

GUIA RÁPIDO DE INÍCIO

⚡ 3 Passos para Começar

1 Configure a API Key do Gemini

Crie um arquivo `.env` na raiz do projeto:

```
GEMINI_API_KEY=sua_chave_aqui
```

🔑 Como obter a chave:

1. Acesse: <https://aistudio.google.com/app/apikey>
2. Faça login com sua conta Google
3. Clique em "Create API Key"
4. Copie a chave gerada

2 Execute o Projeto

```
npm run dev
```

3 Acesse no Navegador


Abra: **<http://localhost:3000>**

Como Usar a Pipeline

Etapa 1: Upload de PDF

1. Clique em "Escolher arquivos"
2. Selecione um ou mais PDFs
3. Aguarde a extração do texto

Etapa 2: Enriquecer com IA (Recomendado)

1. Clique em " Enriquecer com Gemini AI"
2. Aguarde o processamento (pode levar alguns minutos)
3. Veja as palavras-chave e classificações geradas

Etapa 3: Gerar Embeddings

Opção A (Melhor Qualidade):

- Selecione "Gemini text-embedding-004"

- Clique em "Gerar Embeddings"

Opção B (Mais Rápido):

- Selecione "Sentence-BERT" ou "USE"
- Clique em "Gerar Embeddings"

Etapa 4: Refinamento CNN (Opcional)

1. Configure os parâmetros:
 - Learning Rate: 0.005
 - Margin: 0.2
 - Mining Strategy: hard
 - Epochs: 15
2. Clique em "🔄 Treinar CNN"
3. Acompanhe as métricas de treino

Etapa 5: Clusterização

1. Clique em "Executar Clustering"
2. Visualize a distribuição dos vetores

Etapa 6: Construir Grafo

1. Clique em "Construir Grafo"
2. Explore o grafo interativo
3. Analise as métricas e clusters

Dicas

Para Melhores Resultados:

- ☒ Use documentos bem estruturados (PDFs de artigos, leis, livros)
- ☒ Sempre enriqueça com Gemini AI antes de gerar embeddings
- ☒ Use Gemini text-embedding-004 para máxima qualidade
- ☒ Configure a CNN com estratégia "hard" para melhor refinamento

Resolução de Problemas:

Erro "API Key inválida":

- Verifique se o arquivo `.env` existe
- Verifique se a chave está correta
- Reinicie o servidor (`Ctrl+C` e `npm run dev` novamente)

Processamento muito lento:

- Para documentos grandes, divida em partes menores
- Use embeddings locais (mais rápido, menor qualidade)

- Reduza o número de epochs da CNN

Erro ao processar PDF:

- Verifique se o PDF não está protegido por senha
- Tente converter o PDF para texto antes

Entendendo os Resultados

Métricas do Grafo:

- **Densidade (0-1):** Quanto maior, mais conectado o grafo
- **Grau Médio:** Número médio de conexões por nó
- **Modularidade (0-1):** Força da divisão em comunidades
- **Silhouette Score (-1 a 1):** Qualidade dos clusters (>0.5 = bom)

Análise de Clusters:

- Cada cluster representa um tópico ou tema
- Palavras-chave mostram o conteúdo principal
- Clusters similares compartilham conceitos

Casos de Uso

Análise de Documentos Acadêmicos

- Identificar temas principais
- Descobrir relações entre conceitos
- Mapear estrutura do conhecimento

Análise de Textos Jurídicos

- Identificar artigos relacionados
- Mapear hierarquia legal
- Encontrar precedentes conectados

Análise de Relatórios Técnicos

- Extrair conceitos-chave
- Identificar dependências
- Mapear fluxos de informação

Links Úteis

- **API do Gemini:** <https://ai.google.dev/>
- **Documentação do Projeto:** README.md
- **Melhorias Implementadas:** MELHORIAS_IMPLEMENTADAS.md

Pronto! Agora você está pronto para usar o GraphRAG Pipeline! 🐛