



**LokyApps**

HERRAMIENTAS DE PRODUCTIVIDAD

# Manual de usuario **LokyScan**

Guía detallada para detectar archivos duplicados, versiones, contenido reutilizado y documentos parecidos de forma local.

30 páginas

Paso a paso

Privacidad local

Exportación PDF/CSV/JSON

**Objetivo:** que cualquier usuario pueda cargar archivos o carpetas, escoger el modo de análisis, interpretar los resultados y decidir qué conservar sin borrar información importante por error.



**LokyScan**

Analiza · Compara · Detecta

MANUAL LOKYAPPS

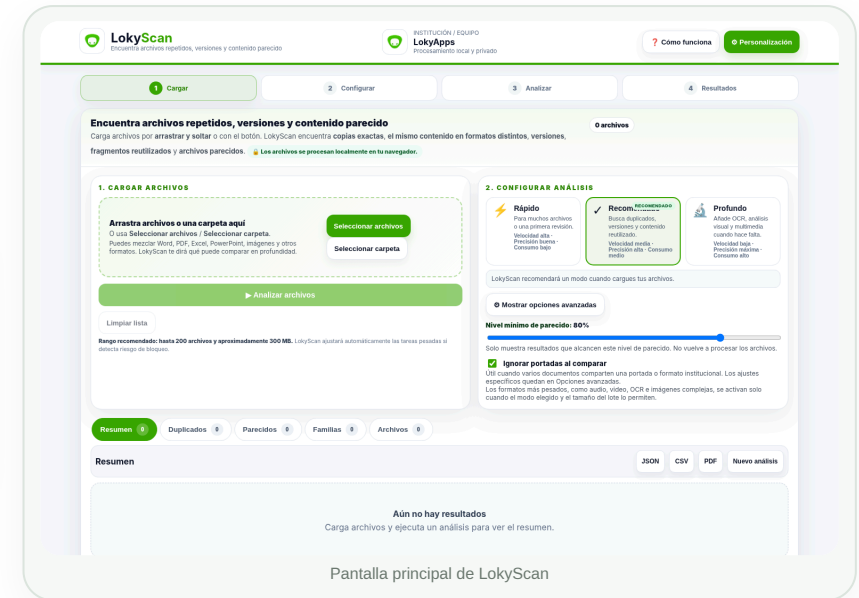
## PROPÓSITO

LokyScan compara archivos dentro del navegador para localizar copias exactas, contenido equivalente, posibles versiones y fragmentos reutilizados.

### Cuándo usarlo

- ✓ Antes de depurar una carpeta grande.
- ✓ Cuando existen varias versiones de informes.
- ✓ Para encontrar el mismo contenido en formatos distintos.
- ✓ Para revisar carpetas históricas antes de archivar.
- ✓ Para estimar espacio recuperable por duplicados.

**Privacidad:** el análisis se realiza localmente. El usuario decide qué archivos cargar y qué acciones tomar después de revisar los resultados.



Pantalla principal de LokyScan



## Duplicado exacto

Dos archivos con contenido binario idéntico; la herramienta utiliza SHA-256 para confirmarlo.

## Mismo contenido

Documentos que conservan esencialmente el mismo contenido aunque el formato o la representación sea diferente.

## Posible versión

Archivos muy parecidos con cambios, adiciones, eliminaciones o reordenamiento.

## Copia parcial

Una parte significativa de un documento aparece dentro de otro.

## Familia de archivos

Agrupación de elementos relacionados que conviene revisar como conjunto.

## Contradicción

Casos donde nombre, contenido u otras señales merecen revisión adicional.

**No equivale a “puedo borrar”.** Un resultado de similitud es una evidencia para revisar; solo los duplicados exactos ofrecen una señal binaria fuerte y aun así conviene verificar ubicación, fecha y uso.



## Antes de empezar

- ✓ Navegador moderno, preferiblemente Chrome o Edge.
- ✓ Memoria suficiente para el tamaño del lote.
- ✓ Archivos accesibles desde el equipo local.
- ✓ Tiempo adicional si activa OCR, imágenes o multimedia.
- ✓ Evite cerrar la pestaña mientras el análisis está en curso.

## Rangos orientativos

| Condición           | Guía práctica                                                                |
|---------------------|------------------------------------------------------------------------------|
| Uso recomendado     | Hasta 200 archivos y aproximadamente 300 MB.                                 |
| Límite de seguridad | La aplicación bloquea lotes extremos de más de 800 archivos o 1,5 GB.        |
| Profundo completo   | Mejor en sublots pequeños; la app puede aplicar análisis profundo selectivo. |
| ZIP                 | Abrir el contenido puede multiplicar el número real de archivos.             |

**Consejo:** si tiene miles de archivos, analice por carpetas o periodos. Luego repita el proceso con los grupos que realmente necesitan comparación profunda.



