Ejercicio 6

Genere un informe que marque las diferencias esenciales (tipo de documento, si es un formato comprimido o no, tamaños, calidades, etc.) entre los siguientes documentos de texto \*.doc, \*.docx, \*.djvu, \*.pdf, \*.epub. Encuentre una aplicación que a un archivo de texto en cualquier formato pueda pasarlo a la mayoría de los restantes.

## \*.doc

***.doc*** es un formato para los documentos de texto creados por Microsoft para Microsoft Word.

Desde las primeras versiones de Word de los años 80, el formato por defecto siempre fue el ***.doc***, cuyo nombre viene de la palabra documento.

Por la amplia difusión del Microsoft Word, este formato se ha convertido en estándar de facto con el que pueden transferirse textos con formato o sin formato, o hasta imágenes, siendo preferido por muchos usuarios. Permite texto enriquecido, tablas, gráficos y macros, pero es menos eficiente en la representación fiel en distintos dispositivos

Este formato posee la desventaja de tener un mayor tamaño al compararlo con algunos otros formatos ya que no realiza ninguna compresión.

## \*.docx

Los ***.docx*** son archivos de texto de que con la publicación de la versión 2007 de Word, el formato se introdujo como nuevo estándar para documentos de texto.

El nombre ***docx*** es una combinación de la ya conocida extensión ***.doc*** y la X de XML.

El objetivo era establecer un formato de archivo basado en XML que necesitara menos capacidad de almacenamiento, pero mantuviera las características de positivas de un .doc. El resultado es un formato que mantiene texto enriquecido, estilos, tablas, imágenes, multimedia y fórmulas matemáticas, posee mayor compatibilidad y portabilidad. Y consume menos espacio.

Un archivo ***.docx*** se comporta como un archivo ZIP que almacena el contenido (texto e imágenes) en datos XML y CSS y, a continuación, los comprime. Al abrir el archivo ***.docx***, el contenido se vuelve a descomprimir.

En conjunto, tanto el proceso de almacenamiento, como el de apertura, de un archivo ***.docx*** es más rápido y ocupa considerablemente menos espacio por archivo que un archivo ***.doc***.

## \*.djvu

***.djvu***, pronunciado como déjà vu, es un formato de archivo informático diseñado principalmente para almacenar imágenes escaneadas.

Se caracteriza por incorporar avanzadas tecnologías tales como:

* Separación de capas de imágenes
* Carga progresiva
* Codificación aritmética
* Compresión con pérdida para imágenes bitonales (dos colores)

Esto permite que imágenes de alta calidad se almacenen en un mínimo de espacio.

***.djvu*** ha sido promovido como una alternativa al .pdf, y en la actualidad supera a este formato en la mayoría de los documentos escaneados.

## \*.pdf

***.pdf*** es un formato de almacenamiento para documentos digitales independiente de plataformas de software o hardware.

Su nombre viene de las siglas de **P**ortable **D**ocument **F**ormat (formato de documento portátil).

Puede contener cualquier combinación de texto, elementos multimedia como vídeos o sonido, elementos de hipertexto como vínculos y marcadores, enlaces y miniaturas de páginas.

Una de sus mejores características es que los archivos ***.pdf*** no pierden el formato con el envío a otros usuarios, como sí sucede cuando se envían documentos de texto (se desordenan las páginas, se desorganizan los párrafos, etc.).

Es el estándar ISO (ISO 19005-1:2005) para ficheros contenedores de documentos electrónicos con vistas a su preservación de larga duración.

Los archivos PDF se pueden comprimir y cada elemento del mismo es comprimido mediante uno u otro algoritmo. Los textos y órdenes de formato PostScript se pueden comprimir usando el algoritmo Lempel Ziv Welch (LZW) y las imágenes mediante JPEG, ZIP o RLE.

## \*.epub

***.epub*** es un formato de archivo de libro electrónico.

El término es un acrónimo de la expresión inglesa **e**lectronic **pub**lication (publicación electrónica).

***.epub*** es un estándar técnico publicado por el International Digital Publishing Forum (IDPF) sustituyendo al Open eBook.

El formato ***.epub*** es implementado como un archivo que contiene archivos XHTML, junto con imágenes y otros archivos de apoyo.

Una publicación ***.epub*** se entrega en un solo archivo. Este archivo es un archivo comprimido sin cifrar que contiene un conjunto de recursos interrelacionados.

Un archivo ***.epub*** es un archivo que contiene, en efecto, un sitio web. Incluye archivos HTML, imágenes, hojas de estilo CSS y otros activos. También contiene metadatos.

El contenedor del ***.epub*** debe contener:

* Al menos un documento de contenido.
* Un documento de navegación.
* Un documento de paquete que enumera todos los recursos de la publicación.

## Convertir archivos

Una pagina interesante que se encontró para convertir distintos tipos de archivos de texto es [convertio.co](https://convertio.co/es/djvu-pdf/)



