

Omnilex Code Assignment: Brazilian Civil Code

To goal is to scrape the page : https://www.planalto.gov.br/ccivil_03/Leis/2002/L10406.htm which is the official Brazilian civil code.

The output should be a .json file with an array of object with the following schema:

```
{
  articleCitation : string;
  articleContent : string;
  sections : string[]
}
```

where **articleCitation** is the citation of the article like “art. 41”. , **articleContent** its content like : “São pessoas jurídicas de direito público interno: I - a União; II - os Estados, o Distrito Federal e os Territórios; III - os Municípios; IV - as autarquias; IV - as autarquias, inclusive as associações públicas; (Redação dada pela Lei nº 11.107, de 2005) V - as demais entidades de caráter público criadas por lei.” and **sections** the list of section this article is in : [“PARTE GERAL”, “LIVRO I DAS PESSOAS”, “TÍTULO II DAS PESSOAS JURÍDICAS”, “CAPÍTULO I Disposições Gerais”].

The main challenge here is in keeping the right nested section path for each article, given the large size of the webpage. as well as avoiding revoked articles.

Constraints :
- Typescript with Playwright

Delivery,
- a link to a github repo with both the code and .json file

