

Introdução

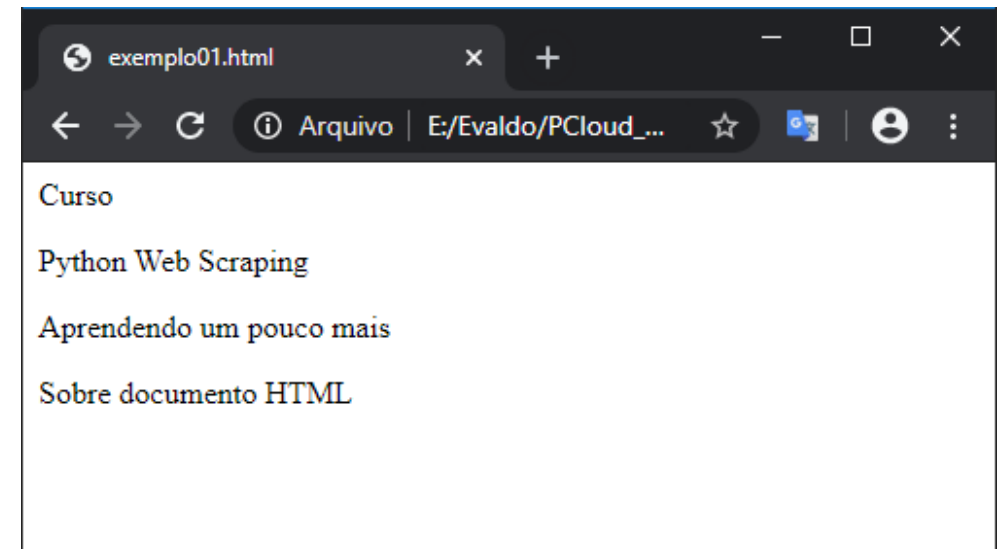
Um pouco mais sobre páginas web

Introdução

Um pouco mais sobre páginas web

Nesta aula você aprenderá um pouco mais sobre arquivos HTML. Veja um arquivo HTML à esquerda e sua exibição no navegador, à direita.

```
1 <html>
2   <body>
3     <div>
4       <p>Curso</p>
5       <p>Python Web Scraping</p>
6     </div>
7     <p>Aprendendo um pouco mais</p>
8     <p>Sobre documento HTML</p>
9   </body>
10 </html>
```



Introdução

Um pouco mais sobre páginas web

Observe as tags que utilizamos:

Tag que indica que é um documento HTML

`<html>...</html>`

Tag que indica o corpo do documento

`<body>...</body>`

```
1 <html>
2   <body>
3     <div>
4       <p>Curso</p>
5       <p>Python Web Scraping</p>
6     </div>
7     <p>Aprendendo um pouco mais</p>
8     <p>Sobre documento HTML</p>
9   </body>
10 </html>
```

Esta tag é um elemento que define uma divisão ou seção em um documento HTML.

Ela é utilizado para definição de blocos, ou seja, ela contém outros elementos.

`<div>...</div>`

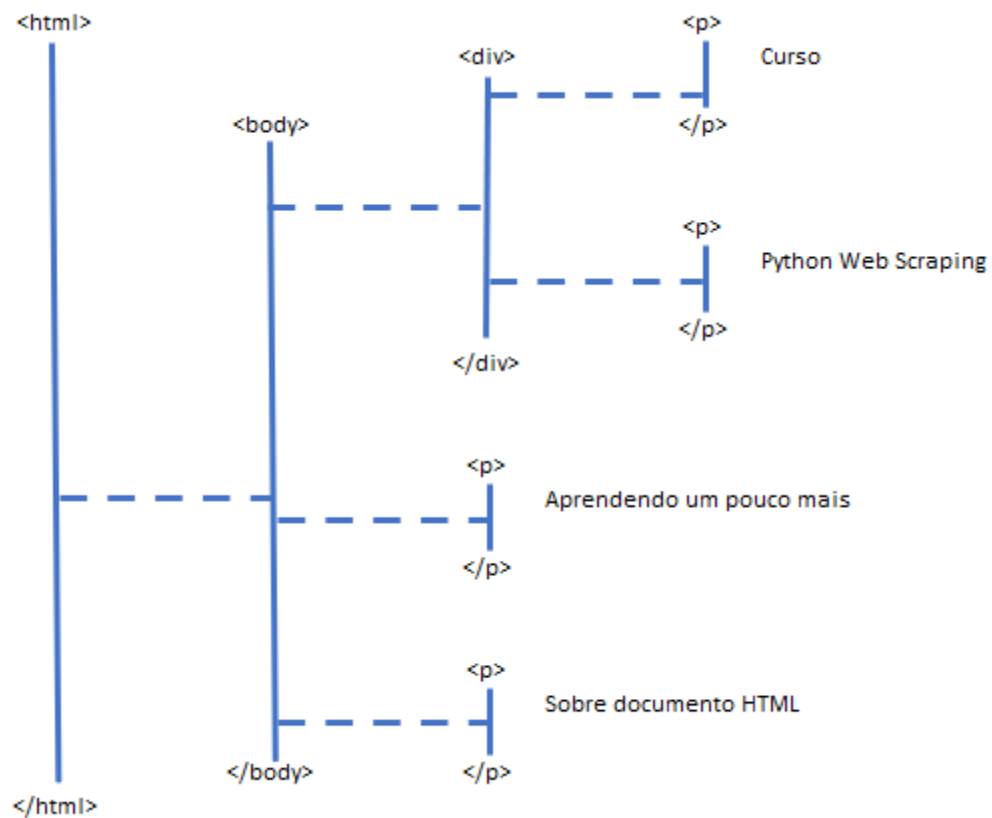
Esta tag é um elemento parágrafo.

