

**UNIVERSIDAD RAFAEL LANDIVAR**  
**FACULTAD DE INGENIERIA**  
**INGENIERIA EN INFORMATICA Y SISTEMAS**  
**MANEJO E IMPLEMENTACION DE ARCHIVOS**



**Universidad**  
**Rafael Landívar**  
Tradición Jesuita en Guatemala

# **REPORTE DE PROYECTO**

**PABLO ISAAC GARCIA ORELLANA 15352-18**

# INTRODUCCION

El formato PDF (Portable Document Format) es en la actualidad el estándar de facto en todo el mundo para el intercambio de documentos tanto a nivel empresarial como de usuario. Su éxito está dado principalmente por ser multiplataforma, puede visualizarse e imprimirse de forma indistinta en sistemas Macintosh, Microsoft Windows, Unix, Linux y otros dispositivos móviles.

Así, este proyecto pretende crear una herramienta de software que sea capaz de leer los metadatos de los documentos PDF, tamaño del archivo, número de página, título del documento, asunto, palabras clave, tipo de archivo, versión de PDF, aplicación con la que fue creada, lista de imágenes que pueda tener el documento y lista de fuentes que incluye el documento.

Se realizó previamente un estudio de métodos para conocer la estructura de los diferentes tipos de archivos PDF y poder obtener la información mencionada.

# MANUAL DE USO

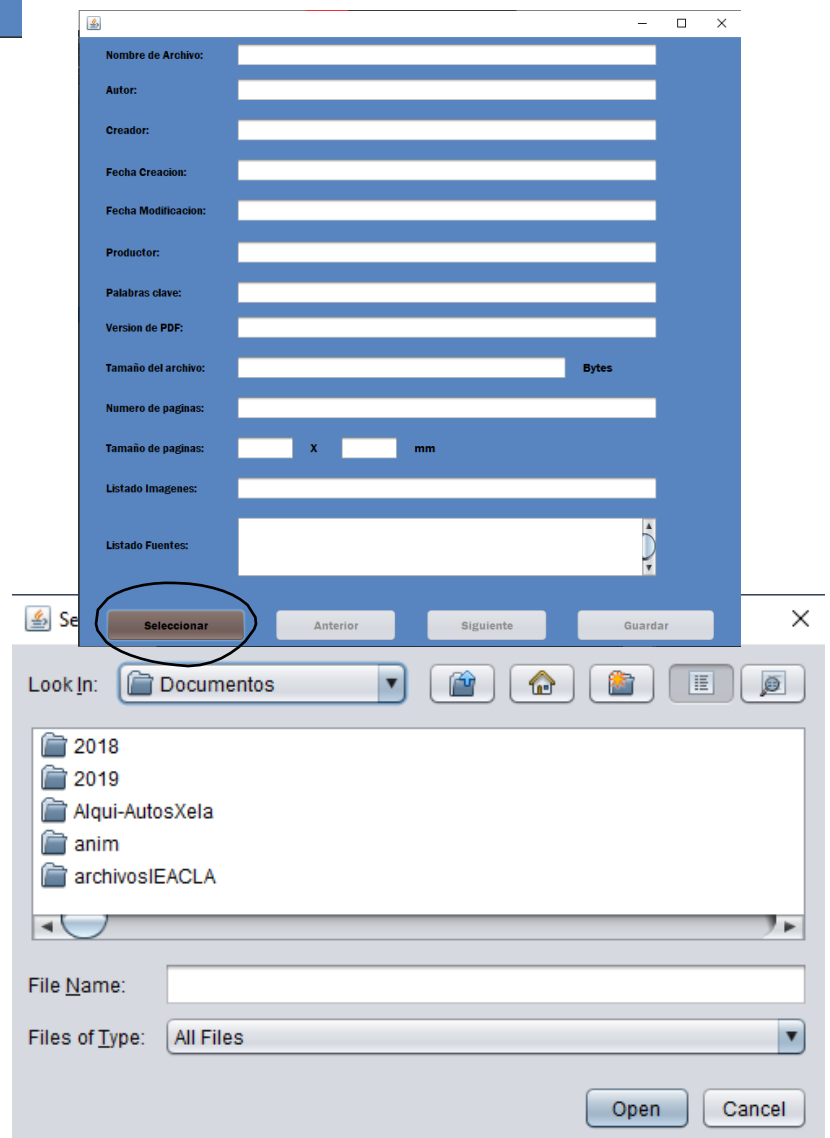


Al momento de ejecutar el programa, aparecerá esta ventana.

Primero necesitaremos  
Seleccionar Archivos

Esta será la segunda ventana que le aparecerá al usuario, Le daremos al botón "Seleccionar" y se abrirá una ventana emergente en donde podrá elegir los archivos PDF que desee leer.

Tendrá la opción de elegir una carpeta y darle open. Si quiere abrir y leer una carpeta que esta adentro de otra darle doble click a la carpeta, al momento de leer la carpeta correcta le aparecerá lo siguiente al usuario.