## CUATRO ETAPAS



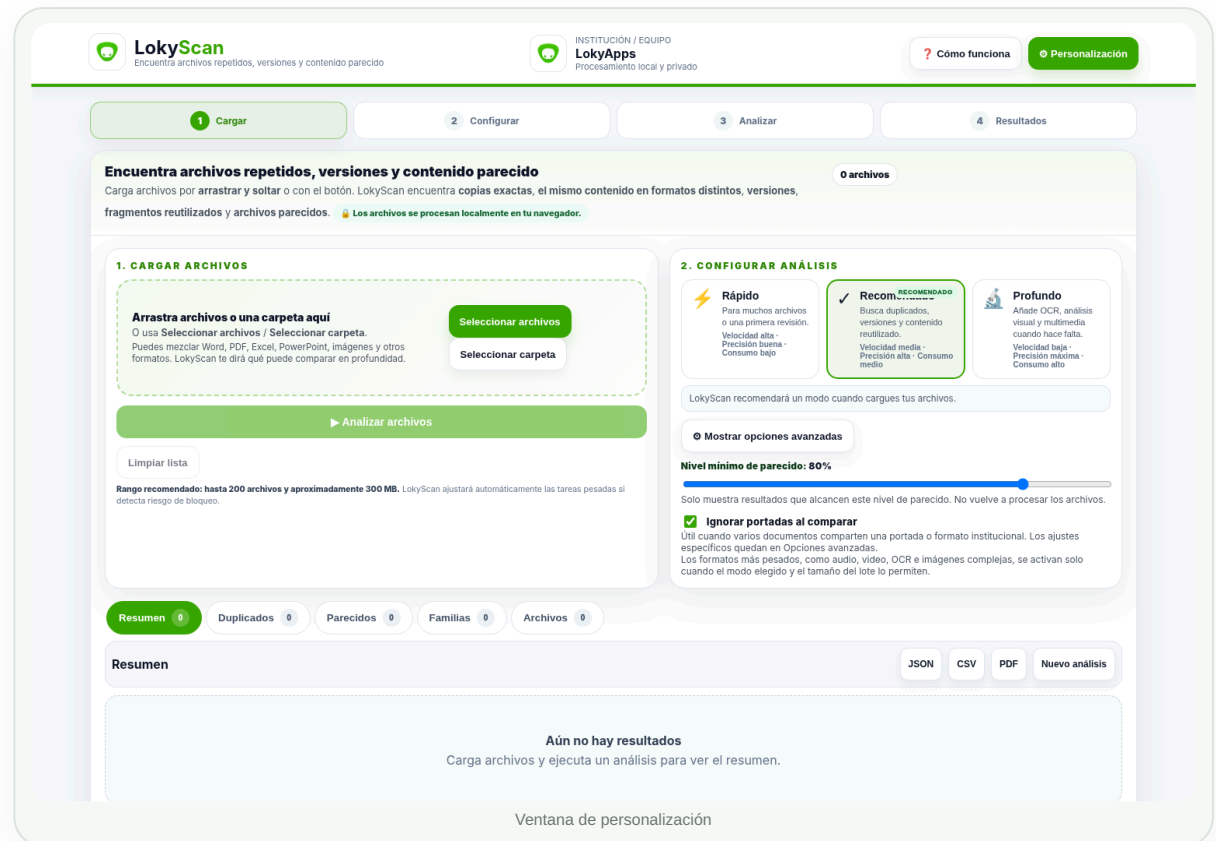
**Ruta para principiantes:** Cargar → Recomendado → Analizar → Resumen → Duplicados → Parecidos → Exportar PDF.

## Qué puede cambiar

- ✓ Nombre de institución o equipo.
- ✓ Logo institucional.
- ✓ Color principal de la interfaz.
- ✓ Restablecer la identidad predeterminada.

- 1 Abra ⚙ Personalización.
- 2 Escriba el nombre del equipo o institución.
- 3 Seleccione el logo si desea sustituir el predeterminado.
- 4 Defina el color principal.
- 5 Pulse Guardar cambios.

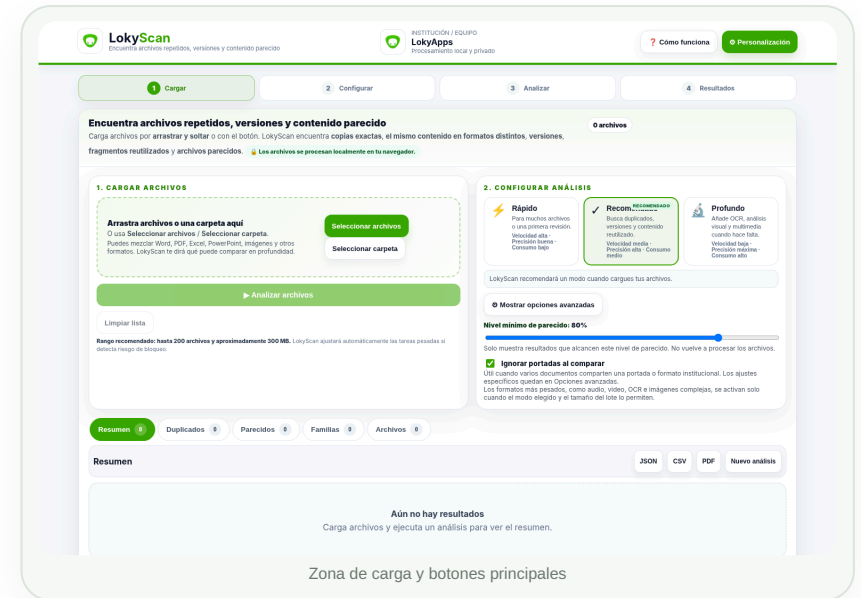
**LokyApps:** para mantener la identidad original, use el logo oficial y el verde #39A900.



- 1 Abra LokyScan.
- 2 Arrastre archivos al área punteada o pulse Seleccionar archivos.
- 3 Para revisar una carpeta completa, use Seleccionar carpeta.
- 4 Puede mezclar distintos formatos en el mismo lote.
- 5 Verifique el contador de archivos antes de iniciar.

## ¿Archivo o carpeta?

| Opción               | Úsela cuando                                                                    |
|----------------------|---------------------------------------------------------------------------------|
| Seleccionar archivos | Solo desea comparar documentos específicos.                                     |
| Seleccionar carpeta  | Quiere revisar un conjunto completo con subrutas disponibles para el navegador. |



