# Archivos Fragmentados y Sobrescritos

Los PDF tienen una estructura clara que facilita su recuperación:

1. **Cabecera (%PDF-x.x):**
   * Siempre está al inicio del archivo y especifica la versión del formato PDF (por ejemplo, %PDF-1.7).
   * Esta cabecera es un patrón único y confiable para identificar el comienzo de un archivo PDF en el disco.
2. **Cuerpo:**
   * Contiene objetos como texto, imágenes, gráficos y metadatos.
   * Los objetos tienen identificadores únicos (Obj ID) y son referenciados en la tabla xref.
3. **Tabla de referencias cruzadas (xref):**
   * Lista las posiciones de los objetos dentro del archivo.
   * Es clave para reconstruir archivos fragmentados porque señala dónde se encuentran las partes del archivo en el disco.
4. **Cola (trailer):**
   * Aparece al final del archivo y apunta al inicio de la tabla xref.
   * Contiene información crítica como el número total de objetos y el objeto raíz.

Incluso si el archivo está fragmentado, estos patrones permiten localizar partes específicas del archivo y reconstruirlo.

## Funcionamiento de los Sistemas de Archivos en Windows

Windows utiliza principalmente el sistema NTFS (New Technology File System). Este sistema organiza los datos en discos de la siguiente manera:

1. **Estructura de bloques:**
   * Los discos se dividen en bloques físicos (típicamente de 512 bytes o múltiplos como 4 KB).
   * Cada archivo ocupa uno o más bloques. Si un archivo es grande o el disco está fragmentado, sus bloques pueden almacenarse en ubicaciones no consecutivas.
2. **Master File Table (MFT):**
   * En NTFS, la MFT almacena información sobre cada archivo, incluyendo:
     + Nombre.
     + Ubicación de los bloques en el disco.
     + Tamaño y atributos.
   * Al eliminar un archivo, su entrada en la MFT se marca como "libre", pero los datos en los bloques físicos permanecen intactos hasta que se sobrescriben.
3. **Fragmentación:**
   * Ocurre cuando un archivo no puede ser almacenado en bloques contiguos. Esto hace que los datos se dispersen por el disco, complicando su recuperación.

## Acceso al Disco en Bajo Nivel

Para recuperar archivos directamente, necesitas ignorar el sistema de archivos y trabajar a nivel de bloques físicos del disco.

1. **Acceso físico al disco:**
   * Lee el contenido del disco directamente, bloque por bloque.
   * Esto implica acceder a los sectores físicos del disco, ignorando la lógica del sistema operativo.
2. **Herramientas técnicas:**
   * Puedes utilizar herramientas de desarrollo como bibliotecas de bajo nivel (por ejemplo, en Python, el acceso directo al disco con os o ctypes).
   * Otra opción es trabajar con imágenes forenses del disco creadas previamente.
3. **Registro de patrones:**
   * Escanea cada bloque del disco buscando la cabecera %PDF.
   * Registra la posición exacta del bloque donde aparece.

Este paso inicial te permite identificar posibles inicios de archivos PDF, incluso si están eliminados o fragmentados.

## Localización y Reconstrucción de Fragmentos

Cuando un archivo está fragmentado, sus partes pueden estar dispersas en el disco. El objetivo es encontrar y unir estos fragmentos.

1. **Identificación de cabeceras (%PDF):**
   * Cada vez que encuentres la cabecera, verifica los sectores contiguos para identificar si pertenecen al mismo archivo.
   * Si detectas patrones como objetos (obj) o tablas xref, probablemente sean fragmentos del archivo.
2. **Búsqueda en sectores vecinos:**
   * Explora los bloques adyacentes al primero que identificaste como parte del PDF.
   * Analiza el contenido en busca de otros patrones típicos de un PDF.
3. **Uso de la tabla xref:**
   * Si encuentras la tabla xref, utilízala para localizar la posición de los objetos dentro del archivo.
   * La xref te permite reconstruir el archivo al indicar qué fragmentos pertenecen al archivo y su posición lógica.
4. **Reconstrucción:**
   * Una vez localizados los fragmentos, organízalos en el orden correcto según la información de la xref o su contenido.

## Consideraciones para Archivos Sobrescritos

La recuperación de archivos sobrescritos depende de cuánto del archivo original permanece intacto:

1. **Sobrescritura parcial:**
   * Si solo algunos bloques fueron sobrescritos, es posible recuperar partes intactas del archivo.
   * Analiza los bloques marcados como "libres" y busca patrones de datos residuales.
2. **Sobrescritura total:**
   * Si todos los bloques han sido reemplazados, la recuperación ya no es posible, ya que los datos originales han desaparecido.

**Medidas clave:**

* Detén cualquier escritura en el disco para evitar más sobreescritura.
* Trabaja con una copia forense del disco para preservar su estado original.

## Validación del Archivo Reconstruido

Una vez que hayas ensamblado los fragmentos, valida la integridad del archivo:

1. **Comprobación estructural:**
   * Verifica que el archivo tenga una cabecera %PDF válida.
   * Asegúrate de que la tabla xref sea consistente con los objetos del cuerpo.
2. **Pruebas funcionales:**
   * Abre el archivo con un lector de PDF para confirmar que sea legible.
   * Si hay errores, busca fragmentos faltantes o corrige la tabla xref.

