MarkItDown: menos tokens con tus PDFs

@StephanoBrooke · Palabra clave: TOKEN

Lo que necesitas

- Claude Desktop (la app de escritorio), porque aquí los servidores MCP se agregan con un archivo de configuración.
- Python 3.10 o más nuevo (hasta la 3.13) instalado. Si no sabes si lo tienes, el prompt 01 le pide a Claude que te lo explique.
- Un PDF, Word o Excel tuyo para probar. Mejor uno con tablas o varias páginas.
- Unos 5 a 15 minutos y paciencia con la terminal (Terminal en Mac, PowerShell en Windows). Claude te guía.

Qué problema resuelve

Cuando subes un PDF, Claude no lee solo el texto limpio. Según cómo esté hecho el archivo, procesa también el formato, las tablas y una imagen de cada página. Eso puede costar mucho más espacio (tokens) del que ocupa lo que dice el documento.

- MarkItDown es una herramienta gratis y de código abierto de Microsoft. Convierte PDF, Word, Excel, PowerPoint y más a Markdown, que es texto limpio con títulos y listas.
- El texto en Markdown suele ocupar menos espacio y es más fácil de leer para un modelo de IA.
- Con un servidor MCP, Claude puede usar MarkItDown como una herramienta propia: tú le pides que convierta un archivo y lo hace.

Tips:

- No te prometo un porcentaje de ahorro: depende de tu archivo. En la sección de prueba mides tú el resultado.

Paso 1: deja que Claude te guíe

Empieza con un chat nuevo y pega esto:

Prompt 01: Guía de instalación a tu medida:

> *"Quiero instalar el servidor MCP de MarkItDown (el paquete markitdown-mcp de Microsoft) en Claude Desktop. Primero pregúntame si uso Mac o Windows y si ya tengo Python. Después guíame paso a paso, uno a la vez: instalar el paquete, abrir el archivo claude_desktop_config.json y qué pegar. Espera mi 'listo' antes de pasar al siguiente paso."*

Qué pasa después: Claude te pregunta tu sistema y te va dando un paso por mensaje.

Paso 2: instalar MarkItDown MCP

Estos son los pasos de fondo. Claude te los explica con tu sistema.

Prompt 02: Pedir el texto exacto de configuración:

> *"Dame el texto exacto que debo pegar en claude_desktop_config.json para agregar el servidor markitdown-mcp (el comando es markitdown-mcp). Si ya tengo otros servidores en el archivo, explícame cómo agregarlo sin borrarlos y revisa que las comas queden bien."*

Abre la terminal e instala el paquete con: pip install markitdown-mcp

En Claude Desktop abre Ajustes, Desarrollador, Editar configuración. Se abre el archivo claude_desktop_config.json.

Agrega el servidor de MarkItDown (el prompt 02 te da el texto exacto) y guarda el archivo.

Cierra Claude Desktop por completo y ábrelo otra vez.

Qué pasa después: Al volver a abrir Claude Desktop, el servidor aparece en las herramientas disponibles.

Tips:

- Existe otra forma oficial, con Docker, que Microsoft recomienda para Claude Desktop. Si ya usas Docker, el prompt 01 también puede guiarte con esa.
- Si algo falla, copia el mensaje de error y pégalo en Claude: el prompt 03 le pide que lo interprete.

Paso 3: pruébalo con tu archivo

El servidor tiene una herramienta (convert_to_markdown) que recibe la dirección del archivo. Tú le dices a Claude cuál usar:

Prompt 03: Si algo no funciona:

> *"El servidor MarkItDown no aparece o da este error: [pega el error]. Dime qué significa en palabras simples y qué reviso, empezando por lo más probable."*

Prompt 04: Convertir y resumir un archivo:

> *"Usa la herramienta de MarkItDown para convertir a Markdown este archivo: file:///[ruta completa de tu archivo, por ejemplo /Users/tu_nombre/Documents/contrato.pdf]. Luego resúmelo en 5 puntos y dime si las tablas se leyeron completas."*

Prompt 05: Medir si valió la pena:

> *"Con el archivo ya convertido, dime cuántas palabras tiene el Markdown y si perdiste información importante respecto al original (tablas, números, notas al pie). Si algo no se leyó bien, dímelo."*

Qué pasa después: Claude convierte el archivo, lo lee en texto limpio y te dice si algo se perdió.

