

# Documentação do Projeto: Web Crawler de Notícias

---

## Introdução

Este projeto consiste em um Web Crawler desenvolvido em Java para extrair notícias de três diferentes fontes (G1, Terra e UOL) e gerar um relatório com as notícias mais relevantes de cada fonte.

## Estrutura do Projeto

O projeto é composto por várias classes, cada uma com uma responsabilidade específica. Abaixo está a descrição de cada classe e sua função:

### Application.java

Esta classe é o ponto de entrada do projeto e é responsável por coordenar a execução do Web Crawler. Ela define as URLs das fontes de notícias, extrai as notícias de cada fonte, combina todas as notícias em uma lista única e gera um relatório com as notícias mais relevantes.

### G1NewsCrawler.java

Esta classe é responsável por extrair notícias do portal G1. Ela se conecta ao site do G1, extrai os links das notícias principais e cria uma lista de objetos `NewsData` contendo o título, o link e a fonte de cada notícia.

### TerraNewsCrawler.java

Esta classe é responsável por extrair notícias do portal Terra. Assim como a classe `G1NewsCrawler`, ela se conecta ao site do Terra, extrai os links das notícias principais e cria uma lista de objetos `NewsData`.

### UOLNewsCrawler.java

Esta classe é responsável por extrair notícias do portal UOL. Da mesma forma que as classes anteriores, ela se conecta ao site do UOL, extrai os links das notícias principais e cria uma lista de objetos `NewsData`.

## NewsData.java

Esta classe é uma classe de dados simples que representa uma notícia. Ela possui três atributos: `source` (fonte da notícia), `title` (título da notícia) e `link` (link para a notícia).

## Funcionamento do Projeto

1. O método `main()` da classe `Application.java` é executado.
2. Para cada fonte de notícias (G1, Terra e UOL), o Web Crawler correspondente é chamado para extrair as notícias.
3. As notícias de todas as fontes são combinadas em uma lista única.
4. Um relatório contendo as notícias mais relevantes é gerado e salvo em um arquivo de texto chamado `relatorio_noticias.txt`.

## Finalidade do Projeto

O objetivo deste projeto é demonstrar a utilização de Web Crawlers para extrair informações de diferentes fontes da web e realizar uma análise comparativa entre elas. Ele pode ser utilizado como base para o desenvolvimento de sistemas de monitoramento de notícias, análise de tendências e outras aplicações que envolvam a coleta e processamento de dados da web.

---

Esse documento fornece uma visão geral do projeto, suas classes e funcionalidades, além de sua finalidade e possíveis aplicações.