

## 1. Informações do Grupo

Nome do grupo: chatgpt lovers

Código do grupo: HIPER-212

Integrantes:

Artur Valladares Hernandez Giacummo

Augusto Esteves Carrera

Lorenzo Percussi

## 2. Visão Geral do Sistema e Seu Propósito:

O sistema é uma aplicação web, projetada para análise de arquivos CSV contendo informações sobre aplicações. Seu propósito é permitir que os usuários usem arquivos pré-disponibilizados ou carreguem seus próprios e então a aplicação realiza a análise desses dados e gera visualizações das relações entre as aplicações e um texto.

## 3. Visão de Alto Nível e Interações:

O software interage com os usuários através de uma interface web, permitindo o upload de arquivos CSV. Ele também pode interagir com outros sistemas ou componentes externos para processar os dados ou gerar visualizações.

## 4. Estado Atual e Descrição da Arquitetura :

Atualmente, o sistema é composto por:

- **Frontend:** Desenvolvido com Streamlit, oferecendo uma interface de usuário intuitiva para upload de arquivos e interação com os resultados.
- **Backend:** Funcionalidades de limpeza de dados e geração de arquivos são implementadas em Python, utilizando Pandas, Pydotplus e outros módulos Python. Além disso, é usada a linguagem C para realizar a análise dos dados após a limpeza.

## 5. Decisões Arquitetônicas:

- Utilização do Streamlit para desenvolvimento rápido da interface do usuário, além de ser responsivo (se adaptar às dimensões dos dispositivos), acessível na rede, fácil manipulação e aprendizado e conhecimento prévio da biblioteca.
- Uso do Python e suas bibliotecas para limpeza de dados devido à sua popularidade, poder de processamento, conhecimento da biblioteca e simplicidade.
- Escolha de dependências específicas, como Pandas e Pydotplus, com base em suas capacidades de manipulação de dados.
- Escolha de código em C para realizar a análise dos dados, visto que é possível criar estruturas de dados e ter maior controle da memória.

## 6. Dependências:

- Streamlit: A biblioteca Streamlit é utilizada para criar a interface do usuário.

- pandas: Utilizada para a manipulação e visualização de dados tabulares.
- pydotplus: Utilizada para a conversão de arquivos DOT em JSON para visualização do grafo.
- re: Módulo padrão do Python para expressões regulares, utilizado para verificação de padrões de pastas.

## 7. Instruções de Instalação, Configuração e Operação:

- Instalação: Requer Python instalado no sistema. Use o pip para instalar as dependências listadas usando os seguintes comandos:

*pip install streamlit*

*pip install pandas*

*pip install streamlit-aggrid*

*pip install streamlit\_modal*

*pip install pydotplus*

*pip install regex*

As bibliotecas os, csv, json e subprocess são nativas do python, logo não precisam ser instaladas via comando. Certifique-se de ter permissões de administrador (no Windows) ou use `sudo` (no Linux/macOS) ao executar os comandos acima, se necessário.

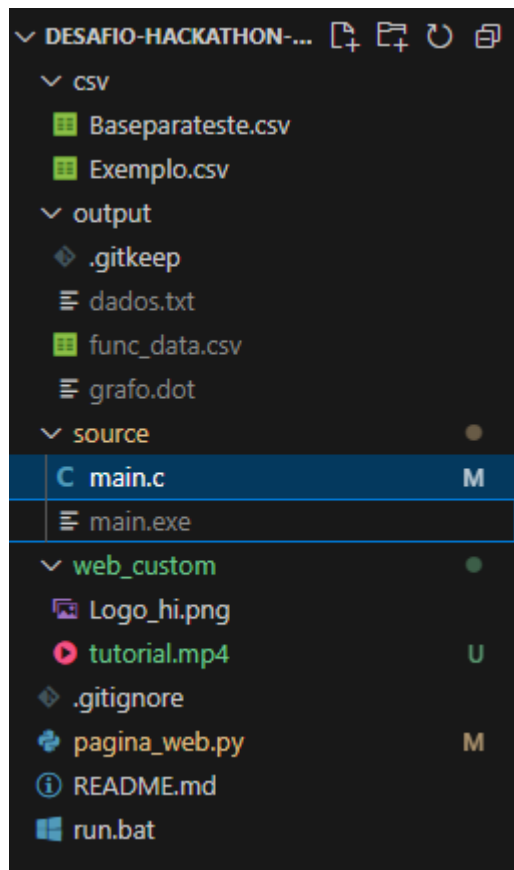
Além disso, o programa usa a linguagem C para fazer a análise de dados, portanto seu sistema operacional deve ser capaz de reconhecer o comando gcc.