## Desafíos Comunes

* **Fragmentación extrema:** Si los fragmentos están muy dispersos, puede ser difícil identificar cuáles pertenecen al archivo.
* **Archivos grandes:** Los archivos PDF grandes suelen tener más fragmentos, aumentando la complejidad de la reconstrucción.
* **Daño en la tabla xref:** Si falta esta tabla, debes usar métodos heurísticos para unir los fragmentos.

# Firmas para Recuperación en General

La recuperación de datos mediante firmas, también conocida como "data carving", es una técnica avanzada utilizada para restaurar archivos cuando la información del sistema de archivos está dañada o es inaccesible. Esta metodología se basa en identificar patrones específicos, o "firmas", que son secuencias únicas de bytes al inicio o final de un archivo, permitiendo así su reconstrucción sin depender de la estructura del sistema de archivos.

## Principios de la Recuperación Basada en Firmas

A diferencia de los métodos tradicionales que dependen de tablas de asignación de archivos o directorios, la recuperación basada en firmas escanea directamente el contenido completo del medio de almacenamiento. Busca secuencias de bytes que correspondan a las firmas conocidas de tipos de archivos específicos. Una vez identificada una firma, el software intenta reconstruir el archivo completo, basándose en la información obtenida de estas firmas. Este enfoque es especialmente útil en situaciones donde:

* El sistema de archivos está severamente dañado o corrupto.
* Se han eliminado particiones o formateado el disco.
* Los metadatos del sistema de archivos no están disponibles.

## Proceso de Data Carving

* Escaneo Completo: El software lee la superficie completa del medio de almacenamiento, analizando cada sector en busca de firmas conocidas.
* Identificación de Firmas: Al detectar una secuencia de bytes que coincide con una firma predefinida, se marca el inicio potencial de un archivo.
* Reconstrucción del Archivo: A partir de la firma identificada, el software intenta reconstruir el archivo completo, determinando su tamaño y estructura basándose en la firma y otros patrones conocidos.

## Limitaciones

Aunque el data carving es una herramienta poderosa, tiene ciertas limitaciones:

* Fragmentación: Si un archivo está fragmentado en el disco, la recuperación completa puede ser difícil, ya que las partes del archivo pueden no estar contiguas.
* Firmas Desconocidas: Si el tipo de archivo no tiene una firma bien definida o es poco común, el software puede no reconocerlo.
* Archivos Parciales: En algunos casos, solo se pueden recuperar partes de archivos si el resto de los datos ha sido sobrescrito o está dañado.

## Herramientas de Recuperación Basadas en Firmas para Windows

Existen varias herramientas especializadas que implementan técnicas de data carving para la recuperación de datos en sistemas Windows:

* Windows File Recovery: Una herramienta oficial de Microsoft que permite recuperar archivos eliminados utilizando diferentes modos de recuperación, incluyendo el basado en firmas.
* PhotoRec: Un software de código abierto diseñado para recuperar archivos de diversos formatos mediante técnicas de data carving.
* Ontrack EasyRecovery: Ofrece recuperación precisa de archivos perdidos debido a eliminación, reformateo y otros escenarios de pérdida de datos.

# Referencias

[1] "Repair NTFS File System and Recover Data." Recoverit Wondershare. https://recoverit.wondershare.es/file-recovery/repair-ntfs-file-system-and-recover-data.html (accedido el 1 de noviembre de 2024).

[2] "Recovery Chances on Windows." Raise Data Recovery. https://www.raisedr.com/resources/general/recovery-chances-windows/ (accedido el 1 de noviembre de 2024).

[3] "Windows File Recovery." Microsoft Support. https://support.microsoft.com/en-us/windows/windows-file-recovery-61f5b28a-f5b8-3cc2-0f8e-a63cb4e1d4c4 (accedido el 1 de noviembre de 2024).

[4] "¿Cómo funciona la recuperación de archivos?" Busindre. https://www.busindre.com/como\_funciona\_la\_recuperacion\_de\_archivos (accedido el 1 de noviembre de 2024).

[5] "Catalog Files." Microsoft Learn. https://learn.microsoft.com/es-es/windows-hardware/drivers/install/catalog-files (accedido el 1 de noviembre de 2024).

[6] "Windows File Recovery." Microsoft Support. https://support.microsoft.com/en-us/windows/windows-file-recovery-61f5b28a-f5b8-3cc2-0f8e-a63cb4e1d4c4 (accedido el 1 de noviembre de 2024).

[7] "Backup and Restore with File History." Microsoft Support. https://support.microsoft.com/en-us/windows/backup-and-restore-with-file-history-7bf065bf-f1ea-0a78-c1cf-7dcf51cc8bfc (accedido el 1 de noviembre de 2024).

[8] "Los cinco mejores programas de recuperación de datos." Huffington Post. https://www.huffingtonpost.es/tecnologia/los-cinco-mejores-programas-recuperacion-datos.html (accedido el 1 de noviembre de 2024).

[9] "Power Search File Types." DiskInternals. https://www.diskinternals.com/power-search/file\_types/ (accedido el 1 de noviembre de 2024).

[10] "Top 5 Data Recovery Software for Windows." CleverFiles. https://www.cleverfiles.com/howto/top-5-data-recovery-software-windows.html (accedido el 1 de noviembre de 2024).

[11] "Ontrack EasyRecovery for Windows." Ontrack. https://www.ontrack.com/en-us/software/easyrecovery/windows (accedido el 1 de noviembre de 2024).