

Curso Superior de Tecnologia em Análise e Desenvolvimento de Sistemas - Trabalho prático

Disciplina:	Tópicos Especiais I	Valor:	30 pontos	Pontuação:	
Professor(a):	Pedro David Netto Silveira	Data:		Período:	5º
Aluno:					

Automatização de páginas web com Python

O objetivo deste trabalho é desenvolver a competência relacionada a alguns conceitos básicos de programação com web scraping usando a linguagem Python e biblioteca BeautifulSoup. Neste script, os seguintes aspectos devem ser considerados:

1. Deve-se utilizar a biblioteca BeautifulSoup + python.
2. O script deve produzir uma função para “raspar” informações do site OLX.

Requisitos da função:

Parâmetro de entrada: Uma string que contenha um link para uma página de busca da OLX.

Exemplos:

```
url = 'https://www.olx.com.br/autos-e-pecas/carros-vans-e-utilitarios/seda/estado-es/norte-do-espirito-santo?me=40000&pe=60000&rs=65'
```

```
url = 'https://www.olx.com.br/autos-e-pecas/carros-vans-e-utilitarios/estado-es/norte-do-espirito-santo/vila-velha?me=30000&pe=45000&rs=68'
```

Retorno: Um array que contenha os dados raspados.

Exemplos:















```
anuncios = [['VOLKSWAGEN', 'VOLKSWAGEN VOYAGE 1.6 MSI FLEX 16V 4P AUT', '2019', 'Manual', 'Sedã', 'Prata', '49999', 'Vila Velha'], ['FORD', 'FORD KA+ SEDAN 1.0 TIVCT FLEX 4P', '2019', 'Manual', 'Sedã', 'Prata', '55000', 'Vila Velha'], ['VOLKSWAGEN', 'VOLKSWAGEN VOYAGE TRENDLINE 1.6 T.FLEX 8V 4P', '2018', 'Manual', 'Sedã', 'Branco', '53900', 'Vitória'], ['CHEVROLET', 'CHEVROLET PRISMA SED. LT 1.4 8V FLEXPOWER 4P AUT.', '2019', 'Automático', 'Sedã', 'Preto', '49990', 'Serra'], ['FORD', 'FORD KA 1.5 SEDAN SE 12V FLEX 4P MEC.', '2020', 'Manual', 'Sedã', 'Cinza', '56900', 'Vila Velha'], ['HYUNDAI', 'HYUNDAI HB20 COMFORT PLUS 1.0 TB FLEX 12V MEC.', '2017', 'Manual', 'Sedã', 'Branco', '55000', 'Colatina'], ['RENAULT', 'RENAULT LOGAN ZEN FLEX 1.0 12V 4P MEC.', '2020', 'Manual', 'Sedã', 'Branco', '51990', 'Vitória'], ['VOLKSWAGEN', 'VOLKSWAGEN VOYAGE 1.6 MSI FLEX 8V 4P', '2020', 'Manual', 'Sedã', 'Branco', '59500', 'Vila Velha'], ['FIAT', 'FIAT CRONOS DRIVE 1.8 16V FLEX AUT', '2020', 'Automático', 'Sedã', 'Cinza', '60000', 'Serra'], ['CHEVROLET', 'CHEVROLET ONIX SED. PLUS PREM. 1.0 12V TB FLEX AUT', '2020', 'Automático', 'Sedã', '4 portas', '600', 'Cariacica'], ['FORD', 'FORD KA+ SEDAN 1.0 SEL TICVT FLEX 4P', '2018', 'Manual', 'Sedã', 'Preto', '50000', 'Jaguaré']]
```

```
anuncios = [['FORD', 'FORD KA 1.5 SE PLUS 12V FLEX 5P MEC.', '2020', 'Manual', 'Hatch', 'Branco', '59900', 'Serra'], ['HYUNDAI', 'HYUNDAI HB20 SENSE 1.0 FLEX 12V MEC.', '2018', 'Manual', 'Hatch', 'Branco', '55850', 'Serra'], ['VOLKSWAGEN', 'VOLKSWAGEN FOX CONNECT 1.6 FLEX 8V 5P', '2019', 'Manual', 'Hatch', 'Branco', '58000', 'Serra'], ['CHEVROLET', 'CHEVROLET ONIX HATCH JOY 1.0 8V FLEX 5P MEC.', '2020', 'Manual', 'Hatch', 'Branco', '56600', 'Vila Velha'], ['VOLKSWAGEN', 'VOLKSWAGEN FOX CONNECT 1.6 FLEX 8V 5P', '2018', 'Manual', 'Hatch', 'Branco', '58000', 'Serra'], ['HYUNDAI', 'HYUNDAI HB20 COMF./C.PLUS/C.STYLE 1.0 FLEX 12V', '2018', 'Manual', 'Hatch', 'Preto', '56900', 'Serra'], ['CHEVROLET', 'CHEVROLET ONIX HATCH LT 1.0 8V FLEXPOWER 5P MEC.', '2019', 'Manual', 'Hatch', 'Prata', '57500', 'Cariacica'], ['VOLKSWAGEN', 'VOLKSWAGEN FOX CONNECT I MOTION 1.6 FLEX 8V 5P', '2019', 'Semi-Automático', 'Hatch', 'Preto', '56900', 'Linhares'], ['FIAT', 'FIAT ARGO DRIVE 1.0 6V FLEX', '2020', 'Manual', 'Hatch', 'Prata', '59900', 'Serra'], ['FORD', 'FORD KA 1.5 SE PLUS 12V FLEX 5P MEC.', '2020', 'Manual', 'Hatch', 'Branco', '57900', 'Vila Velha'], ['HYUNDAI', 'HYUNDAI HB20 UNIQUE 1.0 FLEX 12V MEC.', '2019', 'Branco', 'Hatch', '4 portas', '57000', 'Vila Velha']]
```

Requisitos:

1. O script deve buscar os dados dos veículos na URL fornecidas a saber: (i) marca, (ii) modelo, (iii) ano, (iv) cambioAutomatico, (v) se é sedan ou hatch, (vi) cor, (vii) valor e (viii) municipio.
2. A maioria dessas informações encontram-se no link de cada anuncio, na seção “Características” como apresentado no exemplo da imagem abaixo:

Características

 Categoria Carros, vans e utilitários	 Modelo VOLKSWAGEN VIRTUS COMFORT. 200 TSI 1.0 FLEX 12V AUT	 Marca VOLKSWAGEN	 Tipo de veículo Sedã
 Ano 2022	 Quilometragem 8200	 Potência do motor 1.0	 Combustível Flex
 Possui Kit GNV Não	 Câmbio Automático	 Cor Cinza	 Portas 4 portas
 Final de placa 5	 Tipo de direção Hidráulica		
Opcionais			
✓ Alarme	✓ Air bag	✓ Ar condicionado	✓ Câmera de ré
✓ Sensor de ré	✓ Som	✓ Trava elétrica	✓ Vidro elétrico

3. Essa “raspagem” pode ser feita com a biblioteca *BeautifulSoup* com a ajuda de *Selenium*.

Observações e entrega

- O trabalho deve ser realizado em dupla.
- Apenas um arquivo .py deve ser submetido no AVA até a data definida.
- Após a entrega, poderá ocorrer uma avaliação individual (na forma de apresentação oral) sobre a solução adotada.