Nombre de Archivo: Almacenaje de archivos.pdf

Autor: Claudia Chum

Creador: IW00r0d0 p0a0r0a0 M0i0e0r0o0s0o0f0t0 030605

Fecha Creacion: D:20211105113824-06'00'

Fecha Modificacion: D:20211105113824-06'00'

Productor: IW00r0d0 p0a0r0a0 M0i0e0r0o0s0o0f0t0 030605

Palabras clave:

Version de PDF: %PDF-1.7

Tamaño del archivo: 44095 Bytes

Numero de paginas: 1

Tamaño de paginas: 0.0 X 0.0 mm

Listado Imagenes: 0

Listado Fuentes: BCDEEE+CourierNewPSMT/ArialMT/BCDFEE+Centu

Seleccionar Anterior Siguiente

Si abre una carpeta podrán navegar sobre todos los archivos con los botones "Siguiente" y "anterior" y hasta llegar al último archivo se habilitará el botón de "Guardar"

Si abre un solo archivo, los botones de "Siguiente" y "Anterior" estarán bloqueados y solamente podrá guardar el documento.

En esta ventana es en donde se podrá modificar el nombre o el autor del archivo. Una vez que se guarde ya no se podrá modificar nuevamente.

Al guardarlo mostrará un mensaje emergente en donde confirmará que se

Nombre de Archivo: ArchivoPrueba.pdf

Autor: Claudia Chum

Creador: IW00r0d0 p0a0r0a0 M0i0e0r0o0s0o0f0t0 030605

Fecha Creacion: D:20210927195111-06'00'

Fecha Modificacion: D:20210927195111-06'00'

Productor: IW00r0d0 p0a0r0a0 M0i0e0r0o0s0o0f0t0 030605

Palabras clave:

Version de PDF: %PDF-1.7

Tamaño del archivo: 41351 Bytes

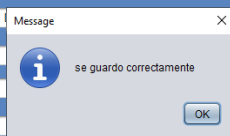
Numero de paginas: 1

Tamaño de paginas: 0.0 X 0.0 mm

Listado Imagenes: 0

Listado Fuentes: BCDEEE+Calibri/

Seleccionar Anterior Siguiente Guardar



Un solo archivo

Regresaremos al formulario principal y ahora podremos cargar los archivos guardados.

En este formulario tendremos la opción de cargar los datos de todos los archivos pdf que están almacenados o de buscar el archivo por el nombre incluyendo ".pd" o por el autor.

Cargar Archivos Guardados

Seleccionar Archivos PDF

0.0 X 0.0 mm

0

CIDFont+F1/

ArchivoPrueba.pdf

Claudia Chum

000M0i0e0r0o0s0o0f0t0000W00r0d0 p0a0r0a0 M0i0e0r0o0s0o0f0t0 030605

D:20210927195111-06'00'

D:20210927195111-06'00'

000M0i0e0r0o0s0o0f0t0000W00r0d0 p0a0r0a0 M0i0e0r0o0s0o0f0t0 030605

%PDF-1.7

41351

1

0.0 X 0.0 mm

0

BCDEEE+Calibri/

Buscar Cargar

Input

Ingrese Nombre del archivo como nombre.pdf o autor

ArchivoPrueba.pdf

OK Cancel

ArchivoPrueba.pdf

Claudia Chum

000M0i0e0r0o0s0o0f0t0000W00r0d0 p0a0r0a0 M0i0e0r0o0s0o0f0t0 030605

D:20210927195111-06'00'

D:20210927195111-06'00'

000M0i0e0r0o0s0o0f0t0000W00r0d0 p0a0r0a0 M0i0e0r0o0s0o0f0t0 030605

%PDF-1.7

41351

1

0.0 X 0.0 mm

0

BCDEEE+Calibri/

Buscar Cargar

# ESTRUCTURA DE ALMACENAMIENTO

Lectura Archivo			
VARIABLE	No. Bytes	TIPO	JUSTIFICACIÓN
Nombre	El # de bytes varían	Byte[]	Se usa de tipo array, porque varia la cantidad de caracteres.
Autor	El # de bytes varían	Byte[]	Se usa de tipo array, porque varia la cantidad de caracteres.
Creador	El # de bytes varían	Byte[]	Se usa de tipo array, porque varia la cantidad de caracteres.
Fecha Creación	23 bytes	Byte[]	Se usa de tipo array, y se le asigna un valor por que son los que nosotros creemos que utilizará.
Fecha Modificación	23 bytes	Byte[]	Se usa de tipo array, y se le asigna un valor porque son los que nosotros creemos que utilizará.
Productor	El # de bytes varían	Byte[]	Se usa de tipo array, porque varia la cantidad de caracteres.
Versión	8 bytes	Byte[]	Se usa de tipo array, y se le asigna un valor por que son los que nosotros creemos que son los que utilizará.

