

Projeto de Desenvolvimento: Web Scraping de Notícias do Agronegócio

Objetivo

O objetivo deste projeto é criar um sistema de web scraping com Node.js para capturar notícias relacionadas ao agronegócio de sites específicos. As notícias serão pré-processadas e armazenadas em um banco de dados não relacional, que será acessado por meio de uma API em Node.js. A API será consumida por uma aplicação web em React.

Tecnologias Utilizadas

Node.js

Express.js

MongoDB (Banco de Dados Não Relacional)

React

Fluxo do Sistema

O sistema fará a captura das notícias de sites selecionados usando técnicas de web scraping com Node.js.

As notícias serão pré-processadas e armazenadas em um banco de dados MongoDB.

A API em Node.js será responsável por expor as notícias armazenadas no banco de dados em formato JSON.

A aplicação em React consumirá a API e exibirá as notícias em um formato amigável ao usuário.

Requisitos Funcionais

O sistema deve ser capaz de capturar notícias de sites selecionados e pré-processá-las.

O sistema deve ser capaz de armazenar as notícias pré-processadas em um banco de dados MongoDB.

A API em Node.js deve ser capaz de expor as notícias armazenadas em formato JSON.

A aplicação em React deve ser capaz de consumir a API e exibir as notícias em um formato amigável ao usuário.

Requisitos Não Funcionais

O sistema deve ser capaz de lidar com grande volume de notícias.

O sistema deve ser capaz de lidar com requisições concorrentes.

O sistema deve ser seguro e resistente a ataques.

A aplicação em React deve ser responsiva e ter boa usabilidade.

Arquitetura do Sistema

Web Scraping

O web scraping será implementado usando a biblioteca Cheerio, que permite analisar e manipular documentos HTML usando a sintaxe do jQuery. Serão criados scripts para cada site selecionado, que farão a captura das notícias e o pré-processamento das mesmas.

Banco de Dados

O MongoDB será utilizado como banco de dados não relacional para armazenar as notícias pré-processadas. Será criada uma coleção para armazenar as notícias, com campos como título, conteúdo, data e fonte.

API em Node.js

A API em Node.js será implementada usando a biblioteca Express.js. Serão criadas rotas para retornar as notícias armazenadas no banco de dados. A API também será responsável por lidar com requisições concorrentes e garantir a segurança do sistema.

Aplicação em React

A aplicação em React será responsável por consumir a API em Node.js e exibir as notícias em um formato amigável ao usuário. Serão utilizadas bibliotecas como Axios para consumir a API e Material UI para estilizar a aplicação.

Considerações Finais

Este projeto de desenvolvimento tem como objetivo criar um sistema de web scraping com Node.js para capturar notícias relacionadas ao agronegócio de sites específicos. As notícias serão pré-processadas e armazenadas em um banco de dados não relacional, que será acessado por meio de uma api feita em Node.js e servido a um site de notícias referente ao agronegócio feito em React