

API — OCR & Procedimentos Médicos

Esta API processa PDFs e imagens de pedidos/exames médicos, realiza OCR e identifica procedimentos existentes dentro da tabela procedimentos.csv, além de registrar logs detalhados de decisão.

Principais Funcionalidades

- Upload de arquivos (PDF / imagens)
- Processamento assíncrono (Celery + Redis)
- OCR automático com EasyOCR
- Identificação por código, IA (embeddings) e heurísticas
- Logs de auditoria com rastreabilidade completa

Fluxo de Processamento

- Envio do arquivo → retorna job_id
- Worker executa OCR
- Matcher identifica procedimentos
- Resultado e logs podem ser consultados por job_id

Recursos Técnicos

- Python 3.10
- FastAPI
- Celery
- Redis
- EasyOCR
- Sentence Transformers (all-mpnet-base-v2)
- Faiss

Estrutura do CSV

Arquivo obrigatório: /app/data/procedimentos.csv

Colunas:
CODIGO
DESCRICAO

Delimitador detectado automaticamente (; ou ,).

Variáveis de Ambiente

```
PROCEDURES_CSV=/app/data/procedimentos.csv
DATA_DIR=/app/data
UPLOAD_DIR=/app/data/uploads
CELERY_BROKER_URL=redis://ocr_redis:6379/0
CELERY_RESULT_BACKEND=redis://ocr_redis:6379/1
MATCHER_THRESHOLD=0.88
```

Observações

Heurísticas e IA trabalham juntas:

- correção de quebras de linha do OCR
- regras inteligentes para siglas
- filtragem de palavras genéricas
- fallback via embeddings