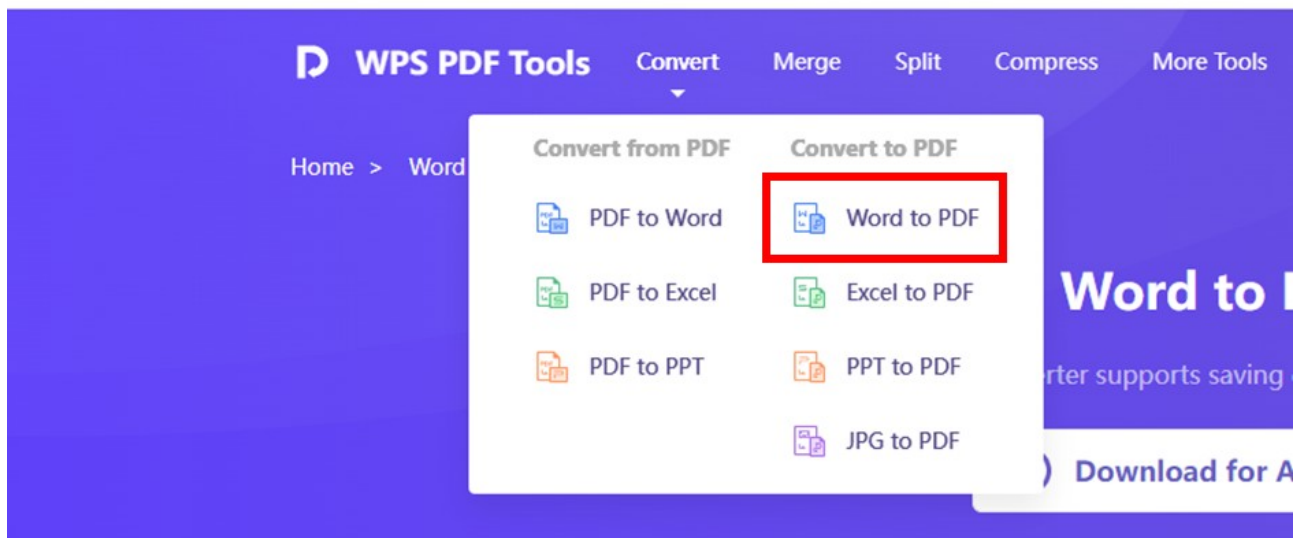


iLovePDF es una herramienta online y completamente gratuita que permite unir ficheros PDF, separarlos, comprimirlos y convertir documentos Office a PDF, PDF a JPG y JPG a PDF, entre otras funcionalidades. No se requiere de instalación para su utilización. [6]