Zona de carga y botones principales



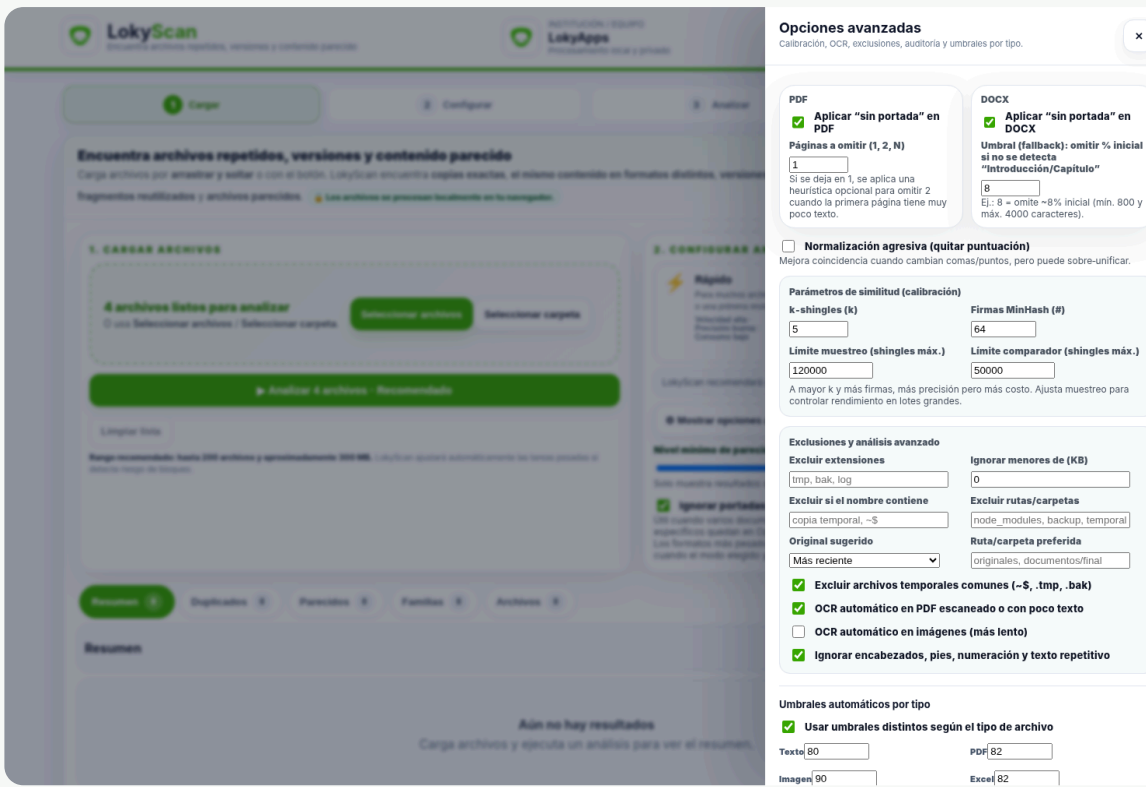
| Familia        | Ejemplos                                          | Profundidad esperada                        |
|----------------|---------------------------------------------------|---------------------------------------------|
| Texto y código | TXT, MD, CSV, JSON, XML, HTML, JS, TS, PY, SQL... | Comparación textual.                        |
| PDF            | PDF digital o escaneado                           | Texto, estructura y OCR si corresponde.     |
| Word           | DOCX; DOC con limitaciones                        | DOCX permite extracción más rica.           |
| Excel          | XLSX, XLS                                         | Análisis estructural/contenido.             |
| PowerPoint     | PPTX; PPT con limitaciones                        | Texto y señales visuales en PPTX.           |
| Imagen         | PNG, JPG, WEBP, TIFF...                           | Huellas perceptuales y visuales.            |
| Audio          | MP3, WAV, FLAC, AAC...                            | Señales multimedia según modo.              |
| Video          | MP4, WEBM, MOV, AVI, MKV...                       | Señales multimedia según modo.              |
| Comprimidos    | ZIP; otros archivos pueden cargarse               | ZIP puede inspeccionarse con restricciones. |

**Importante:** que un formato pueda cargarse no significa que todos sus contenidos puedan extraerse con la misma precisión. Los formatos modernos abiertos como DOCX/XLSX/PPTX ofrecen más información que algunos formatos heredados.

## Antes de analizar

- ✓ Confirme que el contador coincide con lo esperado.
- ✓ Revise nombres y extensiones.
- ✓ Quite archivos innecesarios si el lote es demasiado grande.
- ✓ Use Limpiar lista para empezar de nuevo.
- ✓ No confunda “cargado” con “analizado”: todavía no hay resultados.

**Ejemplo:** una carpeta puede contener informe\_original, una copia exacta, una versión editada y un resumen que reutiliza párrafos. Ese conjunto permite ver las cuatro clases de relación.



**Opciones avanzadas**  
Calibración, OCR, exclusiones, auditoría y umbrales por tipo.

**PDF**  
☒ Aplicar “sin portada” en PDF  
Páginas a omitir (1, 2, N)  
1  
Si se deja en 1, se aplica una heurística opcional para omitir 2 cuando la primera página tiene muy poco texto.

**DOCX**  
☒ Aplicar “sin portada” en DOCX  
Umbral (fallback): omitir % inicial si no se detecta “Introducción/Capítulo”  
8  
Ej.: 8 = omitir ~8% inicial (min. 800 y máx. 4000 caracteres).

☐ Normalización agresiva (quitar puntuación)  
Mejora coincidencia cuando cambian comas/puntos, pero puede sobre-unificar.

**Parámetros de similitud (calibración)**  
k-shingles (k): 5  
Firmas MinHash (#): 64  
Limite muestreo (shingles máx.): 120000  
Limite comparador (shingles máx.): 50000  
A mayor k y más firmas, más precisión pero más costo. Ajusta muestreo para controlar rendimiento en lotes grandes.

**Exclusiones y análisis avanzado**  
Excluir extensiones: tmp, bak, log  
Ignorar menores de (KB): 0  
Excluir si el nombre contiene: copia temporal, --\$  
Excluir rutas/carpetas: node\_modules, backup, temporal  
Original sugerido: Más reciente  
Ruta/carpeta preferida: originales, documentos/final  
☒ Excluir archivos temporales comunes (--\$, .tmp, .bak)  
☒ OCR automático en PDF escaneado o con poco texto  
☐ OCR automático en imágenes (más lento)  
☒ Ignorar encabezados, pies, numeración y texto repetitivo

**Umbrales automáticos por tipo**  
☒ Usar umbrales distintos según el tipo de archivo  
Texto: 80  
PDF: 82  
Imagen: 90  
Excel: 82

Ejemplo demostrativo con archivos preparados para análisis



# Elegir el modo de análisis



## Rápido

Ideal para muchos archivos o una primera revisión. Prioriza velocidad y bajo consumo.



