

Guia Técnico e de Contingência – Projeto BIA e PINE

1. Visão Geral do Projeto

O projeto consiste em dois módulos principais:

- **BIA (Dicionário de Dados)**

Gera um documento Word a partir de um recurso específico de um portal CKAN.

O documento final é baseado em um template Word e contém informações sobre o recurso, incluindo:

- Nome da base
- Descrição do recurso
- Lista de colunas e tipo de dados identificado (Numérico, Alfanumérico ou Texto)

- **PINE (Monitoramento)**

Faz a varredura de todos os datasets de um portal CKAN, coleta informações (nome, órgão, formatos dos recursos, datas) e atualiza uma planilha existente no Google Sheets usando a API do Google.

Ambos foram integrados em um aplicativo Streamlit (app.py), que permite escolher no menu lateral entre executar a BIA ou o PINE.

2. Estrutura de Arquivos e Pastas

Todos os arquivos do projeto devem estar na mesma pasta. Por exemplo:

pgsql

CopiarEditar

bia_pine_app/

└─ app.py

└─ bia.py

└─ pine.py

└─ modelo_bia2_pronto_para_preencher.docx

└─ credenciaisbiapine.json

└─ requirements.txt

Descrição dos arquivos:

- **app.py**
Interface Streamlit que une as funções da BIA e do PINE.
 - **bia.py**
Contém a função gerar_dicionario_word(recurso_url, template_path) responsável pela BIA.
 - **pine.py**
Contém a função atualizar_planilha(portal_url, verificar_urls) responsável pelo PINE.
 - **modelo_bia2_pronto_para_preencher.docx**
Template Word utilizado pela BIA.
É obrigatório que esteja presente na pasta.
 - **credenciaisbiapine.json**
Arquivo de credenciais do Google Service Account.
É obrigatório para o PINE conseguir editar a planilha.
 - **requirements.txt**
Lista de dependências Python necessárias para o projeto.
-

3. Dependências e Instalação

O projeto depende das seguintes bibliotecas:

nginx

CopiarEditar

streamlit

pandas

requests

python-docx

gsread

oauth2client

Para instalar:

1. Abra o terminal na pasta do projeto.
2. Execute:

nginx

CopiarEditar

pip install -r requirements.txt

4. Como Executar

Dentro da pasta do projeto:

arduino

CopiarEditar

python -m streamlit run app.py

O navegador abrirá em <http://localhost:8501>.

Use o menu lateral para:

- Aba BIA: inserir link do recurso CKAN e gerar dicionário.
 - Aba PINE: inserir link do portal CKAN e atualizar planilha.
-

5. Funcionamento Interno

BIA

1. Recebe o link de um recurso CKAN.
2. Extrai dataset_id e resource_id.
3. Acessa o CKAN via API e baixa dados do recurso.
4. Analisa cada coluna para determinar tipo:
 - Numérico: apenas números
 - Alfanumérico: mistura de letras e números
 - Texto: apenas texto
5. Preenche o template Word com:
 - Título do documento: DICIONARIO DE DADOS + nome da base limpo
 - Descrição do recurso
 - Tabela com colunas, tipos e descrições
6. Gera arquivo Word final e disponibiliza para download.

PINE

1. Recebe link do portal CKAN.
 2. Acessa a lista de datasets via API.
 3. Para cada dataset, coleta:
 - Nome da base
 - Órgão
 - Datas de criação e modificação
 - Quantidade de recursos por formato (CSV, XLSX, PDF, JSON)
 - Se habilitado, verifica URLs
 4. Monta um DataFrame com essas informações.
 5. Autentica no Google Sheets usando credenciaisbiapine.json.
 6. Limpa a aba definida (ex: Sheet1) e atualiza com os novos dados.
-

6. Pontos Críticos e Erros Comuns

1. **Arquivo ausente ou nome incorreto**
 - O template Word deve existir com o nome modelo_bia2_pronto_para_preencher.docx.
 - O arquivo de credenciais deve existir com o nome credenciaisbiapine.json.
2. **Permissões do Google Sheets**
 - O e-mail client_email do arquivo JSON deve estar adicionado como editor na planilha do Google Sheets.
 - O nome da planilha e da aba devem ser exatamente iguais aos usados no código:

ini

CopiarEditar

```
planilha = client.open("Indicadores PINE")
```

```
aba = planilha.worksheet("Sheet1")
```

Verifique se sua planilha tem exatamente esse nome e aba.

3. Erros de URL

- Para a BIA: o link deve ser o do recurso (contém /resource/).
- Para o PINE: o link deve ser do portal CKAN (por exemplo <https://dadosabertos.go.gov.br>).

4. Organização da pasta

- Todos os arquivos devem estar na mesma pasta ao rodar o app.py.
- Se renomear algum arquivo, atualize o nome correspondente no código.

5. Ambiente Python

- Certifique-se de rodar os comandos no mesmo ambiente onde instalou as dependências (pip install -r requirements.txt).
- Se aparecer mensagem de comando não reconhecido, use:

arduino

CopiarEditar

```
python -m streamlit run app.py
```

7. Recomendações

- Não mova ou renomeie os arquivos sem atualizar o código.
- Mantenha um backup do arquivo de credenciais JSON.
- Verifique sempre o nome exato da planilha e aba no Google Sheets.
- Se precisar trocar de portal CKAN, basta inserir o novo link no campo do aplicativo.
- Antes de qualquer alteração grande, teste em um ambiente separado ou faça cópia da pasta.