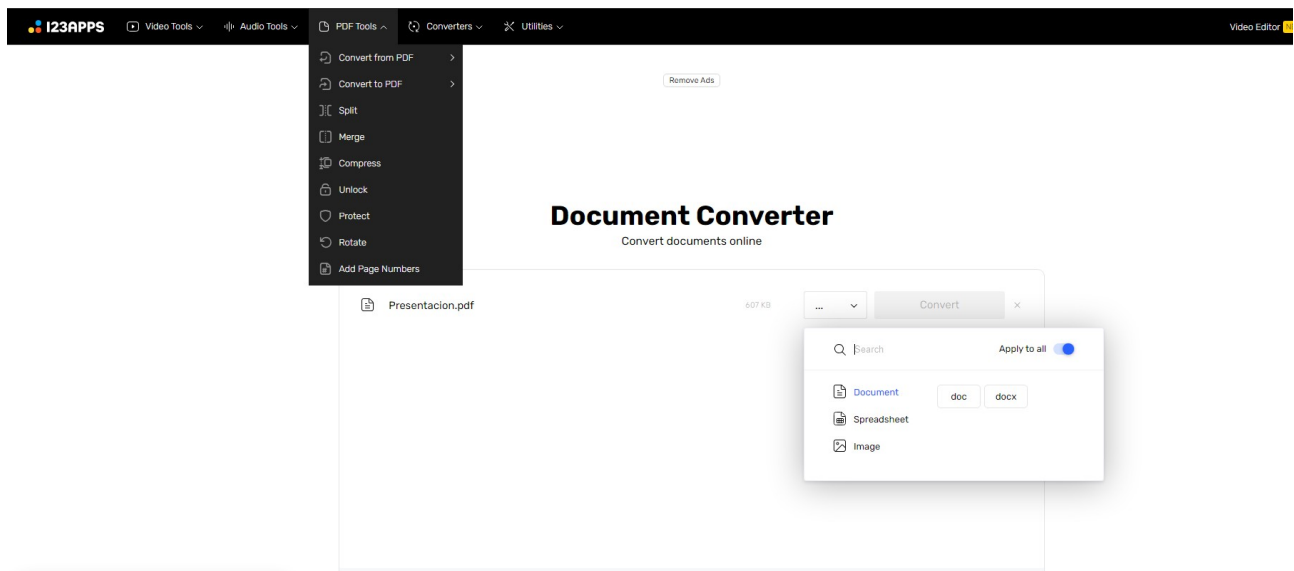
## WPS Office

WPS Office es una herramienta que permite trabajar fácilmente con archivos PDF y DOCX, facilitando la conversión entre estos formatos y otros. [1]



## Convert.io

Al igual que iLovePDF, Convert.io es una herramienta web que permite la conversión entre diferentes archivos, inclusive entre formatos distintos a los mencionados (por ejemplo: admite conversión entre formatos de audio, imágenes, video, etc). También ofrece servicios similares a los de iLovePDF [7]



Djvu2pdf.com

# DjVu to PDF

Djvu2pdf es una herramienta online que permite la conversión de ficheros .djvu a .pdf de forma sencilla y rápida, sin la necesidad de instalación de algún programas. [8]

## DjVu to PDF

Español ▼

Convierta sus archivos DjVu a archivos PDF en línea gratuitamente de forma rápida y sencilla sin tener que instalar ningún programa.

- 1 Haga clic en el botón SUBIR y seleccione hasta un máximo de 20 archivos DjVu para convertir.
- 2 Espere que el proceso de Conversión finalice y descargue los archivos por separado, usando las miniaturas o agrupados en un archivo ZIP.



## Epub2pdf.com

Tiene un funcionamiento similar a la herramienta anterior, solo que aplicada a los ficheros .epub.  
[9]



## 3. Conclusiones

En conclusión, se pudo observar que cada fichero se diferencia esencialmente en las técnicas de compresión que utiliza para reducir su espacio, haciendo que un tipo de fichero tenga menor tamaño que otro (como ocurre con el formato .docx respecto del .pdf, donde en casos generales el .docx tiene mayor tamaño a un formato .pdf) .

También se dejó en evidencia que algunos formatos presentaban limitaciones a la hora de editarlos. Tal como en el caso de los ficheros .doc, .docx y .epub, que presentaban mas opciones y posibilidades de edición en comparación con los formatos .pdf y .djvu

Por último, se dejó en evidencia que en algunos ficheros, como los PDF y .epub, utilizan técnicas más avanzadas de compresión, lo que provoca que su tamaño sea mucho menor que los otros ficheros que emplean técnicas más simples de compresión (como en el caso del formato .doc).

Bibliografía consultada:

<https://es.wps.com/articulos-de-word/diferencias-entre-documentos-docx-y-pdf-una-guia-esencial/> [1]

<https://www.onlyoffice.com/blog/es/2022/07/pdf-vs-docx> [2]

<https://es.wikipedia.org/wiki/DjVu> [3]

<https://www.adobe.com/es/acrobat/resources/document-files/ebook-files/epub.html> [4]

<https://recoverit.wondershare.es/file-recovery/what-is-epub-file.html> [5]

<https://www.ilovepdf.com/es> [6]

<https://convert.io/document-converter> [7]

<https://djvu2pdf.com/es/> [8]

<https://epubtopdf.com/es/> [9]