

Digitalizador documental (OCR local)

Convierte documentos escaneados en texto buscable y editable — sin conexión a internet

PROTOTIPO

Python · Transformers · PyMuPDF · 100% local

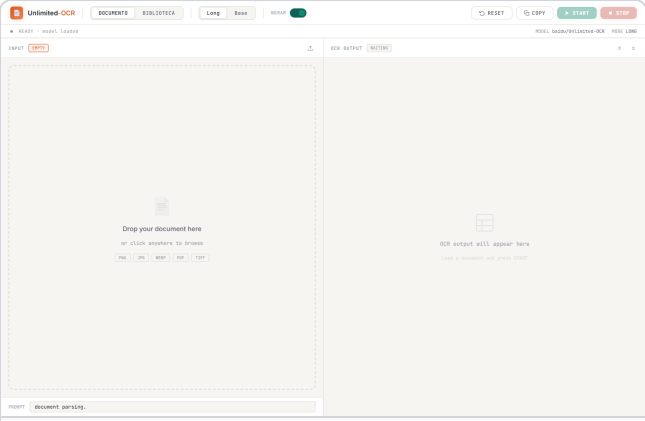
QUÉ HACE

Recibe un PDF escaneado y decide, página a página, si ya contiene texto digital o si necesita reconocimiento óptico de caracteres. El resultado es un documento con el mismo contenido, pero con todo el texto ya buscable, copiable y exportable.

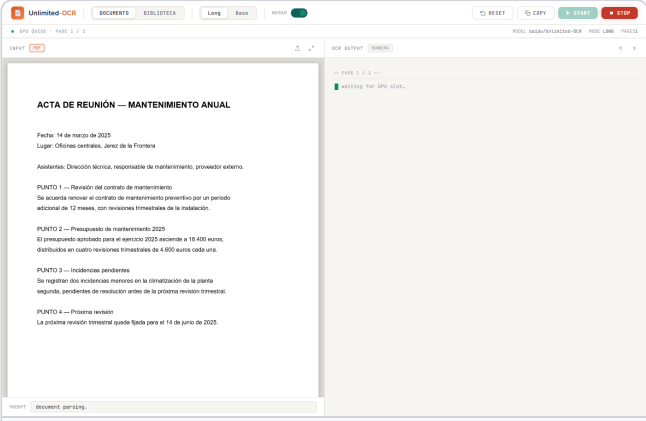
Cada página reconocida queda marcada internamente con su número de origen, de forma que cualquier búsqueda o cita posterior sobre ese texto apunta siempre al lugar exacto del documento original — esto es lo que permite, por ejemplo, que una herramienta de consulta documental (ver RagDesk) pueda citar «página 14» con total fiabilidad.

Cuando el sistema no puede leer un fragmento con suficiente confianza, lo señala explícitamente en vez de rellenarlo con una suposición — una decisión deliberada para no introducir errores silenciosos en un documento que después alguien va a dar por bueno.

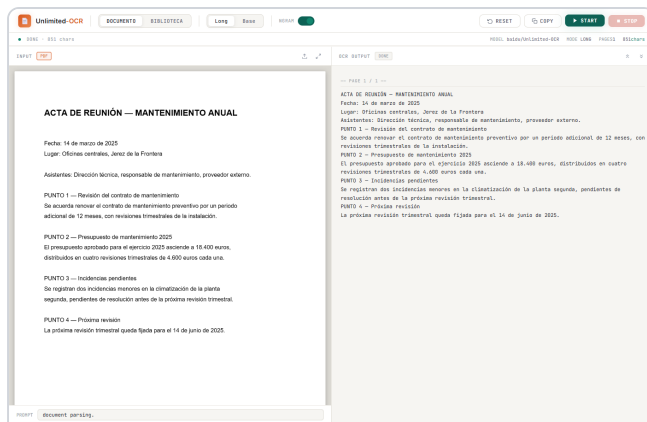
CÓMO SE VE EN USO



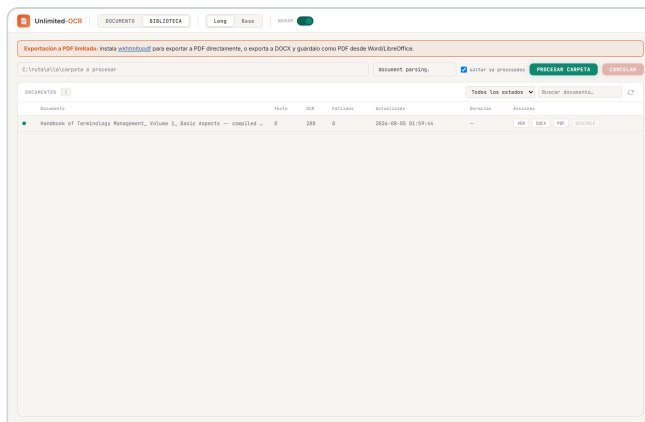
1. Punto de partida — un PDF escaneado, sin texto digital.



2. Procesamiento — cada página se analiza y digitaliza en local.



3. Resultado — el texto queda buscable y verificable página a página.



4. Exportación — el documento final se exporta ya listo para usar.

SERVICIOS QUE PUEDO OFRECER CON ESTO

- **Digitalización de archivo histórico o administrativo** — actas, expedientes, escrituras — para que quede buscable en texto.
- **Migración puntual de papel a formato digital** para despachos, gestorías y notarías.
- **Primer paso de un flujo mayor de consulta documental:** una vez digitalizado, el archivo puede alimentar un asistente de búsqueda en lenguaje natural (ver el manual de RagDesk).