

Teste Scraping

Jadiel Eudes Mendonça Barbosa

Execução:

- Baixar o projeto no repositório: https://github.com/Eudess/teste_scraping.
- instalar o framework scrapy na máquina ou numa máquina virtual pelo comando: `pip install scrapy`.
- navegar pelo terminal/cmd até a pasta do projeto.
- Executar o comando: `scrapy crawl trf5`

Decisões:

- Estudar o funcionamento do framework scrapy para aproveitá-la o máximo possível.
- Estudar o funcionamento do framework selenium para tentar inserir no projeto.
- Pegar toda a tag body para a lista “movimentacoes” para tratá-la posteriormente.

Dificuldades:

- Entender como preencher um formulário com scrapy. Para esta dificuldade procurei por exemplos e ler a documentação para conseguir implementar.
- Entender como funciona a função xpath, também foi difícil encontrar bons exemplos. Para esta dificuldade fui analisar os exemplos que tinham para entender um pouco da estrutura.
- Preencher um valor para um radio button no formulário. Para esta dificuldade li a documentação e utilizei o parâmetro `formid` da função `FormRequest.from_response` que utilizei para preencher todo o formulário com os números do processo.
- Realizar a busca por CNPJ, pois não consegui preencher o valor de dois radio button. Para esta dificuldade pesquisei em diversos lugares, li a documentação, procurei em livros específicos de scrapy, mas não encontrei uma resposta satisfatória para esta situação. Outra alternativa que pensei foi integrar o framework selenium ao projeto, porém não consegui fazer uma conexão satisfatória do selenium com o scrapy.