|  |  |  |
| --- | --- | --- |
|  | **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 |
| **ALUNO(S):** | **Jackson Pontes de Mesquita**  **Jonas Felipe Freitas da Silva** |

**Avaliação 1**

1. Implemente um programa que capture pelo menos 10 notícias do site [https://br.yahoo.com](https://br.yahoo.com/) e as imprima na tela.

**Arquivo:** Avaliação01\_Questão01.py

**Comando:** scrapy runspider Avaliação01\_Questão02.py

Devido a formato da página, o resultado exibe inicialmente as categorias das notícias e só depois os títulos das manchetes. Mantive a solução mas a questão foi anulada em sala de aula.

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.

**Arquivo:** Avaliação01\_Questão02.py

**Comando:** scrapy runspider --nolog Avaliação01\_Questão02.py

Para atender ao requisito da não exibição do log, deve executar a chamada do mesmo utilizando o comando com o parâmetro “--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.

**Arquivo:** Avaliação01\_Questão03.py

**Comando:** scrapy runspider Avaliação01\_Questão03.py

A saída lista o nome do produto e em seguida o valor do mesmo.

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

Alguns aspectos legais que devem ser atentados ao capturar, os cuidados são os seguintes:

* Os dados coletados não devem possuir direitos autorias;
* A captura não deve operar em serviços do site;
* Não deve violar termos de uso do site;
* Jamais deve se coletar dados sensíveis.

1. Quais ameaças capturas automáticas proporcionam para sistemas web?

Existem diversas ameaças que são causadas pelo uso de capturas automáticas, os principais são:

* Exposição de dados, dependendo do nível de segurança do serviço, dados sensíveis podem ser expostos prejudicando a imagem da empresa e clientes;
* Negação de serviço (Denial of service ou DoS), o serviço travar devido a agressividade do sistema de captura impedindo que usuário utilizem a página.

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

Sim, a tarefa principal pode ser divindades em subtarefas e essas tarefas podem ser executadas de em paralelo uma das outras já que a ideia é que cada sub tarefa seja independente, um exemplo prático e a captura de informações de uma página web simples, um cada sub tarefa pode ser captura uma informação especifica dessa página (Uma subtarefa seria responsável por pegar o conteúdo textual e outra responsável pelas imagens).

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