



UNIVERSIDADE FEDERAL DO RIO GRANDE DO NORTE
PROGRAMA DE PÓS GRADUAÇÃO EM ENGENHARIA ELÉTRICA E DE COMPUTAÇÃO

Projeto da Disciplina Tópicos Especiais – Aprendizado de Máquinas
Criação de Aplicação do tipo FAQ

Professor:
Ivanovitch Silva

Alunos:
Hilário José Silveira Castro
Walter Soares Barbosa Rocha Filho

Introdução

LangChain é um framework para desenvolvimento de aplicações usando Grandes Modelos de Linguagem (Large Language Model - LLMs). Com a combinação das LLMs com outros ativos é possível gerar aplicações que envolvem gerenciamento de prompt, encadeamento (chain), geração de dados aumentados (data argumentation), designação de sequência de operações, etc.

No contexto apresentado, o Langchain desempenha um papel fundamental na integração com a API do OpenAI para permitir que os usuários façam perguntas em linguagem natural sobre o conteúdo de arquivos .pdf. Ele oferece recursos específicos para lidar com a estrutura e o formato desses documentos, permitindo a extração de informações de maneira eficiente.

Objetivos

O objetivo deste projeto é mostrar como a combinação da API do OpenAI com a tecnologia Langchain possibilita aos usuários fazer perguntas em linguagem natural sobre o conteúdo de arquivos .pdf. Essa abordagem visa superar os desafios enfrentados na extração de informações estruturadas desses arquivos, que muitas vezes são complexos e não estão prontamente acessíveis.



Aspectos Negativos Enfrentados na Pesquisa

Embora a abordagem de usar a API do OpenAI em conjunto com Langchain apresente benefícios significativos, existem alguns pontos negativos a serem considerados. Um deles é o processamento lento quando se lida com grandes volumes de dados ou documentos complexos. Essa limitação pode afetar a eficiência do sistema em cenários de processamento intensivo.

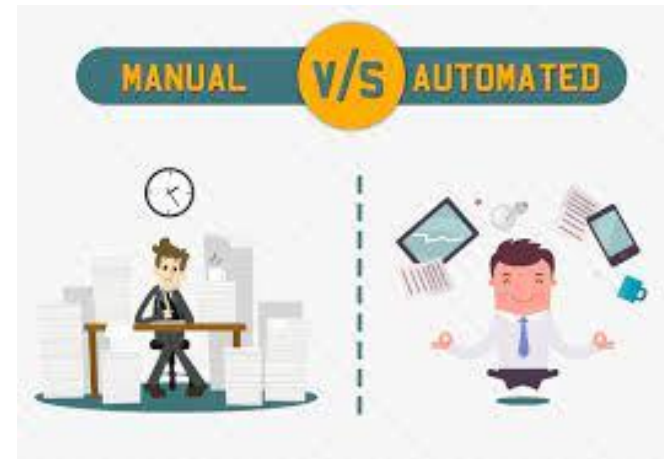
Além disso, a tecnologia atual pode apresentar dificuldades em extrair informações de arquivos .pdf com estruturas complexas ou layouts não padronizados. Esses desafios representam oportunidades para melhorias futuras, buscando aprimorar a eficiência e a capacidade de lidar com uma variedade mais ampla de documentos.



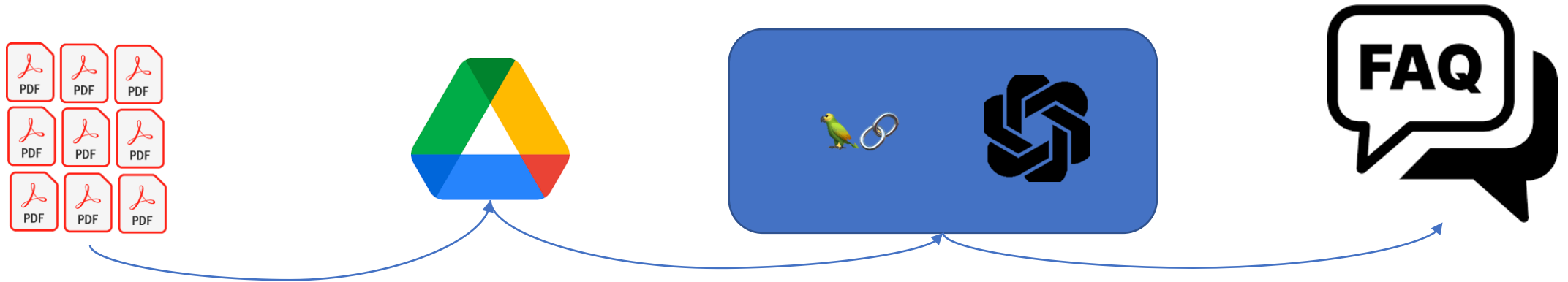
Aspectos Positivos Descobertos na Pesquisa

A integração da API do OpenAI e Langchain oferece uma série de pontos positivos. Primeiramente, a API do OpenAI fornece poderosas capacidades de processamento de linguagem natural, permitindo que as perguntas sejam formuladas de maneira natural e intuitiva. Por sua vez, a tecnologia Langchain é especializada no processamento de arquivos .pdf, facilitando a extração de informações relevantes desses documentos. Com essa combinação, os usuários podem automatizar tarefas de análise de documentos, economizando tempo e esforço.

Além disso, a capacidade de fazer perguntas em linguagem natural torna o processo mais acessível, eliminando a necessidade de conhecimentos técnicos avançados para interagir com os arquivos .pdf.



Processo de Desenvolvimento



Conclusão

Em suma, a combinação da API do OpenAI com Langchain para fazer perguntas em linguagem natural sobre arquivos .pdf oferece vantagens significativas, como a automação de processos e a facilidade de uso. No entanto, é importante reconhecer as limitações atuais, como o processamento lento e a dificuldade em extrair informações de documentos complexos. Esses pontos negativos apontam para áreas de pesquisa e desenvolvimento futuros, visando melhorar a eficiência e superar esses desafios.



