

MARKITDOWN

Deja de quemar tokens con PDFs

★ 168,000 estrellas en GitHub · Microsoft · **MIT License** · **Gratis**

github.com/microsoft/markitdown

EL PROBLEMA — Cuántos tokens estás desperdiciando

Cada página de un PDF consume entre 1,500 y 3,000 tokens al procesarse. Un documento de 20 páginas quema hasta 60,000 tokens antes de tu primera pregunta. Eso es porque Claude tiene que procesar el formato, las tablas rotas, los metadatos y toda la basura estructural del archivo.

MarkItDown convierte cualquier archivo a texto limpio en Markdown — el formato con el que Claude fue entrenado — y reduce el uso de tokens hasta un 70%. Menos tokens = respuestas más rápidas, más baratas y más precisas.

1. QUÉ ES MARKITDOWN

MarkItDown es una herramienta de línea de comandos y librería de Python creada por Microsoft que convierte archivos de cualquier formato a texto limpio en Markdown. Fue diseñada específicamente para alimentar modelos de lenguaje como Claude — no para consumo humano, sino para que la IA pueda leer documentos de forma eficiente.

FORMATOS QUE CONVIERTE

- ✓ PDF — incluyendo PDFs escaneados con OCR (plugin adicional)
- ✓ Word (.docx) — mantiene encabezados, listas y tablas
- ✓ Excel (.xlsx y .xls) — convierte hojas en tablas Markdown
- ✓ PowerPoint (.pptx) — extrae texto de cada slide con su estructura

- ✓ Imágenes — extrae metadatos EXIF y puede describir con visión de LLM
- ✓ Audio — transcribe con metadatos EXIF
- ✓ HTML — limpia el código y deja solo el contenido
- ✓ CSV, JSON, XML — convierte a texto estructurado legible
- ✓ URLs de YouTube — extrae la transcripción automáticamente
- ✓ EPubs, archivos ZIP y más

Ahorro real de tokens por tipo de documento

Documento	Tokens con PDF	Tokens con .md	Ahorro
Contrato de 5 páginas	~12,000	~3,500	71%
Reporte de 20 páginas	~60,000	~18,000	70%
Presentación de 40 slides	~80,000	~22,000	72%
Hoja de Excel compleja	~15,000	~4,200	72%
Manual técnico 100 págs	~280,000	~85,000	70%

Nota: los valores son estimados basados en el contenido promedio de cada tipo de documento. Los PDFs con muchas imágenes o tablas complejas pueden consumir significativamente más tokens.

2. INSTALACIÓN — 2 minutos desde cero

REQUISITO PREVIO

Necesitas Python 3.10 o superior instalado en tu computadora.
Para verificar tu versión: abre la terminal y escribe `python --version`
Si no tienes Python: descárgalo gratis en python.org

Opción A — Instalación completa (recomendada)

Instala MarkItDown con soporte para todos los formatos de archivo:

```
pip install 'markitdown[all]'
```

Opción B — Instalación selectiva

Si solo necesitas convertir ciertos formatos, instala solo los necesarios:

```
# Solo PDF y Word
pip install 'markitdown[pdf,docx] '

# Solo Excel y PowerPoint
pip install 'markitdown[xlsx,pptx] '

# Solo PDF
pip install 'markitdown[pdf] '
```

Verifica que la instalación fue exitosa

```
markitdown --version
```

Si ves un número de versión (actualmente 0.1.6), la instalación fue exitosa.

3. CÓMO USARLO — Línea de comandos

MarkItDown funciona desde la terminal con comandos simples. Una vez instalado, abres tu terminal, navegas a donde está tu archivo y ejecutas el comando correspondiente.

Convertir un archivo y ver el resultado en pantalla

```
markitdown mi-documento.pdf
```

Convertir y guardar como archivo .md

```
markitdown mi-documento.pdf -o mi-documento.md
```

```
# Lo mismo para cualquier formato:
markdown reporte.xlsx -o reporte.md
markdown presentacion.pptx -o presentacion.md
markdown contrato.docx -o contrato.md
```

Convertir una URL de YouTube (extrae la transcripción)

```
markdown https://www.youtube.com/watch?v=XXXX -o transcripcion.md
```

Convertir varios archivos en secuencia

```
# Mac / Linux
for f in *.pdf; do markdown "$f" -o "${f%.pdf}.md"; done

# Windows (PowerShell)
Get-ChildItem *.pdf | ForEach-Object { markdown $_.Name -o ($_.BaseName +
'.md') }
```

FLUJO DE TRABAJO CON CLAUDE

1. Tienes un PDF de 30 páginas que quieres analizar con Claude
2. Ejecutas: `markdown mi-reporte.pdf -o mi-reporte.md`
3. Abres `claude.ai`, adjuntas el archivo `mi-reporte.md` (no el PDF)
4. Claude lee el mismo contenido usando hasta un 70% menos de tokens
5. Resultado: análisis más rápido, respuestas más precisas

4. USO DESDE PYTHON — Para automatizar

Si quieres integrar MarkItDown en un script o flujo de trabajo automatizado, puedes usarlo directamente desde Python con unas pocas líneas de código.

