

Explicação do código utilizando ferramentas de web scraper em python para um site de concursos públicos (PCI Concursos)

1. Objetivo do Código

O objetivo principal do código é realizar a extração de informações de concursos públicos divulgados no site [PCI Concursos](#), organizar esses dados e armazená-los em arquivos no formato JSON. Para isso, o script realiza a coleta de informações relevantes sobre os concursos, como instituição, estado, quantidade de vagas, salário, cargo, escolaridade, data limite de inscrição e URL da notícia.

2. Funcionamento Geral

O funcionamento do código pode ser dividido em três etapas principais:

- Extração de dados: usando urllib e BeautifulSoup, o script acessa a página de concursos e percorre os elementos HTML para extrair os dados necessários.
- Processamento de dados: os concursos são organizados em dicionários Python para padronizar as informações coletadas.
- Armazenamento: os dados extraídos e processados são salvos em três arquivos JSON distintos: TodosConcursos.json, ConcursosAbertos.json e ConcursosRS.json.

3. Descrição das Funções

3.1 ExtrairTodosConcursos()

Essa função realiza a conexão com o site da PCI Concursos e extrai informações de todos os concursos disponíveis.

Principais etapas:

- Tenta acessar a URL e processar o conteúdo HTML.
- Em caso de erro de rede (HTTPError, URLError) ou erro inesperado, captura a exceção e retorna uma lista vazia.
- Navega pelos elementos <div> da página com classes específicas para identificar os concursos.

- Extrai informações como instituição, estado, quantidade de vagas, salário, cargo, escolaridade, data limite e URL.
- Trata casos em que algum campo esteja ausente, preenchendo-os com None se necessário.
- Armazena os dados extraídos no arquivo TodosConcursos.json.

3.2 MostrarConcursosAbertos(todos_concursos)

Essa função filtra os concursos obtidos, mantendo apenas aqueles cujo prazo final de inscrição ainda não expirou.

Principais etapas:

- Compara a data limite de inscrição de cada concurso com a data atual do sistema.
- Salva o resultado filtrado no arquivo ConcursosAbertos.json.

3.3 MostrarConcursosRioGrandeDoSul(todos_concursos)

Essa função filtra os concursos obtidos, mantendo apenas aqueles que são destinados ao estado do Rio Grande do Sul (RS).

Principais etapas:

- Filtra concursos cuja sigla do estado seja "RS".
- Salva o resultado filtrado no arquivo ConcursosRS.json.

4. Tratamento de Casos Omissos

Estes são os principais casos omissos tratados:

- Falha de conexão com o site: Uso de blocos try/except para capturar e tratar erros como HTTPError, URLError e exceções genéricas.
- Ausência de elementos HTML: Utilização de try/except no loop de extração de concursos para capturar AttributeError e continuar a execução ignorando concursos com estrutura inesperada.

- Campos vazios ou inconsistentes: Adoção de verificações condicionais (if) para preencher os campos com None caso o valor extraído esteja vazio ou seja inexistente.
- Erro ao salvar arquivos JSON: Inclusão de blocos try/except nas funções que gravam arquivos para tratar problemas durante a escrita no disco.
- Datas fora do padrão: Uso de expressões regulares (regex) para capturar apenas datas no formato dd/mm/yyyy, evitando formatos incorretos.