8. Procedimento em Caso de Erro

Caso algo não funcione:

1. Verifique se os arquivos obrigatórios ainda estão na pasta.
2. Verifique se a planilha existe e se o nome bate.
3. Verifique se a conta de serviço continua com permissão de edição.

4. Revise o terminal onde rodou o streamlit run app.py para mensagens de erro detalhadas.

5. Caso apareça mensagem de “não reconhecido” para streamlit, use:

arduino

CopiarEditar

```
python -m streamlit run app.py
```

6. Caso a instalação de dependências falhe, rode novamente:

nginx

CopiarEditar

```
pip install -r requirements.txt
```

9. Manutenção

- Sempre que for alterar algo no código, salve uma cópia dos arquivos originais.
- Teste o aplicativo após alterações pequenas, para identificar rapidamente onde ocorreu o problema.
- Atualize o requirements.txt caso adicione novas bibliotecas ao projeto.

GUIA PARA INICIAR E RODAR O PROJETO BIA E PINE

1. Estrutura dos arquivos

Antes de tudo, verifique se você tem esta estrutura na mesma pasta (exemplo: C:\Users\SeuNome\Desktop\bia_pine_app):

pgsql

CopiarEditar

bia_pine_app/

├─ app.py

├─ bia.py

├─ pine.py

└─ modelo_bia2_pronto_para_preencher.docx

└─ credenciaisbiapine.json

└─ requirements.txt

Todos esses arquivos devem estar juntos na mesma pasta.

2. Abrir o terminal na pasta do projeto

No Windows:

1. Abra a pasta bia_pine_app.
2. Segure a tecla **Shift** no teclado e clique com o **botão direito do mouse** em um espaço vazio da pasta.
3. No menu que aparece, clique em **Abrir janela do PowerShell aqui** ou **Abrir janela do Terminal aqui**.

Aparecerá uma janela preta ou azul com algo assim no caminho:

mathematica

CopiarEditar

PS C:\Users\SeuNome\Desktop\bia_pine_app>

3. Verificar se o Python está instalado

No terminal que abriu, digite:

css

CopiarEditar

python --version

e pressione **Enter**.

Se aparecer algo como:

nginx

CopiarEditar

Python 3.x.x

está tudo certo.

Caso dê erro, será necessário instalar o Python primeiro no seu computador.

4. Instalar as dependências

Ainda no terminal, digite:

```
nginx
```

CopiarEditar

```
pip install -r requirements.txt
```

e pressione **Enter**.

Esse comando vai instalar todas as bibliotecas necessárias (streamlit, pandas, requests, gspread, etc.).

Espere até finalizar. Pode demorar alguns minutos na primeira vez.

5. Executar o aplicativo

Depois que as dependências forem instaladas, execute o comando:

```
arduino
```

CopiarEditar

```
python -m streamlit run app.py
```

Espere alguns segundos.

Se aparecer uma mensagem pedindo e-mail, basta pressionar **Enter** (não é obrigatório).

O Streamlit irá inicializar e mostrar no final:

```
nginx
```

CopiarEditar

Local URL: <http://localhost:8501>

Seu navegador deve abrir sozinho.

Caso não abra, copie o endereço e cole no navegador.

6. Usar o aplicativo

Quando abrir no navegador:

- **Aba BIA – Dicionário de Dados**

Cole o link do recurso CKAN que você quer gerar o dicionário e clique em **Gerar**

Dicionário de Dados.

O sistema irá gerar um arquivo Word e você poderá baixar.

- **Aba PINE – Monitoramento**

Cole o link do portal CKAN (exemplo: <https://dadosabertos.go.gov.br>) e marque ou desmarque a opção de verificar URLs.

Clique em **Atualizar Planilha**.

A barra de progresso irá aparecer e mostrará também o tempo decorrido enquanto processa.

Quando terminar, atualizará a planilha no Google Sheets configurada.

7. Possíveis problemas e soluções rápidas

Problema: comando streamlit não reconhecido

Use este comando no lugar:

arduino

CopiarEditar

```
python -m streamlit run app.py
```

Problema: erro ao atualizar planilha

Verifique:

- O arquivo credenciaisbiapine.json está na pasta correta?
- O nome da planilha e da aba no código são exatamente iguais aos do Google Sheets?
- A planilha foi compartilhada com o e-mail do client_email dentro do JSON?

Problema: arquivo ausente

Confirme se todos os arquivos citados estão na pasta:

- app.py
 - bia.py
 - pine.py
 - modelo_bia2_pronto_para_preencher.docx
 - credenciaisbiapine.json
 - requirements.txt
-

8. Fluxo resumido para uso diário

1. Abra a pasta `bia_pine_app`.
2. Clique com o botão direito segurando Shift → **Abrir Terminal aqui**.
3. Rode o comando:

`arduino`

CopiarEditar

`python -m streamlit run app.py`

4. Use o navegador que abrir:
 - Aba BIA para gerar dicionários.
 - Aba PINE para atualizar a planilha.

9. Mantenha este guia guardado

Sempre que você for usar ou compartilhar o projeto, consulte este guia para lembrar os passos corretos.

Caso apareça qualquer erro diferente, copie a mensagem que aparece no terminal ou na tela do Streamlit e procure ajuda.