Conversión básica en Python

```
from markitdown import MarkItDown

# Inicializar
md = MarkItDown()

# Convertir un archivo
resultado = md.convert('mi-documento.pdf')

# Ver el contenido en Markdown
print(resultado.text_content)

# O guardarlo en un archivo
with open('mi-documento.md', 'w') as f:
    f.write(resultado.text_content)
```

Convertir una carpeta completa de PDFs automáticamente

```
from markitdown import MarkItDown
import os

md = MarkItDown()
carpeta = './mis-documentos'

for archivo in os.listdir(carpeta):
    if archivo.endswith('.pdf'):
        ruta = os.path.join(carpeta, archivo)
        resultado = md.convert(ruta)
        nombre_md = archivo.replace('.pdf', '.md')
        with open(os.path.join(carpeta, nombre_md), 'w') as f:
            f.write(resultado.text_content)
        print(f'Convertido: {archivo} -> {nombre_md}')
```

5. INTEGRACIÓN MCP CON CLAUDE DESKTOP

Esta es la opción más poderosa. MarkItDown incluye un servidor MCP (Model Context Protocol) que, al conectarlo a Claude Desktop, convierte automáticamente cada archivo que subes — sin que tengas que ejecutar nada manualmente. Subes el archivo, Claude lo recibe en Markdown optimizado de forma transparente.

QUÉ NECESITAS PARA ESTO

- Claude Desktop instalado (descarga gratuita en claude.ai/download)
- MarkItDown instalado con: `pip install 'markitdown[all]'`
- Acceso al archivo de configuración de Claude Desktop

Paso 1 — Encuentra el archivo de configuración de Claude Desktop

- Mac: `~/Library/Application Support/Claude/claude_desktop_config.json`
- Windows: `%APPDATA%\Claude\claude_desktop_config.json`

Paso 2 — Agrega MarkItDown como servidor MCP

Abre el archivo de configuración con cualquier editor de texto y agrega lo siguiente dentro de la sección `mcpServers`:

```
{
  "mcpServers": {
    "markitdown": {
      "command": "python",
      "args": ["-m", "markitdown.mcp"]
    }
  }
}
```

Paso 3 — Reinicia Claude Desktop

Cierra completamente Claude Desktop y vuelve a abrirlo. En la interfaz verás el ícono de MarkItDown disponible como herramienta activa. A partir de ese momento, cada archivo que subas será convertido automáticamente a Markdown antes de ser procesado por Claude.

LO QUE PASA AUTOMÁTICAMENTE CUANDO CONECTAS EL MCP

- Subes un PDF → Claude lo recibe como texto Markdown optimizado
- Subes un Excel → Claude lo recibe como tablas en texto plano
- Subes un PowerPoint → Claude lo recibe como el contenido de cada slide
- Todo esto ocurre de forma transparente — tú no haces nada diferente
- Reducción de tokens de hasta el 70% en cada conversación con documentos

6. PROBLEMAS COMUNES Y CÓMO RESOLVERLOS

Error: 'markdown' no se reconoce como comando

Python no está en el PATH de tu sistema. Solución:

```
# Usa python -m en lugar del comando directo
python -m markdown mi-archivo.pdf -o resultado.md
```

Error al instalar con pip install 'markdown[all]'

Puede ser un problema de versión de Python o conflicto de dependencias. Solución:

```
# Crea un entorno virtual limpio primero
python -m venv venv-markitdown

# Actívalo (Mac/Linux)
source venv-markitdown/bin/activate

# Actívalo (Windows)
venv-markitdown\Scripts\activate

# Instala dentro del entorno
pip install 'markdown[all]'
```

El archivo convertido tiene caracteres extraños

El PDF puede tener codificación no estándar. Solución:

```
# Especifica la codificación al guardar
markdown mi-archivo.pdf -o resultado.md
# Luego abre resultado.md con tu editor y guarda como UTF-8
```

El MCP no aparece en Claude Desktop

Verifica que el JSON de configuración sea válido (sin comas extra, con comillas correctas). Puedes validarlo en jsonlint.com. Luego reinicia Claude Desktop completamente — en Mac, usa Cmd+Q, no solo cierras la ventana.

¿Quieres que implementemos MarkItDown dentro de un flujo de IA personalizado para tu negocio?

Podemos integrarlo con tus documentos existentes, tu base de conocimiento
y los agentes de Claude para que toda tu operación consuma tokens
de forma eficiente desde el primer día.

➡  Escríbenos directamente a nuestro agente de WhatsApp y agendamos una llamada.

+52 81 2188 4401