## Profundo

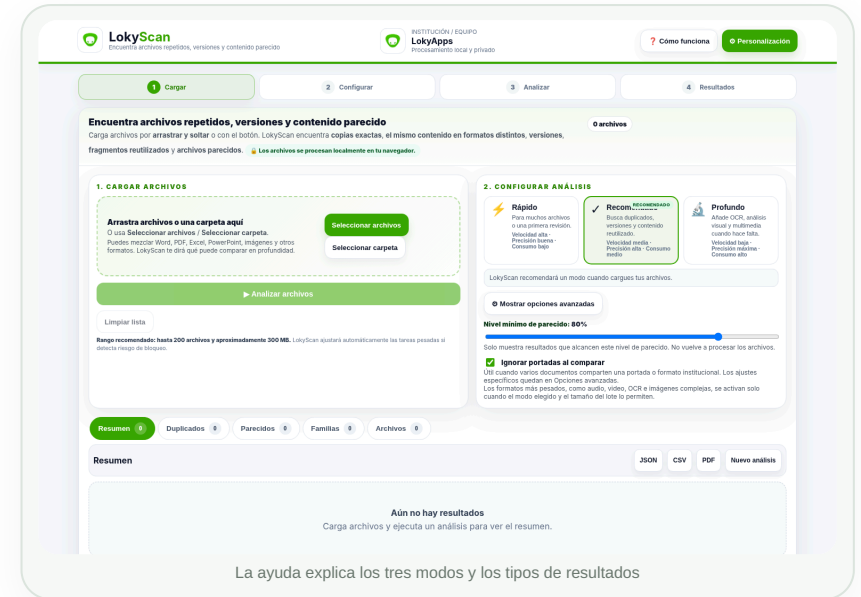
Para lotes pequeños con OCR, imágenes, audio o video. Consume más memoria y tiempo.



## Recomendado

Mejor opción general. Busca duplicados, versiones, párrafos reutilizados y parecido Word/PDF.

**Para empezar:** use Recomendado. Cambie a Profundo solamente cuando el análisis normal no sea suficiente o cuando necesite señales visuales/multimedia.



La ayuda explica los tres modos y los tipos de resultados



## Qué evalúa LokyScan antes de comenzar

- ✓ Cantidad de archivos.
- ✓ Peso total del lote.
- ✓ Tipos de archivo presentes.
- ✓ Necesidad de OCR o señales visuales.
- ✓ Presencia de ZIP.
- ✓ Riesgo de consumo excesivo de memoria.

## Decisiones automáticas

- ✓ Reducir tareas costosas en lotes grandes.
- ✓ Aplicar análisis profundo selectivo.
- ✓ Limitar expansión de ZIP.
- ✓ Evitar OCR masivo cuando exista riesgo de bloqueo.
- ✓ Permitir "Forzar análisis completo" solo como decisión consciente.

**Si aparece una advertencia de carga alta:** acepte el modo protegido salvo que conozca la capacidad del equipo. Dividir el lote suele ser más seguro que forzar.



## Qué significa el porcentaje

El umbral determina qué comparaciones se muestran como suficientemente parecidas. Un umbral alto reduce resultados dudosos; uno bajo descubre más candidatos pero exige mayor revisión humana.

| Umbral     | Uso sugerido                                            |
|------------|---------------------------------------------------------|
| 90 % o más | Versiones casi iguales.                                 |
| 80 %       | Punto de partida razonable para documentos similares.   |
| 70–79 %    | Exploración más amplia; aumenta falsos positivos.       |
| Muy bajo   | Solo para investigación específica y lotes controlados. |

## Regla práctica

# 80%

Es una referencia útil para iniciar, no una garantía matemática de equivalencia documental.

## Opciones que ayudan a reducir falsos positivos

- ✓ Ignorar portadas al comparar.
- ✓ Aplicar "sin portada" en PDF o DOCX.
- ✓ Indicar páginas específicas a omitir.
- ✓ Ignorar encabezados, pies, numeración y texto repetitivo.
- ✓ Usar umbral de texto repetido global para detectar boilerplate.

**Ejemplo:** si 100 informes tienen la misma portada institucional, esa portada no debería hacer que todos parezcan versiones entre sí.



**Opciones avanzadas**  
Calibración, OCR, exclusiones, auditoría y umbrales por tipo.

**PDF**

- ☒ Aplicar "sin portada" en PDF
- Páginas a omitir (1, 2, N)  
1
- Si se deja en 1, se aplica una heurística opcional para omitir 2 cuando la primera página tiene muy poco texto.

**DOCX**

- ☒ Aplicar "sin portada" en DOCX
- Umbral (fallback): omitir % inicial si no se detecta "Introducción/Capítulo"  
8
- Ej.: 8 = omitir ~8% inicial (min. 800 y máx. 4000 caracteres).

☐ Normalización agresiva (quitar puntuación)  
Mejora coincidencia cuando cambian comas/puntos, pero puede sobre-unificar.

**Parámetros de similitud (calibración)**

**k-shingles (k)**  
5

**Firmas MinHash (#)**  
64

**Límite muestreo (shingles máx.)**  
120000

**Límite comparador (shingles máx.)**  
50000

A mayor k y más firmas, más precisión pero más costo. Ajusta muestreo para controlar rendimiento en lotes grandes.

**Exclusiones y análisis avanzado**

**Excluir extensiones**  
tmp, bak, log

**Ignorar menores de (KB)**  
0

**Excluir si el nombre contiene**  
copia temporal, -s

**Excluir rutas/carpetas**  
node\_modules, backup, temporal

**Original sugerido**  
Más reciente

**Ruta/carpetas preferida**  
originales, documentos/final

- ☒ Excluir archivos temporales comunes (-s, .tmp, .bak)
- ☒ OCR automático en PDF escaneado o con poco texto
- ☐ OCR automático en imágenes (más lento)
- ☒ Ignorar encabezados, pies, numeración y texto repetitivo

**Umbrales automáticos por tipo**

- ☒ Usar umbrales distintos según el tipo de archivo

**Texto** 80 **PDF** 82

**Imagen** 90 **Excel** 82

Configuración avanzada y opciones de comparación

| Opción                 | Qué modifica                                             | Cuándo tocarla                                                         |
|------------------------|----------------------------------------------------------|------------------------------------------------------------------------|
| Normalización agresiva | Reduce puntuación y diferencias superficiales.           | Cuando pequeñas diferencias de formato esconden contenido equivalente. |
| k-shingles             | Tamaño de grupos de palabras usados para comparar.       | Normalmente no necesita modificación.                                  |
| Firmas MinHash         | Cantidad de firmas para aproximar similitud.             | Aumentar mejora precisión pero consume más.                            |
| Límite muestreo        | Máximo de shingles usados al crear huellas.              | Mantenga el preset salvo lotes especiales.                             |
| Límite comparador      | Cantidad máxima de shingles para comparaciones directas. | Aumentar solo en documentos grandes y equipos capaces.                 |

**Para usuarios no técnicos:** no cambie estos valores. Los modos Rápido, Recomendado y Profundo ya configuran presets coherentes.