- Configuração: Certifique-se de ter permissões adequadas e as bibliotecas instaladas.
- Operação: Execute o arquivo run.bat que está na pasta, após isso ele irá abrir um prompt de comando e então uma página web com endereço localhost:porta.

## 8. Funcionamento

Link do vídeo no youtube mostrando funcionamento: <https://youtu.be/Xy8BK6DnmbU>

Antes de explicar o funcionamento, vale ressaltar que para o arquivo funcionar a hierarquia de arquivos do git deve ser mantida, ou seja, a estrutura dos arquivos deve ser essa:



.git	25/04/2024 13:58	Pasta de arquivos	
csv	25/04/2024 17:56	Pasta de arquivos	
output	25/04/2024 17:55	Pasta de arquivos	
source	25/04/2024 17:55	Pasta de arquivos	
web_custom	25/04/2024 16:29	Pasta de arquivos	
.gitignore	25/04/2024 13:58	Documento de Te...	1 KB
pagina_web.py	25/04/2024 17:41	Arquivo Fonte Pyt...	16 KB
README.md	25/04/2024 13:58	Arquivo Fonte Ma...	1 KB
run.bat	25/04/2024 13:58	Arquivo em Lotes ...	1 KB

O código pode ser executado de duas maneiras:

Ao executar o arquivo “run.bat” abrirá um prompt de comando que se parecerá com o abaixo:

```
C:\Windows\system32\cmd.exe - streamlit run pagina_web.py

You can now view your Streamlit app in your browser.

Local URL: http://localhost:8501
Network URL: http://192.168.15.9:8501
```

Ou pelo prompt de comando, mas para isso você deve ir até a pasta onde esta a aplicação e então digitar `streamlit run paginaweb.py`:

```
Prompt de Comando - streamlit run pagina_web.py
Microsoft Windows [versão 10.0.19045.4291]
(c) Microsoft Corporation. Todos os direitos reservados.

C:\Users\Staation>cd C:\Users\Staation\Desktop\b\t
C:\Users\Staation\Desktop\b\t>cd Desafio-Hackathon-Hiperstream
C:\Users\Staation\Desktop\b\t\Desafio-Hackathon-Hiperstream>streamlit run pagina_web.py

You can now view your Streamlit app in your browser.

Local URL: http://localhost:8501
Network URL: http://192.168.15.9:8501
```

E então será aberto uma página web:



Se não houver nenhum arquivo .csv na pasta csv será necessário fazer o upload de um e a página se parecerá com a seguinte:



Caso o .csv selecionado seja vazio (sem nenhum conteúdo dentro) o programa informara que ele é invalido e se parecerá com isso:

# Hiperstream Hackathon

[Precisa de Ajuda?](#)

## Análise de Arquivos CSV

Faça o upload de um arquivo CSV



Drag and drop file here

Limit 200MB per file

[Browse files](#)

Ou use um dos arquivos CSV abaixo

Selecione um arquivo CSV:

base.csv



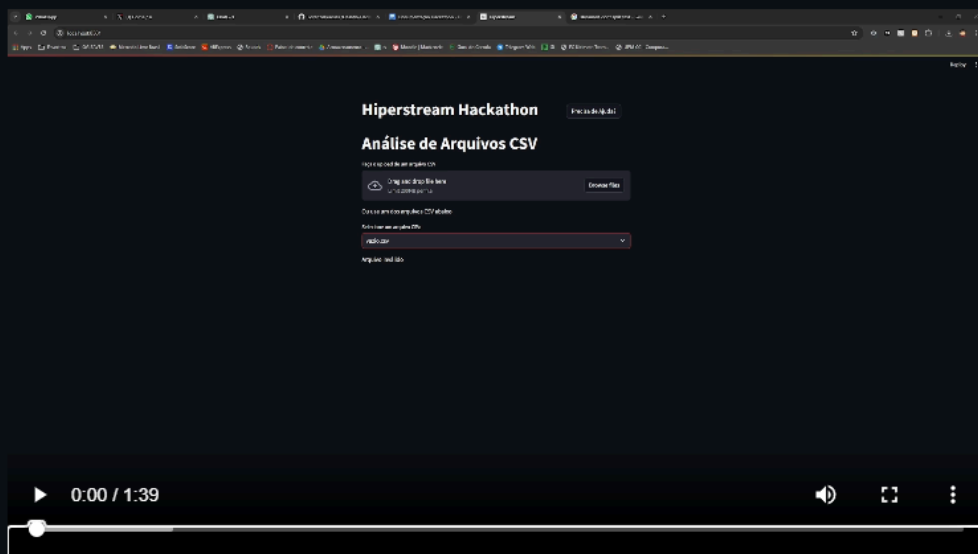
Arquivo inválido

Botão “Precisa de Ajuda?”:

## Tutorial

X

## Video



Se o vídeo não carregar clique [aqui](#)

## Escrito

- 1- Escolha um arquivo da lista ou faça upload do seu próprio.
- 2- Clique em "Analisar dados" para fazer a análise do arquivo escolhido.
- 3- Após a análise será um grafo e um texto resposta.
- 4- O grafo é interativo, ou seja, você pode navegar e editá-lo.
- 5- Se você parar o mouse sobre um vertice, será mostrado o ID.
- 6- Se você parar o mouse sobre uma aresta, será mostrado o diretorio e tipo da relação.

Nesta página Web o usuário pode clicar no botão “Precisa de Ajuda?” para entender o funcionamento do site, escolher arquivos .csv já existentes ou então fazer upload de um personalizado.

Quando um arquivo é selecionado ele mostra a pré-visualização da tabela.

Ao clicar em “Analisar dados” o programa pode chegar em 2 cenários: erro ou mostrar a resposta obtida. O primeiro parecerá com a foto abaixo:

# Hiperstream Hackathon

Precisa de Ajuda?

## Análise de Arquivos CSV

Faça o upload de um arquivo CSV

 Drag and drop file here  
Limit 200MB per file

Browse files

Ou use um dos arquivos CSV abaixo

Selecione um arquivo CSV:

base.csv

Preview da Tabela:

	ID	Nome	PastaOrigem

Analisar Dados

## Resultados

Arquivo Invalido

Vale ressaltar que qualquer coisa que ele identificar como errado será informado para o usuário como no exemplo a seguir:



# Hiperstream Hackathon

Precisa de Ajuda?

## Análise de Arquivos CSV

Faça o upload de um arquivo CSV



Drag and drop file here

Limit 200MB per file

Browse files

Ou use um dos arquivos CSV abaixo

Selecione um arquivo CSV:

base (1).csv

Preview da Tabela:

	ID	Nome	PastaOrigem
0	3,036	Cliente-FATURA-HSJ	E:\PRODUCAO\Cliente\FATURA\HSJ\ENTRADA\
1	3,037	Cliente-FATURA-WFD-EMAIL	E:\PRODUCAO\Cliente\FATURA\FORMATAR_FATU
2	3,036	Cliente-FATURA-HSJ	E:\PRODUCAO\Cliente\FATURA\HSJ\ENTRADA\
3	3,039	Cliente-FATURA-WFD-ARMAZENAMENTO	E:\PRODUCAO\Cliente\FATURA\FORMATAR_FATU
4	3,040	Cliente-FATURA-WFD-ARMAZENAMNTO-ALEATORIO	E:\PRODUCAO\Cliente\FATURA\FORMATAR_FATU
5	3,041	Cliente-FATURA-QUEBRAPDF-EMAIL	E:\PRODUCAO\Cliente\FATURA\SAIDAS\PDF_UN
6	3,042	Cliente-FATURA-QUEBRAPDF-ARMAZENAMENTO	E:\PRODUCAO\Cliente\FATURA\SAIDAS\PDF_UN
7	3,043	Cliente-FATURA-ACTION-EMAIL 1a VIA	E:\PRODUCAO\Cliente\FATURA\ACTION\ENTRAD
8	3,044	Cliente-FATURA-INPUBLOG-DIGITAL	E:\PRODUCAO\Cliente\FATURA\INPUTBLOG\
9	3,045	Cliente-FATURA-CARGAECM-DIGITAL	E:\PRODUCAO\Cliente\FATURA\CARGA_ECM\

Analisar Dados

## Resultados

ID repetido encontrado: 3036

A pasta de origem é igual à pasta de backup. ID: 3046

Arquivo Invalido

No entanto quando é obtido um resultado se parecerá com isso:

# Hiperstream Hackathon

Precisa de Ajuda?

