Informe de las diferencias de los archivos de documentos de texto

Universidad Nacional de San Juan



Facultad de Ciencias Exactas, Físicas y Naturales



Licenciatura en Ciencias de la Computación

Teoría de la Información

2025

Docentes:

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

Integrantes:

* Pereyra Nicolas, Registro: E010-97
* Ledesma David, Registro: E010-100

# Documento de texto \*.doc

Un archivo doc es un documento de texto creado por Microsoft en el año 1983, utilizado para documentos de procesamientos de textos almacenados en el formato de archivo binario Microsoft Word. Estos documentos pueden contener texto formateado, imágenes, gráficos y diagramas, una vez guardados se pueden ajustar y editar tantas veces como uno quiera. Debido a esto, mientras más elementos contengan estos tipos de documentos su tamaño va ir incrementando. El formato de archivo doc es propiedad de Microsoft, por lo que otras aplicaciones de procesamiento de texto pueden tener dificultades para leerlo

Debido a que es tipo de archivo muy antiguo (básicamente se dejó de implementar a finales del 2006), su seguridad no es la mejor de todas, ya que en aquellos tiempos el cifrado que se utilizaba era muy débil, además de que contenían macros lo que los hacía muy vulnerable a virus y malwares dañinos.

A pesar de haber quedado en desuso, es posible convertirlo en otros tipos de documentos de texto como pdf, docx e incluso odt permitiendo así que la extensión .doc no quedara obsoleta del todo.

# Documento de texto \*.docx

Los archivos docx son el equivalente moderno de los archivos doc, una actualización de estos últimos tipos de archivos. Fueron introducidos en Microsoft Word 2007 como parte del estándar Office Open XML (OOXML). La aplicación comprime toda la información del documento en una carpeta zip, dentro de esta se encuentran múltiples ficheros XML más los recursos (como imágenes, estilos, etc.) organizados en carpetas. Esto permite que estos tipos de archivos sean en promedio entre 2 y 10 veces más liviano que un doc. En pocas palabras, el formato de archivo docx puede contener más datos sin ralentizar su máquina, lo que hace que sea mucho más fácil compartir o almacenar estos datos sin ocupar mucho espacio.

Lo usual es trabajar este tipo de archivos con Word, pero también se puede acceder a archivos docx en Google Drive (Google Docs), Apple Pages y LibreOffice. Docx también admite una amplia gama de funciones, incluidos estilos, imágenes, tablas, encabezados, pies de página, hipervínculos y configuraciones de diseño avanzadas. Sin embargo, los usuarios de Mac pueden tener dificultades para acceder a los archivos docx. Además, el formato no siempre se muestra correctamente. Por esto se recomienda convertir el archivo docx a pdf, txt, odt e incluso html según sea la necesidad de uno.

Respecto a doc, desde el punto de vista de la seguridad, es mucho más seguro frente a ataques de recuperación de claves, menos vulnerable a virus y posee algoritmos de cifrados modernos como AES, SHA-1, entre otros.

# Documento de texto \*.djvu

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](https://es.wikipedia.org/wiki/Algoritmo_de_compresi%C3%B3n_con_p%C3%A9rdida) para imágenes bitonales (dos colores), permitiendo que imágenes de alta calidad se almacenen en un mínimo de espacio. Fue creado por AT&T Labs en los años 90 para ser una alternativa más ligera que PDF para documentos escaneados. Este formato de archivo utiliza un sistema de comprensión mixta muy avanzado: JB2 para texto en blanco y negro, IW44 para imágenes a color y Zlib para capas adicionales, esto permite reducir drásticamente el tamaño manteniendo la legibilidad.

Por ejemplo, las páginas escaneadas a 300 DPI a todo color se pueden comprimir hasta archivos de 30 a 100 KB desde 25 MB, mientras que las páginas en blanco y negro rondan los 5 y 30 KB cuando se comprimen. Mientras que un libro escaneado en pdf puede ocupar entre 50 y 100 MB, con djvu puede pesar entre 5 y 10 MB, ¡¡o sea se reduce el tamaño hasta 10 veces menos!!

Para poder visualizar un archivo de este tipo es necesario aplicaciones específicas, como WinDjView, DjView u Okular. Aunque si es posible convertirlo a pdf, png o jpg (si se exportan páginas como imágenes). Y con respecto a la seguridad, es prácticamente nula, no incluye de forma nativa sistemas avanzados de cifrado ni protección por contraseña como pdf, pero si se pueden aplicar medidas externas.