## Puede excluir por

- ✓ Extensión.
- ✓ Tamaño mínimo en KB.
- ✓ Texto contenido en el nombre.
- ✓ Ruta o carpeta.
- ✓ Archivos temporales comunes como ~\$ , .tmp o .bak.

## Por qué conviene excluir

- ✓ Reduce tiempo de análisis.
- ✓ Evita archivos irrelevantes.
- ✓ Disminuye ruido en resultados.
- ✓ Protege el navegador en carpetas muy grandes.
- ✓ Permite centrarse en documentos de negocio.

**Ejemplo:** si analiza una carpeta de proyecto y no le interesan logs ni respaldos temporales, exclúyalos antes de iniciar.



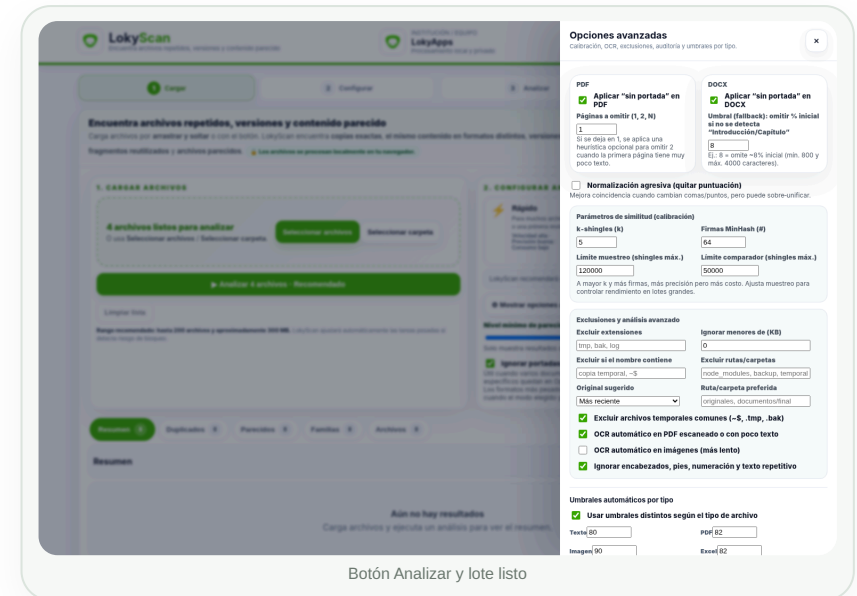
| Función          | Para qué sirve                                   | Costo                      |
|------------------|--------------------------------------------------|----------------------------|
| OCR PDF          | Extraer texto de PDF escaneado o con poco texto. | Alto.                      |
| OCR imágenes     | Reconocer texto dentro de imágenes.              | Alto.                      |
| Huellas visuales | Comparar apariencia de imágenes y PPTX.          | Medio/alto.                |
| Audio / video    | Usar señales multimedia cuando aplica.           | Alto.                      |
| Inspección ZIP   | Agregar archivos internos al lote de revisión.   | Puede multiplicar volumen. |

**Recomendación:** active funciones profundas solo cuando el tipo de evidencia lo requiera. Para una carpeta documental normal, el modo Recomendado suele ser suficiente.

- 1 Revise archivos y modo seleccionado.
- 2 Pulse Analizar.
- 3 Si aparece el plan de análisis, lea la estimación y las advertencias.
- 4 Pulse Comenzar análisis.
- 5 Observe fase, progreso y tiempo aproximado.
- 6 Si nota un problema, use Detener; luego reduzca el lote o la profundidad.

## Durante el proceso

No cierre la pestaña ni cambie archivos del lote. En operaciones pesadas, la interfaz puede tardar entre fases sin que eso signifique necesariamente que se bloqueó.