## Análise de Arquivos CSV

Faça o upload de um arquivo CSV



Drag and drop file here  
Limit 200MB per file

Browse files

Ou use um dos arquivos CSV abaixo

Selecione um arquivo CSV:

Exemplo.csv



Preview da Tabela:

	ID	Nome	PastaOrigem
0	1	Aplicacao1	C:\EntradaA\
1	2	Aplicacao2	C:\Entrada2\
2	3	Aplicacao3	C:\Voa
3	4	Aplicacao4	C:\Entrada2\
4	5	Aplicacao5	C:\monitorada\3k\
5	6	Aplicacao6	C:\Aplicacao2Dest\

Analisar Dados

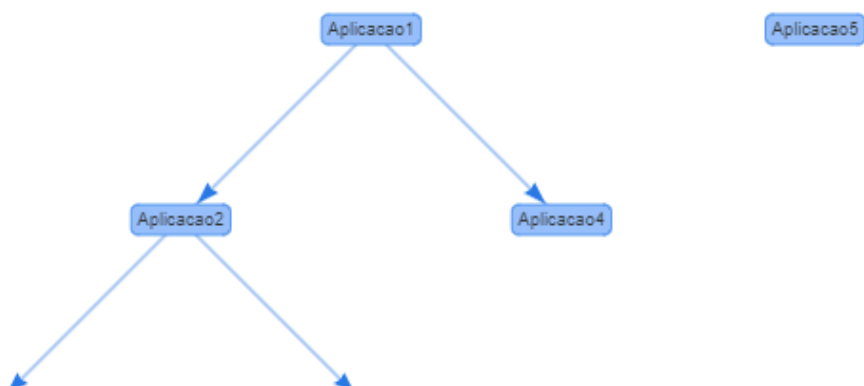
## Resultados

Grafo Dados

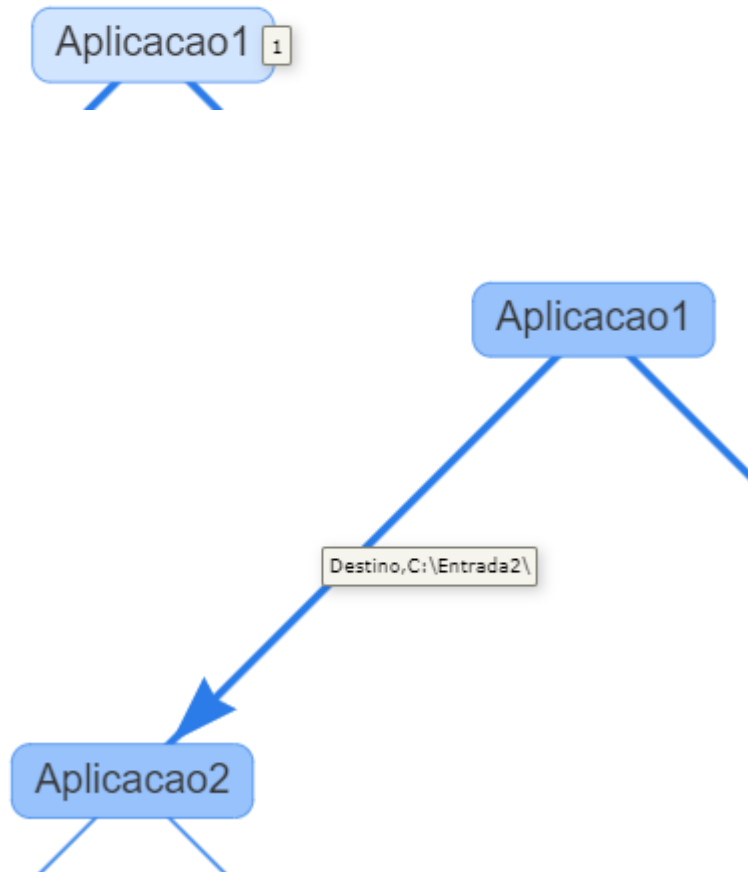
### Grafo gerado



Edit

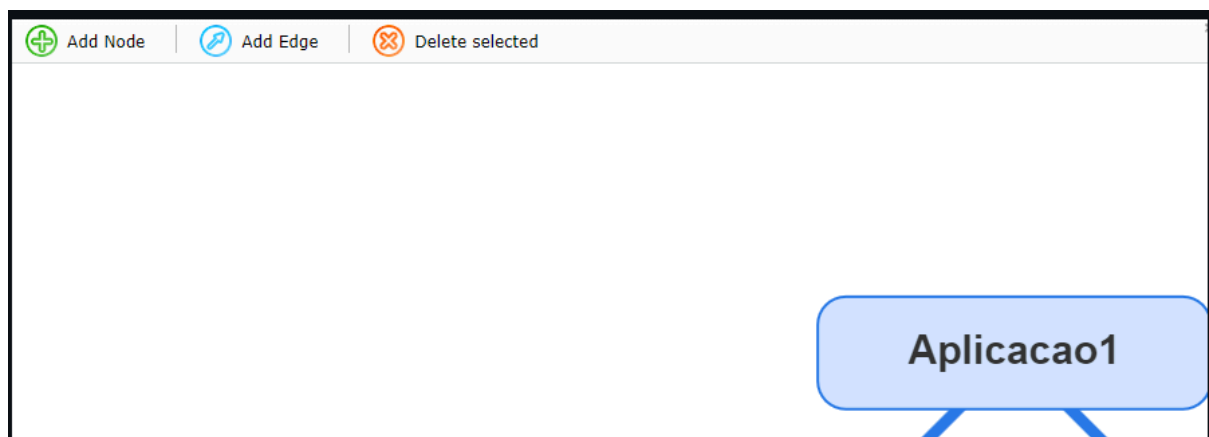


A aba grafo mostra o grafo obtido mapeando as relações entre aplicações. Quando o mouse fica parado sobre um vértice seu id é mostrado. Já quando o mouse fica parado sobre uma aresta a pasta que liga as duas é informada e qual tipo de pasta é (destino ou backup)



Além de tudo o grafo ainda é editável, quando clicar em “Edit”, no entanto as alterações não são carregadas para o csv nem para a análise em texto.

Editar vértice:



Editar aresta:



Por fim, a aba de texto mostra uma transcrição do grafo e se parece com isso:

# Hiperstream Hackathon

[Precisa de Ajuda?](#)

## Análise de Arquivos CSV

Faça o upload de um arquivo CSV



Drag and drop file here

Limit 200MB per file

[Browse files](#)

Ou use um dos arquivos CSV abaixo

Selecione um arquivo CSV:

Exemplo.csv



Preview da Tabela:

	ID	Nome	PastaOrigem
0	1	Aplicacao1	C:\EntradaA\
1	2	Aplicacao2	C:\Entrada2\
2	3	Aplicacao3	C:\Voa
3	4	Aplicacao4	C:\Entrada2\
4	5	Aplicacao5	C:\monitorada\3k\
5	6	Aplicacao6	C:\Aplicacao2Dest\

[Analisar Dados](#)

## Resultados

Grafo **Dados**

### Dados gerados

Aplicação Aplicacao1(1)

Fornece dados pela PASTA DESTINO para as aplicações: Aplicacao2(2),Aplicacao4(4)

Aplicação Aplicacao2(2)

Usa dados de: Aplicacao1(1)

Fornece dados pela PASTA DESTINO para as aplicações: Aplicacao6(6)

Fornece dados pela PASTA BACKUP para as aplicações: Aplicacao3(3)

Aplicação Aplicacao3(3)

Usa dados de: Aplicacao2(2)

Aplicação Aplicacao4(4)

Usa dados de: Aplicacao1(1)

Aplicação Aplicacao5(5) não possui relações

Aplicação Aplicacao6(6)

Usa dados de: Aplicacao2(2)