Tamaño	4 bytes	Int	Aquí se utiliza de tipo int, ya que son suficientes los 4 bytes para el tamaño.
Paginas	4 bytes	Int	Se utiliza tipo int, ya que los 4 bytes son suficientes para su uso.
Tamaño X	4 bytes	Float	Se utiliza float por que se da en decimales. Y son suficientes para el uso del tamaño.
Tamaño Y	4 bytes	Float	Se utiliza float por que se da en decimales. Y son suficientes para el uso del tamaño.
Imágenes	4 bytes	Int	Aquí es de tipo int, ya que es suficiente para contar la cantidad de imágenes.
Fuentes	El # de bytes varían	Byte[]	Aquí es un array depende de los caracteres que puede tomar en total.

Escritor Archivo			
Variable	No. Bytes	Tipo	Justificación
Fecha Creación	23 bytes	String	Variable String que guardamos la fecha de creación del archivo.
Nombre	Variable	String	Variable String que guardamos el nombre del archivo.
Autor	Variable	String	Variable String que guardamos la fecha de creación del archivo.
Creador	Variable	String	Variable de tipo String que guarda el creador del archivo.

Fecha Modificación	23 bytes	Char[]	Array con un valor asignado que creemos son lo que utilizará para almacenar la fecha de modificación.
Productor	Variable	String	Variable de tipo String que guarda el productor.
Versión	8 bytes	Char[]	Array con un valor asignado que creemos son lo que utilizará para almacenar la versión del archivo.
Tamaño	4 bytes	Int	Variable tipo int, suficiente para guardar el tamaño.
Paginas	4 bytes	Int	Variable tipo int, suficiente para guardar el número de páginas.
Tamaño X	4 bytes	Float	Variable tipo int, suficiente para guardar el tamaño.
Tamaño Y	4 bytes	Int	Variable tipo int, suficiente para guardar el tamaño.
Imágenes	4 bytes	Int	Variable tipo int, suficiente para guardar las imágenes.
Fuentes	Variable	String	Variable de tipo String que guarda las fuentes.

# ALGORITMOS UTILIZADOS

- **Lectura Metadatos:** este algoritmo consiste en varios ciclos while, condiciones, ciclos for, y variables que leen partes específicas del archivo PDF para poder obtener la información, se tuvo que navegar a través de la estructura del archivo para llegar a los metadatos, para leer ordenadamente los archivos se utilizó una estructura para separar e ir encontrando los datos.
- **Lectura de Versión:** este algoritmo consiste en leer la primera línea del PDF, ya que todos los PDF tienen versión en la primera línea.
- **Lectura de Tamaño:** este algoritmo está estructurado de comandos directos, obtener el tamaño del archivo en bytes.
- **Lectura Numero de Paginas:** este algoritmo consiste en varios ciclos while, condiciones, ciclos for, y variables están estructurados para navegar en el archivo PDF y así solo leer el número de páginas.
- **Lectura Tamaño de la Pagina:** este algoritmo consiste en varios ciclos while, condiciones, ciclos for, y variables están estructuradas para navegar en el archivo PDF y leer solamente el tamaño de página que estaba dada en point y el algoritmo lo convierte en milímetros.
- **Lectura de Imágenes:** este algoritmo consiste en varios ciclos while, condiciones, ciclos for y variables están estructuradas para navegar en el archivo PDF e ir leyendo página por página las imágenes que contiene cada una.
- **Lectura de Fuentes:** este algoritmo consiste en varios ciclos while, condiciones, ciclos for, y variables están estructuradas para navegar en el archivo PDF e ir leyendo las fuentes que contiene cada página diferente.
- **Escritura Aleatoria:** este algoritmo consiste en condiciones, variables que están estructuradas para crear un archivo de tipo .bin que almacena registros de los archivos PDF y un ciclo while para recorrer la lista del índice del mismo.
- **Leer Índice:** este algoritmo consiste en almacenar en una lista todos los índices del archivo .bin para no perder la información de los mismos y sobre escribir al final del archivo



- Búsqueda de Archivo: este algoritmo consiste en varios ciclos while, condiciones, ciclos for y variables estan estructuradas para navegar dentro del archivo creado con registros de PDF, esto para buscar tanto como el nombre o autor.
- Lector Aleatorio: este algoritmo consiste en varios ciclos while, condiciones, ciclos for y variables estan estructuradas para navegar en el archivo, almacena la información del PDF en un .bin, de tal forma, al momento de buscarlo aparece la informacion en una ventana
- Lista, InsertarM, NodoM, MostrarM: este algoritmo consiste en una lista doblemente enlazada en donde se puede recorrer los diferentes registros de archivos PDF, con un botón de siguiente y anterior que son lo que nos ayudan a recorrer dentro de una carpeta con varios archivos PDF.

# COCLUSIONES

La realización del estudio de métodos en el proyecto Lector, Escritura, Buscador y Modificación del pdf es importante debido a que, a partir de éste, se puede elaborar un sistema de mejoras a los procesos estudiados, haciendo que sean óptimos y que a su vez permita aumentar los niveles de productividad de la generación.

Durante y posterior a la construcción del proyecto se realizaron pruebas en el sistema para garantizar la calidad del funcionamiento de los mismos respecto a los criterios solicitados

Con la realización de este proyecto se han cumplido los objetivos que se fijaron al comienzo del mismo, obteniendo resultados que los validan.