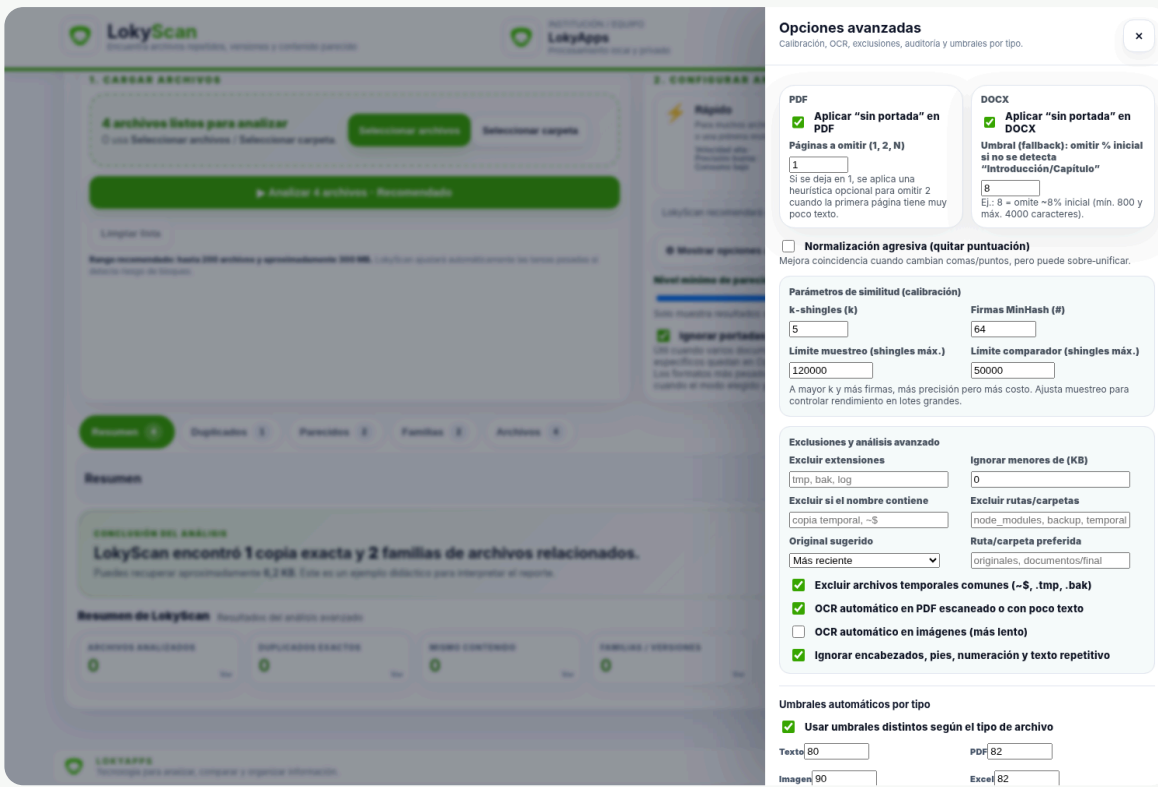


Botón Analizar y lote listo

## Empiece siempre por Resumen

- ✓ Número de copias exactas.
- ✓ Familias de archivos relacionados.
- ✓ Espacio potencialmente recuperable.
- ✓ Contradicciones que conviene revisar.
- ✓ Notas sobre texto repetitivo excluido.
- ✓ Estado de auditoría si existe referencia.

**Orden sugerido:** Resumen → Duplicados → Parecidos → Familias → Archivos.



The screenshot displays the LokyScan web interface. The main panel shows the 'Resumen' (Summary) tab, which includes a section for '4 archivos listos para analizar' (4 files ready for analysis) and a 'Resumen de LokyScan' (LokyScan Summary) section. The summary section indicates that LokyScan found 1 exact copy and 2 families of related files. The sidebar on the right contains the 'Opciones avanzadas' (Advanced Options) menu, which is currently open. This menu includes settings for PDF and DOCX files, such as 'Aplicar "sin portada" en PDF' (Apply "no cover" in PDF) and 'Aplicar "sin portada" en DOCX' (Apply "no cover" in DOCX). It also features a 'Normalización agresiva' (Aggressive normalization) option and a 'Parámetros de similitud' (Similarity parameters) section with fields for 'k-shingles (k)' and 'Firmas MinHash (#)'. The sidebar also includes 'Exclusiones y análisis avanzado' (Advanced exclusions and analysis) and 'Umbral automático por tipo' (Automatic threshold by type) settings.

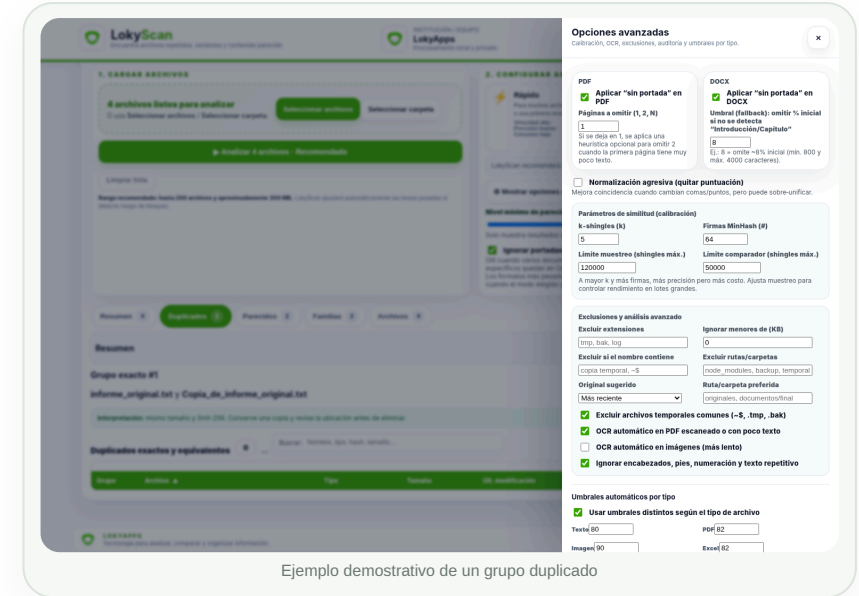
Resumen demostrativo del análisis

## LA SEÑAL MÁS FUERTE

Un duplicado exacto comparte el mismo contenido binario y el mismo SHA-256.

- 1 Abra la pestaña Duplicados.
- 2 Revise cada grupo.
- 3 Identifique la copia recomendada para conservar.
- 4 Compare ruta, nombre, fecha y contexto.
- 5 Marque una decisión externa; LokyScan no debe sustituir su criterio documental.

**Espacio redundante:** el reporte puede estimar cuánto almacenamiento se recuperaría conservando una sola copia de cada grupo exacto.



Ejemplo demostrativo de un grupo duplicado



## Mismo contenido en formatos diferentes

### Caso típico

Un informe DOCX fue exportado a PDF sin cambios relevantes.

### Qué compara

Texto normalizado, bloques, estructura y otras señales según el tipo.

### Qué hacer

Conserve ambos si cumplen funciones distintas: editable vs versión firmada o distribuida.

### Riesgo

"Mismo contenido" no significa "mismo propósito documental".

**Ejemplo de negocio:** el Word editable puede ser necesario para futuras correcciones y el PDF puede ser el registro oficial entregado. No elimine uno solo porque el contenido sea equivalente.

