

Python para Dados

AT

Python para Dados	1
AT	1
Introdução	1
Etapas	1
Carregamento e Normalização de Dados	1
Extração de Plataformas	1
Download de Páginas da Wikipedia	1
Parsing das Páginas HTML	2
Extração de URLs e E-mails	2
Associação de Jogos aos Usuários	2
Atualização do Banco de Dados	2
Consulta ao Banco de Dados	2
Dificuldades Encontradas	2
Competências Aplicadas	2
Percepção e Feedback sobre o Curso	2

Introdução

O projeto consistiu em extrair, processar e armazenar informações sobre plataformas de jogos, dados de usuários e jogos extraídos de páginas da Wikipedia. O fluxo envolveu a utilização de técnicas de manipulação de dados com Pandas, scraping com BeautifulSoup, manipulação de arquivos JSON, interação com APIs HTTP e uso de banco de dados com SQLAlchemy.

Etapas

Carregamento e Normalização de Dados

O arquivo INFwebNet_Data.json é convertido em um DataFrame do Pandas, preenchendo valores ausentes com "Não Informado".

Extração de Plataformas

Foi utilizado um loop para percorrer os jogos e extrair as plataformas, armazenando-as em um arquivo de texto.

Download de Páginas da Wikipedia

A biblioteca urllib foi utilizada para acessar URLs dinâmicas do Wikipedia geradas com base nos nomes das plataformas. O HTML é salvo localmente.

Parsing das Páginas HTML

Utilizando o BeautifulSoup, as tabelas com a classe wikitable foram parseadas, e os dados estruturados em dicionários.

Extração de URLs e E-mails

Foram utilizadas expressões regulares (regex) para localizar padrões de URLs e e-mails.

Associação de Jogos aos Usuários

Um loop cruzou os jogos fornecidos pelos usuários com os dados parseados.

Atualização do Banco de Dados

Utilizou-se SQLAlchemy para criar uma tabela Jogos_Plataformas e inserir os registros dos jogos extraídos.

Consulta ao Banco de Dados

Foi criada uma query SQL para buscar usuários na tabela Usuarios filtrando pelo jogo fornecido no input do terminal.

Dificuldades Encontradas

Erros de HTTP e URLs, algumas páginas não existiam, gerando erros que precisaram ser tratados com logs.

Na normalização de Strings as diferenças de acentuação e formatação exigiram a criação de uma função específica.

Associação de Dados ao cruzar as informações de jogos entre usuários e páginas da Wikipedia demandou um tratamento cuidadoso para evitar inconsistências.

Competências Aplicadas

- Manipulação de Dados com Pandas
- Web Scraping (urllib/BeautifulSoup)
- Expressões Regulares
- SQLAlchemy
- Tratamento de Exceções

Percepção e Feedback sobre o Curso

O projeto permitiu consolidar diversas competências em manipulação de dados, scraping e banco de dados. A integração de diferentes bibliotecas e técnicas mostrou-se essencial para resolver problemas reais, com ênfase no tratamento de erros e dados inconsistentes.

O curso proporcionou uma base sólida, especialmente em Python aplicado a dados com o uso do Pandas e em boas práticas de desenvolvimento, como a escrita de código limpo e organizado. Recomendo aprofundamento em web scraping, banco de dados e otimização de algoritmos.