



INSTITUTO FEDERAL DO TRIÂNGULO MINEIRO

CAMPUS

ITUIUTABA

PROGRAMAÇÃO PARALELA E DISTRIBUÍDA

COMPARAR OS TEMPOS DE EXECUÇÃO PARA UMA TAREFA DE WEB SCRAPING

AENDER SOUZA

GUILHERME DANTAS

ITUIUTABA - MG

2023

Introdução

Esse trabalho visa comparar os tempos de execução para uma tarefa de web scraping usando a biblioteca Selenium com Python utilizando o método sequencial, método paralelo com uma instância do browser e método paralelo com mais de uma instância do browser.

Desenvolvendo o trabalho

Como já dito anteriormente utilizamos a linguagem python juntamente com a biblioteca Selenium para realizar o web scraping do site flashscore.com pegando algumas informações dos times e gerando um arquivo CSV para guardar essas informações. Primeiramente utilizamos o código sequencial que teve o tempo **50 minutos e 55 segundos**. Com intenção de diminuir esse tempo utilizamos outro método com as funções ThreadPoolExecutor e conseguimos baixar para o tempo **45 minutos e 20 segundos**. E por fim com intenção de reduzir mais ainda o tempo aplicamos ThreadPoolExecutor com mais de uma instância do browser tendo assim um tempo de **33 minutos e 52 segundos**.

Conclusão

Este estudo evidenciou a significativa influência das abordagens de execução na eficiência do web scraping utilizando Selenium com Python. Inicialmente, o método sequencial demandou 50 minutos e 55 segundos, revelando uma base para comparação. A implementação do método paralelo, através da utilização de `ThreadPoolExecutor`, demonstrou uma notável melhoria, reduzindo o tempo para 45 minutos e 20 segundos. Contudo, o método que mais teve uma redução expressiva foi alcançada `ThreadPoolExecutor` com múltiplas instâncias do navegador, resultando em um notável tempo de 33 minutos e 52 segundos.