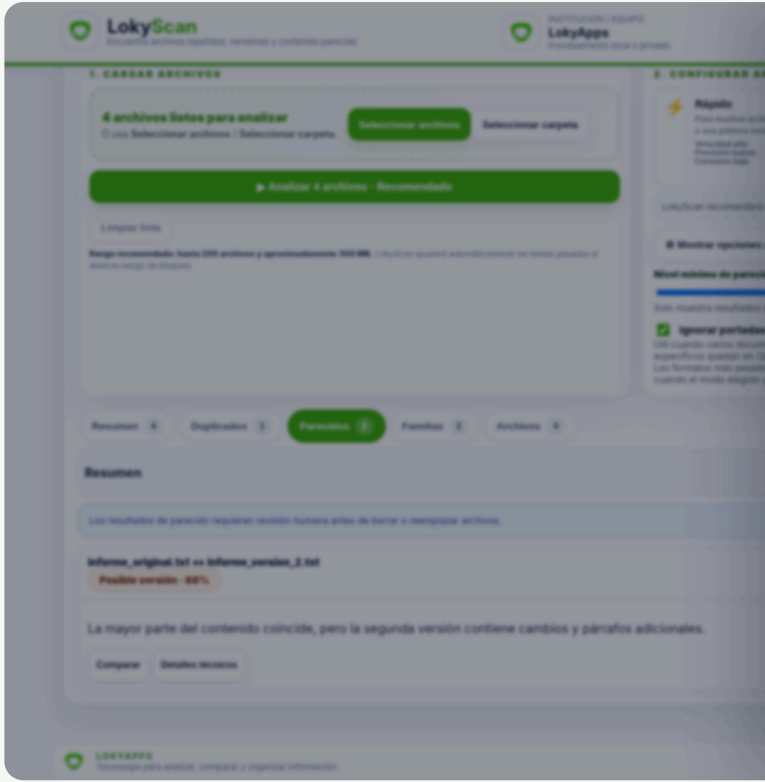
## Posible versión

La mayor parte del archivo coincide, pero existen cambios. Puede ser borrador, versión final, actualización o documento derivado.

## Copia parcial

Una sección importante aparece dentro de otro archivo mayor. Es común en consolidados, informes mensuales y reutilización de apartados.

**Acción correcta:** compare el contenido antes de decidir. La similitud alta puede esconder cambios importantes en conclusiones, cifras o responsables.



**Opciones avanzadas**  
Calibración, OCR, exclusiones, auditoría y umbrales por tipo.

**PDF**  
☒ Aplicar "sin portada" en PDF  
Páginas a omitir (1, 2, N)  
1  
Si se deja en 1, se aplica una heurística opcional para omitir 2 cuando la primera página tiene muy poco texto.

**DOCX**  
☒ Aplicar "sin portada" en DOCX  
Umbral (fallback): omitir % inicial si no se detecta "Introducción/Capítulo"  
8  
Ej: 8 = omitir ~8% inicial (min. 800 y máx. 4000 caracteres).

☐ Normalización agresiva (quitar puntuación)  
Mejora coincidencia cuando cambian comas/puntos, pero puede sobre-unificar.

**Parámetros de similitud (calibración)**  
k-shingles (k): 5  
Firmas MinHash (#): 64  
Limite muestreo (shingles máx.): 120000  
Limite comparador (shingles máx.): 50000  
A mayor k y más firmas, más precisión pero más costo. Ajusta muestreo para controlar rendimiento en lotes grandes.

**Exclusiones y análisis avanzado**  
Excluir extensiones: tmp, bak, log  
Ignorar menores de (KB): 0  
Excluir si el nombre contiene: copia temporal, --\$  
Excluir rutas/carpetas: node\_modules, backup, temporal  
Original sugerido: Más reciente  
Ruta/carpetas preferida: originales, documentos/final  
☒ Excluir archivos temporales comunes (-\$, .tmp, .bak)  
☒ OCR automático en PDF escaneado o con poco texto  
☐ OCR automático en imágenes (más lento)  
☒ Ignorar encabezados, pies, numeración y texto repetitivo

**Umbrales automáticos por tipo**  
☒ Usar umbrales distintos según el tipo de archivo  
Texto: 80  
PDF: 82  
Imagen: 90  
Excel: 82

Ejemplo demostrativo de posible versión



## Qué es una familia

Una agrupación de documentos que presentan señales de contenido equivalente o relación suficiente para revisarlos como conjunto.

- ✓ Versiones sucesivas del mismo informe.
- ✓ Documentos exportados a varios formatos.
- ✓ Copias parciales que comparten una base.
- ✓ Archivos de una misma plantilla con contenido altamente relacionado.

## Cómo usarla

- 1 Abra Familias.
- 2 Revise los nombres y rutas.
- 3 Identifique cuál archivo parece origen o versión más completa.
- 4 Abra comparaciones individuales cuando exista duda.
- 5 Conserve una estructura documental clara, no solo el archivo "más nuevo".



## Para documentos textuales

- ✓ Muestra el archivo A y B.
- ✓ Resalta adiciones, eliminaciones y similitudes.
- ✓ Puede identificar páginas aproximadas coincidentes.
- ✓ Usa comparación por palabras y bloques.
- ✓ Es especialmente útil en PDF, DOCX y texto.

## Cómo interpretar

- ✓ Verde / añadido: contenido presente en una versión.
- ✓ Rojo / eliminado: contenido que desapareció.
- ✓ Similitud: fragmentos equivalentes.
- ✓ No se limite al porcentaje; revise el significado del cambio.

**Pregunta clave:** ¿los cambios son meramente de formato o modifican una decisión, cifra, compromiso, responsable o conclusión?

| Tipo            | Señales adicionales                                                       |
|-----------------|---------------------------------------------------------------------------|
| Imagen          | dHash, pHash, regiones y análisis visual según el modo.                   |
| PPTX            | Texto, estructura y comparación visual simplificada de diapositivas.      |
| Audio           | Señales de similitud multimedia cuando el modo lo permite.                |
| Video           | Comparación multimedia selectiva en análisis profundo.                    |
| Binario / otros | Comparación binaria cuando no puede extraerse contenido semántico fiable. |

**Precaución:** en multimedia el costo computacional aumenta. Divida el lote y utilice Profundo solo con los archivos candidatos que realmente necesita revisar.



