

2. Documentos PDF

1. Introducción

PDF es un formato de archivo para representar documentos de una forma independiente de la aplicación software, hardware o sistema operativo que se haya usado para su creación, así como del dispositivo de salida por el que se muestra o imprime. Un documento PDF consiste en una colección de objetos que juntos, describen la apariencia de una o más páginas, posiblemente acompañados por elementos interactivos adicionales y datos de aplicación de alto nivel. Un archivo PDF contiene los objetos que componen un documento PDF, así como la información estructural asociada, todo ello representado como una única secuencia autocontenida de bytes.

Las páginas de un documento (y otros elementos visuales) pueden contener cualquier combinación de texto, gráficos e imágenes. La apariencia de una página se describe mediante un flujo de contenidos PDF, que contiene una secuencia de objetos gráficos para que sean representados en la pantalla.

Además de describir la apariencia estática de la página, un documento PDF puede contener elementos interactivos que son sólo posibles en una representación electrónica. PDF soporta *anotaciones* (annotations en lenguaje PDF) de muchos tipos, como notas de texto, enlaces de hipertexto, marcas, ficheros adjuntos, sonidos y películas. Un documento puede definir su propia interfaz de usuario; las entradas de ratón y teclado pueden activar ciertas acciones que están especificadas en los objetos PDF. Un documento también puede contener campos de un formulario interactivo para que sean rellenados por el usuario, y puede importar o exportar los valores de esos campos a otras aplicaciones (veremos más adelante que estas características no se pueden incluir en un documento para enviar al IEEE Xplore).

2. Componentes de un archivo PDF

Para conocer la estructura de un documento PDF nos remitiremos a la referencia de Adobe, en la que podemos comprobar como un documento PDF se divide en cuatro partes:

- Objetos: Un documento PDF es una estructura de datos compuesta por un pequeño conjunto de tipos básicos de objetos de datos.
- Estructura de Archivo: La estructura del archivo PDF determina como se almacenan los objetos en un archivo PDF, como se accede a ellos , y como se actualizan. Esta estructura es independiente de la semántica del objeto.
- Estructura de Documento: La estructura de un documento PDF especifica como se usan los tipos básicos de objetos para representar los componentes de un documento PDF, es decir, páginas, anotaciones, fuentes, etc.