#### **Tipos de página**

| Nome | Imagem de exemplo |
| --- | --- |
| Página de edições |  |

#### 

#### **Elementos de interesse**

| Nome | Presente na página | Seletor | Relativo a | Tipo de objeto |
| --- | --- | --- | --- | --- |
| Corpo da tabela | Página de ediões | //div[@class="widget-body"]/div/table/tbody | Raíz | Divisão de página <div> |
| Linhas da tabela | Página de edições | ./tr[i] | Divisão base | Lista de elementos - Row <tr> |
| Edição | Página de edições | ./td/strong/text() | Linha da tabela | Dado - Texto |
| Data de publicação | Página de edições | ./td[1]/text() | Linha da tabela | Dado - Texto |
| Link para baixar PDF | Página de edições | ./td[3]/div/a[1]/@hreh | Linha da tabela | Dado - Link <a> |
| Lista de paginação | Página de edições | //div[@class="row"]/div/nav/ul | Raíz | Lista não ordenada <ul> |
| Link para próxima página | Página de edições | ./li[last()]/a/@href | Último elemento da lista | Link <a> |

#### 

#### **Requisições**

| Nome | De | Para | Tipo de requisição | Exemplo de URL | Exemplo de formdata |
| --- | --- | --- | --- | --- | --- |
| Requisição inicial | N/A | Página de edições | GET | https://diariooficial.portonacional.to.gov.br/edicoes | N/A |
| Acesso a outras páginas | Página legal | Próxima página | POST | https://diariooficial.portonacional.to.gov.br/edicoes?page=j | {"page": "j+1"} |

#### 

#### Estratégia de raspagem

1. Fazer “requisição inicial” para a “Página de edições”
2. Em “Tabela”, buscar “Linha da tabela”
3. Ler e armazenar os textos de “Edição” e “Data de publicação” no “Dado 1”
4. Ler e armazenar o *href* do “Link para baixar PDF” do “Dado 2”
5. Repetir o passo 2 até a última “Linha da tabela” em “Tabela”
6. Em “Lista de páginas”, buscar pelo “Último elemento da lista” e utilizar o *href* do “Link para próxima página” para realizar a paginação

*Alternativa:* Incrementar o valor de “page” em 1 no “acesso a outras páginas” para acessar a próxima página

1. Repetir os passos 2 a 6 até que o texto de “Edição” seja igual a *EDIÇÃO Nº 1*