## Señales técnicas posibles

- ✓ Texto.
- ✓ Bloques.
- ✓ Contención.
- ✓ Penalización de plantilla.
- ✓ Estructura PDF.
- ✓ Visual PPTX.
- ✓ dHash / pHash.
- ✓ OCR, binario, audio o video.

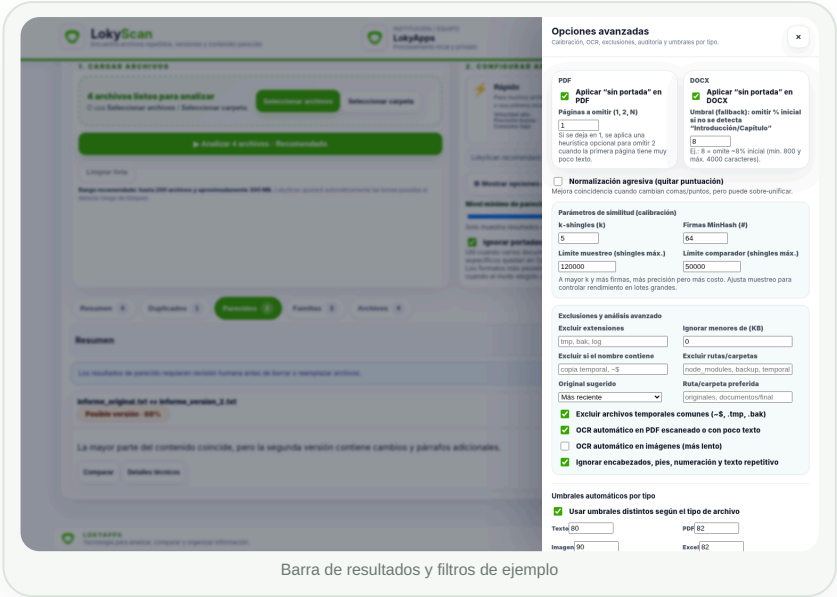
## Principio de decisión

Un porcentaje alto indica que las señales disponibles son muy parecidas; no determina por sí solo cuál documento es oficial, vigente o jurídicamente relevante.

## Revisar

antes de borrar, sobrescribir o mover.

| Herramienta       | Uso                                                         |
|-------------------|-------------------------------------------------------------|
| Pestañas          | Resumen, Duplicados, Parecidos, Familias y Archivos.        |
| Chips             | Versiones, copias parciales, confianza alta, ≥90 %.         |
| Relación          | Filtra por tipo de relación detectada.                      |
| Orden             | Mayor parecido o nombre A-Z.                                |
| Buscar            | Localiza nombres o elementos dentro de resultados.          |
| Detalles técnicos | Muestra hashes, desglose de puntuación y datos adicionales. |





## Qué permite

Guardar un manifiesto de referencia y compararlo con una revisión posterior para identificar archivos sin cambios, modificados, nuevos o eliminados.

- ✓ Usar el último manifiesto guardado.
- ✓ Importar un manifiesto JSON.
- ✓ Exportar una nueva referencia.
- ✓ Revisar diferencias en el resumen.

## Cuándo es útil

- ✓ Auditorías periódicas de carpetas.
- ✓ Control de cambios en repositorios compartidos.
- ✓ Seguimiento de entregables.
- ✓ Verificación antes y después de una depuración.

**Buena práctica:** exporte el manifiesto antes de limpiar una carpeta importante. Le servirá como evidencia de qué archivos existían en ese momento.



| Formato | Uso recomendado                                                       |
|---------|-----------------------------------------------------------------------|
| PDF     | Entregar un informe legible con conclusiones y resultados.            |
| CSV     | Analizar tablas en Excel, Power Query o herramientas BI.              |
| JSON    | Conservar estructura completa para auditoría o procesamiento técnico. |

**Antes de exportar:** aplique filtros si desea revisar un subconjunto, pero conserve también una exportación completa cuando el análisis sea parte de un proceso formal.

**PDF**

Ideal para compartir con responsables y documentar decisiones.

**CSV**

Útil para filtrar, ordenar o cruzar resultados.

**JSON**

Adecuado para respaldo técnico y automatización.

**Nuevo análisis**

Limpia el contexto actual y permite iniciar otro lote.



## Depuración segura

- 1 Exporte el informe.
- 2 Revise duplicados exactos.
- 3 Confirme cuál copia conservar.
- 4 Mueva candidatos a una carpeta temporal antes de borrar definitivamente.
- 5 Abra los documentos críticos después del movimiento.
- 6 Elimine solo cuando haya validado la copia conservada.

## Problemas frecuentes

| Síntoma                       | Qué hacer                                   |
|-------------------------------|---------------------------------------------|
| El análisis tarda demasiado   | Divida el lote o use Rápido/Recomendado.    |
| Pocos resultados              | Baje el umbral con cautela o use Profundo.  |
| Demasiados parecidos          | Ignore portadas/boilerplate y eleve umbral. |
| PDF escaneado sin texto       | Active OCR PDF.                             |
| Pestaña consume mucha memoria | Evite ZIP/OCR masivo; trabaje por sublotes. |



## Ejemplo de principio a fin

- 1 Cargue 4 documentos: original, copia exacta, versión actualizada y resumen parcial.
- 2 Use Recomendado con umbral 80 %.
- 3 Ejecute el análisis.
- 4 Revise Resumen y confirme el duplicado exacto.
- 5 Compare la posible versión antes de decidir.
- 6 Exporte PDF del análisis.
- 7 Mueva la copia exacta a una carpeta temporal y valide el original.

## Checklist antes de terminar

- ✓ El lote analizado corresponde a la carpeta correcta.
- ✓ Se revisaron los duplicados exactos.
- ✓ Las versiones fueron comparadas, no borradas por porcentaje.
- ✓ Se verificó la ruta de la copia a conservar.
- ✓ Se exportó al menos un informe.
- ✓ Los archivos críticos tienen respaldo.
- ✓ La depuración final se hizo después de validar.

**Resultado esperado:** menos duplicados y una estructura documental más clara, sin perder versiones importantes ni evidencia de trabajo.