

Relatório do Projeto: Tour Extraction System

Nome do Grupo

Grupo 287

Integrantes do Grupo

Victor Hugo, Jander Alves, Gleice Clementino, Wagner Lemos, Bruno Leão, Wilson

Descrição do Tema Escolhido

O projeto desenvolvido, chamado Tour Extraction System, é um sistema inteligente que automatiza a extração de dados estruturados a partir de catálogos turísticos em PDF, usando agentes autônomos orquestrados pelo CrewAI. O sistema processa documentos complexos, realiza OCR e utiliza técnicas de RAG para identificar e extrair informações críticas, convertendo centenas de páginas em dados prontos para análise em poucos minutos.

Sugestão: Foque na clareza da transformação (de horas para minutos) e destaque a automação ponta a ponta, do upload do PDF até a exportação pronta para uso.

Público-alvo

- Times de Produto/Operação de marketplaces de turismo (como iFriend)
- Usuários finais que fazem uploads de PDFs e baixam arquivos estruturados
- Gestores de Dados/BI para análises estratégicas e precificação

Sugestão: Ressalte o uso por gestores de BI e operadores que passam a depender menos de tarefas manuais recorrentes.

Justificativa do Tema Escolhido

O sistema responde a uma dor operacional concreta da iFriend: processar manualmente dezenas de catálogos mensais é demorado, caro e propenso a erros (8% de erro humano). A automação entrega redução de mais de 90% no tempo (de 2-3 horas para ~18 minutos) e quase 100% de redução no custo operacional por documento, elevando a escalabilidade e qualidade dos dados.

A flexibilidade multilíngue e a atuação adaptativa dos agentes IA agregam elevado valor estratégico.

Sugestão: Suprimir detalhes secundários sobre alternativas técnicas não adotadas e manter apenas o racional da escolha do modelo principal.

Detalhamento do que foi desenvolvido

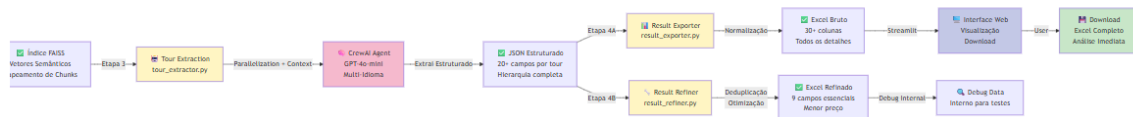
O sistema está estruturado em 4 etapas principais:

1. Pré-processamento do PDF: Usa Docling para ler, aplicar OCR e segmentar texto dos PDFs.
2. Indexação Semântica: Vetoriza o texto com Sentence Transformers e indexa no FAISS.
3. Extração Inteligente: Agentes via CrewAI interagem com o modelo GPT-4o-mini para extrair e estruturar os dados relevantes em JSON.
4. Exportação de Resultados: Utiliza Pandas para gerar arquivos JSON e Excel, prontos para integração e análise.

Operação: CLI e Streamlit para uploads, downloads e visualização de métricas.

Sugestão: Simplifique o detalhamento, priorize clareza, indique limitações atuais e avance os tópicos de evolução.

Elementos adicionais: Tabelas, Gráficos, Diagramas



Link para o repositório do GitHub

https://github.com/victorhss21/I2A2_trabalho_final/tree/main

Sugestão: README atualizado, onboarding facilitado, instruções de execução local/cloud, resumo visual do pipeline.

O que pode ser excluído ou reduzido

- Apêndices extensos de documentação técnica interna
- Comparativos detalhados de modelos desnecessários
- Tópicos genéricos ou repetidos

O que pode ser melhorado ou acrescentado

- Destaque visual dos resultados alcançados
- Depoimentos ou validação do parceiro/cliente
- Explicação leiga sobre como IA supera templates
- Detalhamento dos próximos passos de evolução