

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.

Arquivo: AV1_Q1.py
Comando: scrapy runspider --nolog AV1_Q1.py
Link: https://github.com/joabalves/MBABI-M11-AV01/blob/master/AV1_Q1.py

- 2) 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: AV1_Q2.py
Comando: scrapy runspider AV1_Q2.py
Link: https://github.com/joabalves/MBABI-M11-AV01/blob/master/AV1_Q2.py

- 3) Quais cuidados devem ser observados ao capturar dados de um site?

Dentre os cuidados necessários ao realizar a captura de dados, deve-se considerar que:

1. Os dados que serão coletados não possuam restrição por direitos autorais;
2. A captura não deve operar prejudicando os serviços do site;
3. A captura não viola termos de uso do site;
4. Jamais sejam coletados dados sensíveis.

- 4) Quais ameaças capturas automáticas proporcionam para sistemas web?

As principais ameaças decorrentes de capturas automáticas em sistemas web são: a perda de confidencialidade com exposição de dados, podendo revelar informações sigilosas e prejudicar a imagem da empresa e clientes; a indisponibilidade de serviços (bloqueio de acessos), onde o serviço pode ser interrompido devido a atuação do sistema de captura; captura de dados para realização de fraudes (*phishing*); e atuação de *malwares* responsáveis por roubos ou sequestros de dados por meio de invasões a computadores e bases de dados.

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

Sim, os uso de bots/crawlers podem ser instanciados de forma paralelizável, organizados em subprocessos de uma rotina principal e com objetivos específicos, como por exemplo no utilizado pelos motores de busca (Googlebot da Google, Slurp do Yahoo e o Msnbot do Bing Microsoft), responsáveis por criar uma cópia das páginas web visitadas de forma recursiva, mantendo sua base de dados atualizadas, facilitando a classificação de dados, promovendo buscas mais rápidas e possibilitando a geração de insights.