

# Document Comparator

Uma aplicação web desenvolvida em Streamlit para comparar documentos PDF e Word (.docx) e gerar relatórios detalhados das diferenças encontradas.

## Funcionalidades

- **Suporte a múltiplos formatos:** PDF e Word (.docx)
- **Comparação detalhada:** Identifica diferenças linha por linha
- **Tabela interativa:** Visualize as diferenças com número da página e linha
- **Filtros avançados:** Filtre por tipo de mudança e página/seção
- **Relatório HTML:** Baixe um relatório formatado das diferenças
- **Interface intuitiva:** Upload separado para cada arquivo
- **Performance otimizada:** Barras de progresso e processamento eficiente
- **Tratamento de erros:** Validação robusta de arquivos

## Como usar

### 1. Instalação Local

```
# Clone o repositório
git clone https://github.com/seu-usuario/document-comparator.git
cd document-comparator

# Instale as dependências
pip install -r requirements.txt

# Execute a aplicação
streamlit run app.py
```

### 2. Deploy no Streamlit Cloud

1. Faça fork deste repositório
2. Acesse [share.streamlit.io](https://share.streamlit.io)
3. Conecte sua conta GitHub
4. Selecione o repositório `document-comparator`
5. Configure:
6. **Main file path:** `app.py`
7. **Python version:** 3.11

8. Clique em "Deploy"

### 3. Usando a aplicação

**1. Upload dos arquivos:**

2. Faça upload do documento de referência

3. Faça upload do novo documento para comparação

**4. Comparação:**

5. Clique em "Comparar Documentos"

6. Aguarde o processamento

**7. Visualização:**

8. Veja o resumo das diferenças

9. Explore a tabela detalhada

10. Use os filtros para refinar os resultados

**11. Download:**

12. Baixe o relatório HTML formatado

13. Compartilhe ou imprima conforme necessário

## Requisitos

- Python 3.8+
- Streamlit 1.28.0+
- PyMuPDF 1.23.0+
- python-docx 0.8.11+
- pandas 2.0.0+

## Configuração

### Variáveis de ambiente (opcional)

Crie um arquivo `.streamlit/secrets.toml` para configurações personalizadas:

```
[general]
max_upload_size = 200 # MB
```

### [theme]

primaryColor = "#4CAF50"

backgroundColor = "#FFFFFF"

## Limitações

- **Tamanho máximo:** 200MB por arquivo
- **Formatos suportados:** PDF e DOCX
- **Comparação:** Apenas texto (imagens e formatação não são comparadas)
- **Word:** Documentos são divididos em seções de ~50 parágrafos

## Estrutura do Projeto

```
document-comparator/  
├── app.py          # Aplicação principal  
├── requirements.txt # Dependências Python  
├── .gitignore      # Arquivos ignorados pelo Git  
├── .streamlit/  
│   ├── config.toml # Configurações do Streamlit  
│   └── README.md   # Este arquivo
```

## Contribuindo

1. Faça fork do projeto
2. Crie uma branch para sua feature ( `git checkout -b feature/AmazingFeature` )
3. Commit suas mudanças ( `git commit -m 'Add some AmazingFeature'` )
4. Push para a branch ( `git push origin feature/AmazingFeature` )
5. Abra um Pull Request

## Changelog

### v1.0.0

- Suporte inicial para PDF e Word
- Interface Streamlit completa
- Relatórios HTML formatados
- Filtros avançados
- Tratamento robusto de erros

# Problemas Conhecidos

- Documentos Word muito grandes podem demorar para processar
- Formatação complexa pode não ser detectada corretamente
- Tabelas em documentos podem aparecer como texto contínuo

## Suporte

Se encontrar problemas ou tiver sugestões:

1. Verifique os [Issues existentes](#)
2. Crie um novo Issue com detalhes do problema
3. Inclua exemplos de arquivos (se possível)

## Licença

Este projeto está licenciado sob a MIT License - veja o arquivo [LICENSE](#) para detalhes.

## Agradecimentos

- [Streamlit](#) - Framework web
- [PyMuPDF](#) - Processamento de PDF
- [python-docx](#) - Processamento de Word
- [difflib](#) - Comparação de texto

---

**Desenvolvido com   usando Streamlit**