|  |  |  |
| --- | --- | --- |
|  | **CURSO:** | **MBA em BI e Big Data** |
| **DISCIPLINA:** | Captura e Visualização de Dados |
| **GRADE:** | Teoria, Prática e Introdução a Captura e Visualização (Scrapy, Javascript) |
| **PROFESSOR:** | Diénert Vieira (dienertalencar@gmail.com) |
| **ALUNO(S):** | Felipe Pedro Maia de Sousa  Frederico Augusto Lima Queiroga |

**Avaliação 1**

1. Implemente um programa que entre no site do UOL e imprima apenas a seguinte mensagem: A cotação atual do dólar é: <cotação>, onde <cotação> vai ser o valor capturado do site no momento. Procure uma forma de omitir as mensagens de log na execução do seu programa para aparecer apenas essa mensagem como saída.

*R= Arquivo: Atividade01\_q01.py*

*Comando Execução: scrapy runspider --nolog Atividade01\_q01.py*

*Para não aparecer o log na execução do programa foi adicionado no comando de execução a tag –nolog.*

1. Implemente um programa que receba um produto como parâmetro e liste o nome e o preço de todos esses produtos no mercado livre, com paginação incluída. Busque uma forma de passar um parâmetro para o seu programa.

*R= Arquivo: Atividade01\_q02.py*

*Comando de Execução: scrapy runspider --nolog Atividade01\_q02.py*

1. Quais cuidados devem ser observados ao capturar dados de um site?

*R= Ao criar um scrapy para captura de dados alguns cuidados devem ser tomados:*

* *O conteúdo não pode ter direitos autorais*
* *O scraping não pode onerar os serviços do site*
* *Não pode violar os termos de uso de site*
* *Não pode coletar informações sensíveis*
* *Verificar arquivo robots.txt disponível no site, esse arquivo contém informações tem ou não permissão para realização do scrapy.*

1. Quais ameaças capturas automáticas proporcionam para sistemas web?

*R= Podem ser vistas como ameaças a possível exposição dos dados incluindo informações sensível de um site. Pode haver sobrecarga dos serviços de um site causando lentidão ou a não entrega dos serviços deixando o site indisponível.*

1. Você diria que *bots* ou *crawlers* são programas facilmente paralelizáveis? Se sim, explique como isso seria implementado dando um exemplo.

*R= Bots ou crawlers são paralelizáveis de forma fácil, onde uma tarefa pode ser dividida em varias sub tarefas que são independentes umas das outras. Cada uma das tarefas por exemplo pode capturar uma determinada informação que exista no site.*

Obs.: As questões com códigos fonte devem apresentar o link da solução no seu repositório no Github