

# Guia Rápido de Uso

## Instalação Rápida

Bash

```
# 1. Instalar dependências
pip install pdfplumber openai

# 2. Configurar API OpenAI (se for usar IA)
export OPENAI_API_KEY="sua-chave-aqui"
```

## Uso Básico

### Comando Mais Simples

Bash

```
python extrator_cct_v2.py "sua_cct.pdf" -o "saida.csv"
```

### Sem IA (Mais Rápido)

Bash

```
python extrator_cct_v2.py "sua_cct.pdf" -o "saida.csv" --sem-ia
```

## Workflow Completo

### 1 Processar PDF

Bash

```
python extrator_cct_v2.py "CCT_Medicos_2025.pdf" -o "medicos_2025.csv"
```

### Saída esperada:

Plain Text



## EXTRATOR DE DADOS DE CCTs - VERSÃO 2



Extraindo texto do PDF: CCT\_Medicos\_2025.pdf

✓ Página 1/10 extraída

✓ Página 2/10 extraída

...

✓ Extração concluída: 25296 caracteres



Identificando sindicato e convenção...

Sindicato: SINDICATO DOS MÉDICOS x SINDICATO DAS SANTAS CASAS

Convenção: 2025-2026



Extraindo cláusulas...

Encontradas 29 cláusulas

✓ Cláusula 1/29: CLÁUSULA PRIMEIRA - VIGÊNCIA...

✓ Cláusula 2/29: CLÁUSULA SEGUNDA - ABRANGÊNCIA...

...

✓ Total de 29 cláusulas extraídas



Salvando dados em CSV: medicos\_2025.csv

✓ Arquivo CSV salvo com sucesso!

Total de linhas: 29

Tamanho do arquivo: 32.50 KB



PROCESSO CONCLUÍDO COM SUCESSO!

## 2 Verificar Resultado

Abra o arquivo CSV gerado em Excel, Google Sheets ou qualquer editor de planilhas.

### Verifique:

- ✓ Sindicato identificado corretamente
- ✓ Ano/período correto
- ✓ Todas as cláusulas capturadas
- ✓ Resumos fazem sentido
- ✓ Conteúdo completo está presente

### 3 Integrar com Planilha Mãe

#### Opção A: Copiar e Colar

1. Abra `medicos_2025.csv`
2. Selecione todas as linhas **exceto o cabeçalho**
3. Copie (Ctrl+C)
4. Abra `clausulas_farmaceuticos.csv`
5. Vá até a última linha
6. Cole (Ctrl+V)
7. Salve

#### Opção B: Concatenar via Comando

Bash

```
# Remove cabeçalho e adiciona ao final da planilha mãe  
tail -n +2 medicos_2025.csv >> clausulas_farmaceuticos.csv
```

### 4 Atualizar Streamlit

Bash

```
# Reinicie o aplicativo Streamlit  
streamlit run app.py
```

Pronto! As novas cláusulas estarão disponíveis no aplicativo.

## Exemplos de Comandos

### Processar Vários PDFs

Linux/Mac:

Bash

```
for pdf in *.pdf; do  
    echo "Processando: $pdf"  
    python extrator_cct_v2.py "$pdf" -o "${pdf%.pdf}.csv"  
done
```

Windows (PowerShell):

Plain Text

```
Get-ChildItem *.pdf | ForEach-Object {  
    Write-Host "Processando: $($_.Name)"  
    python extrator_cct_v2.py $_.FullName -o "$($_.BaseName).csv"  
}
```

## Usar Modelo Mais Rápido

Bash

```
python extrator_cct_v2.py "cct.pdf" -o "saida.csv" --modelo gpt-4.1-nano
```

## Processar Sem IA (Teste Rápido)

Bash

```
python extrator_cct_v2.py "cct.pdf" -o "saida.csv" --sem-ia
```

## Dicas Importantes

### ✓ Faça Sempre

- ✓ **Backup** da planilha mãe antes de adicionar dados
- ✓ **Revise** os dados extraídos antes de integrar
- ✓ **Teste** com `--sem-ia` primeiro para verificar extração
- ✓ **Mantenha** os PDFs originais como referência

### ✗ Evite

- ✗ Processar PDFs protegidos ou com senha
- ✗ Adicionar dados sem revisar
- ✗ Usar IA para testes rápidos (gasta créditos)
- ✗ Deletar PDFs originais após processamento

## Solução Rápida de Problemas

Problema	Solução
----------	---------

"Nenhuma cláusula encontrada"	Verifique se o PDF tem texto selecionável
"Erro de API OpenAI"	Use <code>--sem-ia</code> ou configure <code>OPENAI_API_KEY</code>
"Caracteres estranhos no CSV"	Abra com encoding UTF-8
"Poucas cláusulas extraídas"	PDF pode ter formatação não padrão

## Ajuda Rápida

Bash

```
# Ver todas as opções
python extrator_cct_v2.py --help

# Versão simples (sem IA)
python extrator_cct_v2.py arquivo.pdf -o saida.csv --sem-ia






# Versão completa (com IA)
python extrator_cct_v2.py arquivo.pdf -o saida.csv
```

## Estrutura do CSV Gerado

Plain Text

```
Sindicato | Convenção | Título da Cláusula | Resumo | Cláusula Completa
-----|-----|-----|-----|-----
SIND...   | 2025-2026 | CLÁUSULA PRIMEIRA | Define... | As partes...
```

## Próximos Passos

1.  Processar seu primeiro PDF
2.  Verificar o CSV gerado
3.  Integrar com a planilha mãe
4.  Testar no Streamlit
5.  Processar demais CCTs

**Dúvidas?** Consulte o `README.md` completo para mais detalhes.