`<p>...</p>`

Introdução

Um pouco mais sobre páginas web

Árvore HTML.



```
1 <html>
2   <body>
3     <div>
4       <p>Curso</p>
5       <p>Python Web Scraping</p>
6     </div>
7     <p>Aprendendo um pouco mais</p>
8     <p>Sobre documento HTML</p>
9   </body>
10 </html>
```

Introdução

Um pouco mais sobre páginas web

Algumas tags HTML podem ter atributos.

Os atributos possuem nome e algumas informações, por exemplo:

```
<nome-tag nome-atributo="atrib info">
```

```
...conteúdo...
```

```
</nome-tag>
```

Introdução

Um pouco mais sobre páginas web

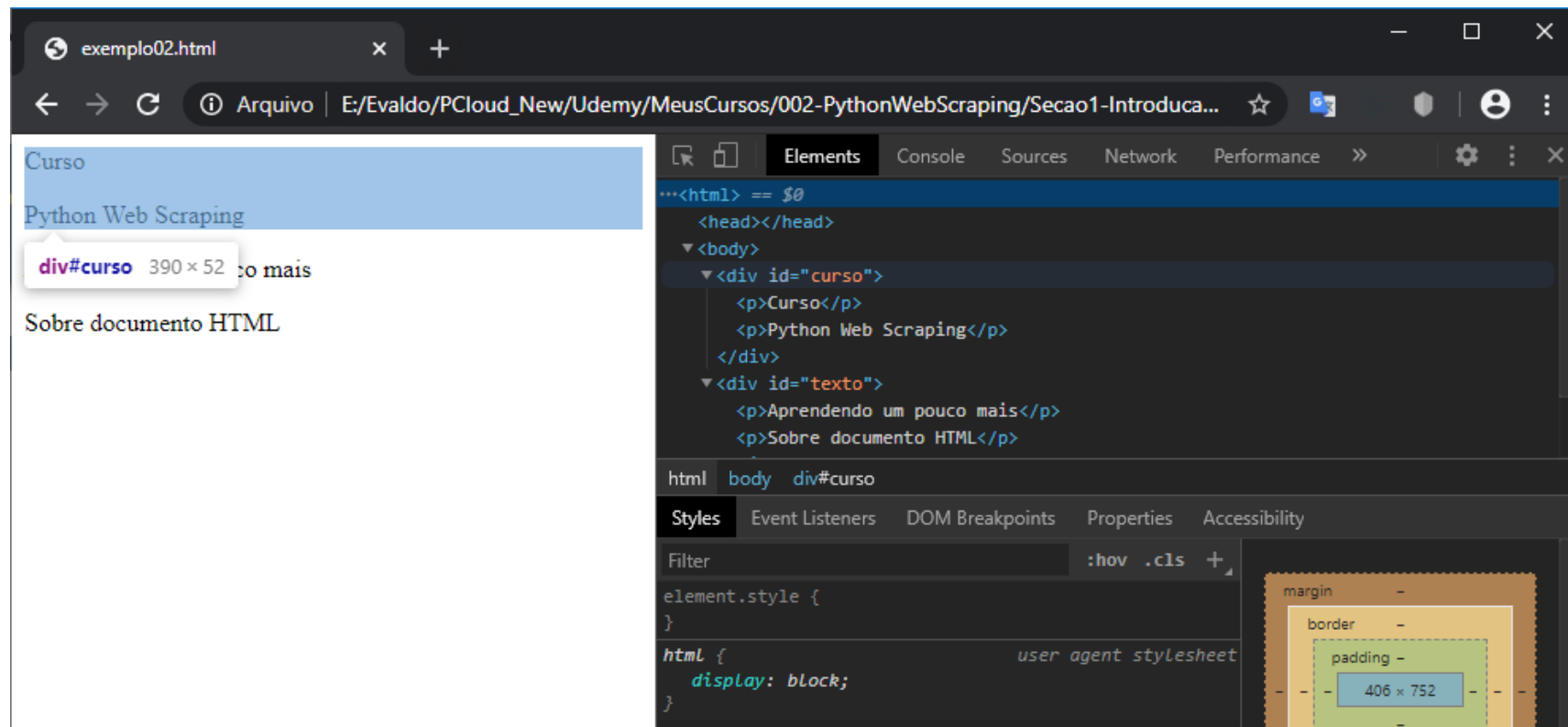
Vamos criar um novo arquivo HTML, adicionando uma nova div e colocando o atributo id, que é um atributo de identificação da tag no documento.

```
1 <html>
2   <body>
3     <div id="curso">
4       <p>Curso</p>
5       <p>Python Web Scraping</p>
6     </div>
7     <div id="texto">
8       <p>Aprendendo um pouco mais</p>
9       <p>Sobre documento HTML</p>
10    </div>
11  </body>
12 </html>
```

Introdução

Um pouco mais sobre páginas web

Inspecionando o documento.



Introdução

Um pouco mais sobre páginas web

A tag que indica um hyperlink para uma página é a tag “a”. Ela possui o atributo href que indica o endereço que será acessado quando o link é clicado. O texto informado é o link exibido na página HTML.

```
<a href='https://www.pythonparatodos.com.br'>
```

Acesse o site Python para Todos

```
</a>
```


Introdução

Um pouco mais sobre páginas web

Analisar e entender as tags HTML do site que queremos realizar o Scraping ajuda e muito na hora de criarmos o nosso algoritmo.

FIM