

# TJMG e PJE Scraper

Este projeto consiste em um conjunto de scripts para extrair dados do site do Tribunal de Justiça de Minas Gerais (TJMG) e do sistema de Processo Judicial Eletrônico (PJE), combinando as informações em um arquivo JSON formatado.

## Índice

1. [Requisitos](#)
2. [Instalação](#)
3. [Configuração](#)
4. [Uso](#)
5. [Estrutura do Projeto](#)
6. [Formato dos Dados](#)
7. [Limitações e Considerações](#)

## Requisitos

- Node.js (v14 ou superior)
- npm (gerenciador de pacotes do Node.js)

## Instalação

1. Clone este repositório ou descompacte o arquivo ZIP fornecido:

```
git clone <url-do-repositorio>
```

1. Navegue até o diretório do projeto:

```
cd tjmg-pje-scraper
```

1. Instale as dependências:

```
npm install
```

# Configuração

O arquivo `config.js` contém todas as configurações do scraper. Você pode editar este arquivo para personalizar o comportamento do script:

```
module.exports = {
  // URLs dos sites
  urls: {
    tjmg: 'https://www8.tjmg.jus.br/juridico/pe/consultaPorEntidadeDevedora.jsf',
    pje: 'https://pje-consulta-publica.tjmg.jus.br/'
  },

  // Parâmetros da consulta
  consulta: {
    entidade: 'MUNICÍPIO DE BELO HORIZONTE', // Nome da entidade devedora
    anoInicio: '', // Ano inicial (deixe vazio para não especificar)
    anoFim: '', // Ano final (deixe vazio para não especificar)
    ocultarFechados: false,
    maxPages: 2 // Limita o número de páginas a serem extraídas (para testes)
  },

  // Configurações do navegador
  browser: {
    headless: true, // Defina como false para visualizar o navegador durante a execução
    width: 1366,
    height: 768,
    userAgent: 'Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/115.0.0.0 Safari/537.36',
    timeout: 60000 // 60 segundos
  }
}
```

## Parâmetros Importantes

- **entidade:** Nome da entidade devedora a ser consultada
- **anoInicio** e **anoFim:** Anos de vencimento (deixe vazios para não especificar)
- **maxPages:** Limite de páginas a serem extraídas (útil para testes)
- **headless:** Define se o navegador será visível durante a execução

# Uso

## Execução Completa

Para executar o fluxo completo (TJMG + PJE):

```
npm start
```

ou

```
node index.js
```

## Execução Individual

Para executar apenas o scraping do TJMG:

```
npm run tjmg
```

ou

```
node tjmg-scraper.js
```

Para executar apenas o scraping do PJE (requer um arquivo JSON do TJMG):

```
npm run pje
```

ou

```
node pje-scraper.js <caminho-do-arquivo-json>
```

## Estrutura do Projeto

- **index.js**: Script principal que integra os scrapers do TJMG e do PJE
- **tjmg-scraper.js**: Script para extrair dados do site do TJMG
- **pje-scraper.js**: Script para extrair dados do sistema PJE
- **utils.js**: Funções utilitárias
- **config.js**: Configurações do scraper
- **output/**: Diretório onde os arquivos JSON são salvos

# Formato dos Dados

## Dados do TJMG

Os dados extraídos do TJMG são salvos em um arquivo JSON com o seguinte formato:

```
[
  {
    "entidade": "MUNICÍPIO DE BELO HORIZONTE",
    "codigo": "695",
    "ano": "2006",
    "natureza": "Alimentar",
    "processo": "0093833-64.2025.8.13.0000",
    "data_protocolo": "02/09/2004",
    "credor": "D.T.S.",
    "status": "Aberto-suspenso",
    "numero": "03737209119958130024",
    "pje_processed": false,
    "assertiva_processed": false
  },
  ...
]
```

## Dados do PJE

Os dados extraídos do PJE são combinados com os dados do TJMG:

```
[
  {
    "entidade": "MUNICÍPIO DE BELO HORIZONTE",
    "codigo": "695",
    "ano": "2006",
    "natureza": "Alimentar",
    "processo": "0093833-64.2025.8.13.0000",
    "data_protocolo": "02/09/2004",
    "credor": "D.T.S.",
    "status": "Aberto-suspenso",
    "numero": "03737209119958130024",
    "pje_processed": true,
    "assertiva_processed": false,
    "partyInfo": [
      {
        "name": "DINAMAR TANURI SANTOS (REQUERENTE)",
        "cpf": "",
        "role": "REQUERENTE",
        "lawyers": "SANTUSA MARILIA UTSCH MOREIRA (CPF: 599.660.766-20, OAB: MG64851); ..."
      }
    ]
  }
]
```

```
]
},
...
]
```

## Dados Formatados

Os dados são formatados conforme o exemplo fornecido:

```
[
  {
    "Credor Principal:": "D.T.S.",
    "Número e Natureza do Precatório:": "Precatório nº 695 - Alimentar",
    "Ano de Vencimento:": "2006",
    "Situação:": "Aberto-suspenso",
    "Valor de formação do Precatório (Valor de Face):": "",
    "Data da última atualização do Valor de Face (Data de Liquidação):": "",
    "Protocolo (Data/Hora):": "02/09/2004",
    "Protocolo (Número/Ano):": "",
    "Processo de Execução nº:": "03737209119958130024",
    "Processo SEI nº:": "",
    "Origem:": "",
    "Ação:": "",
    "processed": true,
    "name": "DINAMAR TANURI SANTOS",
    "role": "REQUERENTE",
    "lawyers": "SANTUSA MARILIA UTSCH MOREIRA (CPF: 599.660.766-20, OAB: MG64851); ..."
  },
  ...
]
```

## Limitações e Considerações

- O script foi desenvolvido para extrair dados específicos do TJMG e do PJE. Alterações na estrutura desses sites podem afetar o funcionamento do script.
- O script utiliza o Puppeteer para automatizar a navegação, o que pode ser detectado como um bot por alguns sites. O plugin Stealth é utilizado para minimizar esse risco.
- Alguns campos do formato de saída podem estar vazios se as informações correspondentes não estiverem disponíveis nos sites.
- Para evitar sobrecarregar os servidores, o script inclui delays aleatórios entre as ações.

- O script foi testado com o MUNICÍPIO DE BELO HORIZONTE como entidade devedora. Outras entidades podem requerer ajustes nos seletores.
- 

## Exemplo de Uso Completo

1. Configure o arquivo `config.js` com os parâmetros desejados
2. Execute o script completo:

```
npm start
```

1. Os arquivos JSON serão salvos no diretório `output/` :
2. `tjmg_MUNICÍPIO_DE_BELO_HORIZONTE_.json` : Dados brutos do TJMG
3. `combined_tjmg_MUNICÍPIO_DE_BELO_HORIZONTE_.json` : Dados combinados do TJMG e PJE
4. `formatted_MUNICÍPIO_DE_BELO_HORIZONTE_.json` : Dados formatados conforme o exemplo