Tips:

- Los PDF escaneados (fotos de hojas) no tienen texto: MarkItDown no los lee bien sin un paso extra de reconocimiento de texto. Si el resultado sale vacío, probablemente es un escaneo.
- Claude llama a la herramienta cuando tú se lo pides. No convierte solo cada archivo que subes al chat: dale la ruta del archivo en tu computadora.
- Para comparar de verdad, sube el mismo PDF directo en otro chat y pregunta lo mismo. Así ves la diferencia en tu caso.

Sin MCP: la forma más simple

Si el MCP te parece demasiado, puedes convertir tú el archivo y subir el resultado:

Instala la herramienta con: pip install 'markitdown[all]'

Convierte con: markitdown archivo.pdf -o archivo.md

Sube el archivo .md al chat de Claude en lugar del PDF.

Qué pasa después: Tienes el mismo texto limpio sin tocar la configuración de Claude Desktop.

Prompts bonus (1 de 2)

Mismo método, para sacarle más a MarkItDown. Copia el prompt y cambia lo que va entre corchetes.

Bonus 01: Convertir varios archivos y compararlos:

> *"Usa MarkItDown para convertir estos tres archivos: file:///[ruta 1], file:///[ruta 2], file:///[ruta 3]. Después compáralos en una tabla: qué dice cada uno, en qué coinciden y en qué se contradicen."*

Bonus 02: Extraer datos de una tabla:

> *"Convierte este archivo con MarkItDown: file:///[ruta]. Saca las tablas y entrégamelas como una tabla limpia. Si alguna celda no se leyó bien, márcala."*

Bonus 03: Revisar un contrato:

> *"Convierte este contrato con MarkItDown: file:///[ruta]. Hazme una lista de fechas, montos, obligaciones y condiciones para cancelar. Marca lo que no entiendas para que yo lo pregunte a un abogado."*

Bonus 04: Resumir una presentación:

> *"Convierte esta presentación de PowerPoint con MarkItDown: file:///[ruta]. Dame el mensaje principal de cada diapositiva en una línea."*

Prompts bonus (2 de 2)

Bonus 05: Preparar apuntes de estudio:

> *"Convierte este documento con MarkItDown: file:///[ruta]. Hazme 10 preguntas de repaso con su respuesta y marca en qué parte del documento está cada una."*

Bonus 06: Limpiar una hoja de Excel grande:

> *"Convierte este Excel con MarkItDown: file:///[ruta]. Describe las columnas, detecta filas repetidas o vacías y dime qué revisaría antes de usar estos datos."*

Bonus 07: Revisar tus propios archivos antes de subirlos:

> *"Antes de seguir, dime qué archivos de mi carpeta [ruta] valdría la pena convertir a Markdown primero por ser muy largos o tener muchas tablas, y por qué."*

La parte honesta

- El reel original dice que MarkItDown 'reduce muchísimo el gasto de tokens'. Puede ayudar, pero depende de tu archivo y no hay un número que te pueda garantizar. Mide tú con el prompt 05 y con una comparación contra subir el PDF directo.
- En el reel se dice que el servidor convierte cada archivo que subes automáticamente. En la práctica, Claude usa la herramienta cuando se lo pides y tú le das la ruta del archivo. No es automático para todo lo que subes al chat.
- Los PDF escaneados (fotos de hojas) salen vacíos o mal. Para esos necesitas un paso de reconocimiento de texto antes.
- Una conversión puede perder cosas: tablas complejas, notas al pie, gráficas. Revisa los números y las fechas importantes en el archivo original antes de decidir algo.
- No conviertas archivos con contraseñas, datos de identidad o información que no quieres que salga de tu computadora. El servidor corre en tu máquina, pero lo que Claude lee queda en la conversación.
- MarkItDown tiene más de 100 mil estrellas en GitHub, según los rastreadores de mediados de 2026. No uses esa cifra como prueba de que es mejor para ti: pruébalo con tus archivos.
- Los pasos de instalación vienen de la documentación oficial de Microsoft y pueden cambiar. Si algo no coincide, revisa el repositorio de markitdown en GitHub o pídele a Claude que te guíe.

Cómo usarlos

1. Pega el prompt 01 en un chat nuevo y deja que Claude te guíe con tu sistema.
2. Instala markitdown-mcp y agrega el servidor con el texto del prompt 02. Reinicia Claude Desktop.
3. Si hay un error, pega el prompt 03 con el mensaje.
4. Prueba con tu archivo usando el prompt 04.
5. Mide el resultado con el prompt 05 y compáralo con subir el PDF directo.
6. Si te gusta, úsalo con los prompts bonus. Si no, usa la forma simple: convertir y subir el .md.