

Argus OCR - README

Descrição Geral

O Argus é uma plataforma de análise documental inteligente para a Defensoria Pública do Distrito Federal (DPDF), integrando o sistema Solar DPDF com serviços cognitivos da Azure para análise automatizada de documentos processuais via OCR e IA. O sistema também suporta observabilidade (Langfuse), automação de workflows (N8N) e fluxos conversacionais (FlowiseAI).

Funcionalidades Principais

- Processamento e análise de documentos processuais (PDF, DOCX, etc.)
- Integração com Solar DPDF para busca e recuperação de processos e documentos
- OCR avançado com Azure Document Intelligence
- Análise inteligente de conteúdo com Azure OpenAI (GPT-4)
- Indexação e busca semântica com Azure Cognitive Search
- Observabilidade e rastreamento de operações de LLM com Langfuse
- Automação de workflows com N8N
- Construção de fluxos conversacionais com FlowiseAI
- Armazenamento seguro em Azure Storage

Módulos e Endpoints Principais

Integração com Solar DPDF

- GET /api/v1/solar/processos — Buscar processos disponíveis
- GET /api/v1/solar/processos/{processo_id} — Detalhes completos do processo
- GET /api/v1/solar/processos/{processo_id}/documentos — Listar documentos do processo
- GET /api/v1/solar/processos/{processo_id}/eventos — Histórico de eventos
- GET /api/v1/solar/processos/{processo_id}/partes — Identificar partes do processo

Processamento de Documentos

- POST /api/v1/documents/upload — Upload ou download de documentos
- POST /api/v1/ocr/analyze — Extração de texto com OCR
- POST /api/v1/ai/documents/analyze — Análise inteligente de documentos (resumo, pontos-chave, etc.)
- POST /api/v1/search/index — Indexação para busca semântica
- GET /api/v1/documents/{document_id}/export — Exportação de resultados (PDF, DOCX, JSON)

Integração com Langfuse

- Rastreamento de operações de LLM
- POST /api/v1/feedback — Avaliação de qualidade das análises

Integração com FlowiseAI

- Configuração de fluxos conversacionais
- POST /api/v1/chat/query — Chatbot de consulta processual

- POST /api/v1/drafts/generate — Geração de minutas com IA

Integração com N8N

- Automação de workflow de análise
- Notificações automáticas
- Integração com sistemas externos via webhooks

Processamento Assíncrono

- Processamento em background com Celery
- Status de tarefas: GET /api/v1/tasks/{task_id}/status

Estrutura do Projeto

```
argus-backend/  
├── main.py  
├── api/  
│   └── v1/  
│       ├── routes_ocr.py  
│       ├── routes_user.py  
│       └── ...  
├── services/  
│   ├── ocr_service.py  
│   ├── document_service.py  
│   └── ...  
├── repositories/  
│   ├── document_repository.py  
│   └── ...  
├── models/  
│   ├── document_model.py  
│   └── ...  
├── schemas/  
│   ├── document_schema.py  
│   └── ...  
├── core/  
│   ├── config.py  
│   ├── database.py  
│   └── security.py  
└── workers/  
    └── ocr_worker.py
```

Tecnologias Utilizadas

- FastAPI — Backend web
- SQLAlchemy — ORM
- PostgreSQL — Banco de dados
- Pydantic — Validação de dados
- Celery — Processamento assíncrono

- Azure Document Intelligence — OCR
- Azure OpenAI — IA para análise
- Azure Cognitive Search — Busca semântica
- Langfuse — Observabilidade de LLM
- N8N — Automação de workflows
- FlowiseAI — Fluxos conversacionais
- Azure Storage — Armazenamento seguro

Arquitetura de Integração

Frontend (React) → Argus Backend (FastAPI) → Solar DPDF, Azure Document Intelligence, Azure OpenAI, Azure Cognitive Search, PostgreSQL, Azure Storage, Langfuse, FlowiseAI, N8N

Como Executar

Subir tudo:

```
chmod +x ./run.sh  
./run.sh
```

Parar e limpar:

```
chmod +x ./stop.sh  
./stop.sh
```

Acessar endpoints de teste:

```
http://localhost:8000/docs  
http://localhost:8000/api/v1/health
```

Observações

- Sempre utilize paginação e filtros ao consumir endpoints do Solar DPDF devido ao grande volume de dados.
- O sistema é extensível para novas integrações e fluxos de automação.
- Consulte o documento de requisitos funcionais para detalhes completos de cada módulo.

Revision #4

Created 6 November 2025 13:11:02 by Carlos

Updated 6 November 2025 13:53:53 by Carlos