# Documento de texto \*.pdf

Pdf es una abreviatura que significa formato de documento portátil (Portable Document Format, en inglés). Es un formato versátil creado por Adobe en 1993 y que ofrece una forma sencilla y segura de presentar e intercambiar documentos con independencia del software, el hardware o el sistema operativo que utilice quien los consulte.

Pdf puede estar comprimido, usando diferentes algoritmos como zip, jpeg para imágenes o ccitt g4 para documentos escaneados en blanco y negro. El nivel de compresión va depender de cómo se genere el pdf, si es texto editable entonces será muy liviano, pero si es escaneado como imagen entonces será mucho más pesado. Por esto su tamaño variará mucho dependiendo del tipo, para el primer caso puede pesar menos de 100 KB mientras que para el segundo caso puede pesar alrededor de 50 y 100 MB.

Respecto a la seguridad, es posible utilizar contraseñas de apertura y edición, soporta cifrado AES de 128 y 256 bits, control de permisos, firmas digitales, integración con sistemas de gestión de derechos digitales (DRM). Sin embargo, la seguridad depende de cómo se configure

# Documento de texto \*.epub

Epub es un formato de archivo de libro electrónico, es compatible con muchos lectores de libros electrónicos como Kobo, Nook, Sony Reader, PocketBook, etc. Este formato puede almacenar palabras, imágenes fuentes, hojas de estilo, metadatos y tablas. Desde un punto de vista técnico, los archivos epub son contenedores zip en los que los distintos datos se agrupan en archivos específicos. Dentro de estos contenedores hay archivos XHTML, HTM, XML, CSS e imágenes (como JGP, PNG, SVG).

Con respecto a su tamaño es muy liviano en comparación con un pdf, un libro promedio de 300 páginas de texto puede pesar en 200 KB y 1 MB. Si incluye imágenes, ilustraciones u otro tipo de contenido multimedia, el tamaño aumenta, pero sigue siendo más eficiente que un pdf escaneado. Es posible convertirlo en pdf, aunque pierde adaptabilidad, o docx si se quiere volver a texto editable.

Y para finalizar, de forma nativa, no trae grandes mecanismos de seguridad, aunque, puede incorporar DRM cuando es distribuido por editoriales o tiendas, limitando así la copia, impresión o apertura en dispositivos no autorizados. Sin DRM es un formato abierto y fácilmente editable.

# Resumen

A continuación, se presentará una tabla donde se puede apreciar con mayor facilidad las diferencias esenciales entre los diferentes tipos de documentos de texto:

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Formato | Tipo de documento | Comprimido | Tamaño | Calidades | Compatibilidad | Conversión | Seguridad |
| .doc | Documento binario | No | Pesado | Buenas, pero problemas en software que no sea de Microsoft | Word | pdf, docx, odt | Cifrado débil |
| .docx | Basado en XML | ZIP | Ligero, hasta 10 veces menor que doc | Excelente, soporta funciones modernas | Universal en Word, Google Docs, LibreOffice, Apple Pages | pdf, odt, txt, html, entre otros | Cifrado moderno, menos riesgo que doc |
| .djvu | Documento escaneado | JB2, IW44, zlib | Muy liviano (hasta 10 veces menos que pdf para el caso de libros) | Buena para escaneos | WinDjView, DjView, Okular | pdf, jpg, png | Nula (de forma nativa) |
| .pdf | Documento portátil | ZIP, JPEG, CCITT, entre otros | Variable, desde KB (puro texto) hasta MB (escaneos) | Excelente fidelidad | Universal | docx, txt, pptx, excel, jpg | Cifrado AES, contraseñas, firmas digitales, DRM, permisos |
| .epub | Libro electrónico | ZIP | Muy liviano, hasta 1 MB para 300 páginas | Excelente para texto | Kobo, Nook, Sony Reader, PocketBook, e-readers en general | pdf, docx | Nativo sin seguridad fuerte |

# Bibliografía

<https://www.adobe.com/es/acrobat/resources/document-files/text-files/doc-file.html>

<https://es.wikipedia.org/wiki/Doc_(inform%C3%A1tica)>

<https://www.adobe.com/acrobat/resources/document-files/doc-vs-docx.html>

<https://www.adobe.com/acrobat/resources/document-files/docx.html>

<http://djvu.org/resources/whatisdjvu.php>

<https://es.wikipedia.org/wiki/DjVu>

<https://www.adobe.com/ar/acrobat/about-adobe-pdf.html>

<https://es.wikipedia.org/wiki/EPUB>

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