



Grupinho sem nome



Time

André da Luz Moraes - RM 93153

Eduardo Nunes - RM 92850

Juan Marco Quezada Camacho - RM 93050

Mariana Freitas de Abreu - RM 92846

Miguel Marcondes - RM 93975



DESAFIO

Esse ano fomos desafiados pela IdWall para criarmos uma aplicação web que ajudasse a combater a lavagem de dinheiro e o financiamento de práticas terroristas, gerenciando as duas listas de pessoas procuradas da Interpol e do FBI.

Técnicas WebScraping & WebCrawler

1ª Parte - Coleta e Busca de Dados:

Nesta primeira fase do projeto, desenvolvemos um sistema privado para as duas organizações e suas filiadas. Criamos um sistema de autenticação com um usuário administrador, que permite o acesso restrito.

A principal funcionalidade desenvolvida nesta etapa é a capacidade de buscar dados de criminosos em um banco de dados. A pesquisa é realizada com base em vários critérios, incluindo nome, idade, tipo de crime, local de nascimento e local do crime. Implementamos filtros avançados para refinar a pesquisa e proporcionar uma experiência de usuário aprimorada. Os resultados da pesquisa são exibidos de forma organizada, facilitando a análise.

Além disso, criamos um painel de controle (dashboard) que exibe gráficos estatísticos, fornecendo uma visão geral das estatísticas relacionadas aos criminosos e crimes. Para implementar essa funcionalidade, utilizamos as técnicas de visualização de dados aprendidas durante o curso.

2ª Parte - Coleta de Dados e Desenvolvimento do Back-End:

Na segunda fase do projeto, focamos na coleta de dados a partir de fontes externas, especificamente das duas instituições relevantes (FBI e Interpol). Para realizar essa coleta de forma eficaz, implementamos um sistema de web crawling, que percorre as páginas oficiais dessas instituições para extrair informações sobre criminosos, suas imagens e os crimes cometidos.

Os dados coletados são então processados e alimentados em nosso banco de dados, atualizando-o com as informações mais recentes. Esse processo de atualização é automatizado para garantir que nossa base de dados esteja sempre atualizada em tempo real.

Além disso, nesta fase, trabalhamos no desenvolvimento do back-end da aplicação, onde os dados coletados são armazenados, gerenciados e servidos aos usuários. Isso envolve a criação de APIs (Application Programming Interfaces) para que a interface do usuário possa acessar os dados de maneira eficiente e rápida.

Em resumo, na segunda parte do projeto, combinamos técnicas de web crawling para coleta de dados externos e desenvolvimento do back-end para alimentar nosso sistema com informações atualizadas, tornando-o funcional e dinâmico.

Web Crawler para coleta de dados

1. Importação de Bibliotecas:

- + O código começa importando as bibliotecas necessárias: "requests" para fazer requisições HTTP e "pandas" para processar os dados e exportá-los para um banco de dados.

2. Definição de Páginas a Serem Coletadas:

- | A variável page é definida para conter uma lista de números de página, indicando as páginas do site do FBI que serão acessadas. Neste caso, apenas as 5 primeiras páginas são acessadas.

3. Loop para Coleta de Dados:

O programa itera sobre cada página definida em page. Para cada página, ele constrói uma URL com parâmetros específicos para acessar uma lista de criminosos procurados. O programa faz uma requisição HTTP à URL usando o método requests.get() e converte a resposta JSON em um objeto Python.

4. Extração e Transformação de Dados:

Para cada criminoso na lista de resultados, o programa extrai as seguintes informações pessoais:

Nome (campo `legat_names`).

Data de Nascimento (campo `dates_of_birth_used`).

Cor do Cabelo (campo `hair`).

Cor dos Olhos (campo `eyes`).

Peso (campo `weight`).

Altura (campo `height_max`).

Profissão (campo `occupations`).

Marcas e Cicatrizes (campo `scars_and_marks`).

Número NCIC (campo `ncic`).

Recompensa Máxima (campo `reward_max`).

Aviso (campo `warning_message`).

Local de Nascimento (campo `place_of_birth`).

Nacionalidade (campo `nationality`).

Sexo (campo `sex`).

Raça (campo `race`).

Esses dados são armazenados em um dicionário Python chamado `entrada_criminoso`.

Web Crawler para coleta de dados

5. Armazenamento de Dados em um Dataframe:

Após a conclusão da coleta de dados, o programa cria um DataFrame do Pandas a partir da lista `lista_criminosos`. Isso facilita o processamento e a manipulação dos dados.

6. Exportação para um Banco de Dados PostgreSQL:

Por fim, o DataFrame é exportado para um banco de dados PostgreSQL. O nome da tabela, esquema e detalhes de conexão são fornecidos como parâmetros para a função `to_sql()`.

Protótipo

