

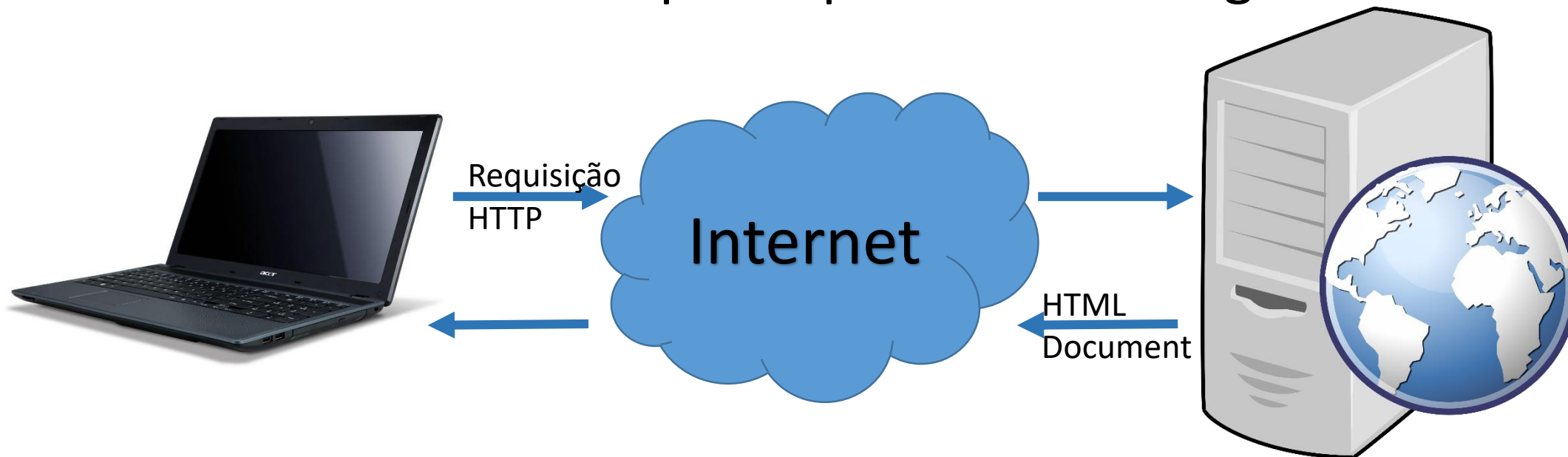
# Introdução

Páginas web

# Introdução – Páginas web

Quando navegamos na internet usando um browser, o mesmo faz requisições de páginas web em servidores.

E os servidores enviam arquivos para o seu navegador.



# Introdução – Páginas web

Veja alguns tipos de arquivos:

HTML: Contém o conteúdo principal da página.

CSS: Define os estilos de exibição (apresentação, aparência da página).

JS: Arquivos JavaScript são responsáveis por tornar as páginas interativas.

JPG e PNG: São imagens exibidas nas páginas.

# Introdução – Páginas web

Assim que o navegador recebe todos arquivos do servidor ele renderiza, ou seja, prepara os arquivos para exibição.

Vários processos são executados em background, porém, quando realizamos um web scraping não precisamos nos preocupar com a maioria deles.

Quando realizamos um web scraping analisamos o HTML retornado.

# Introdução – Páginas web

HTML (HyperText Markup Language – Linguagem de Marcação de HiperTexto) é uma linguagem de marcação que informa ao navegador como deve exibir o conteúdo.

HTML usa “Markup” (marcação) para envolver texto, imagens, e outros conteúdos para renderizar no navegador. Marcadores HTML incluem “elementos” especiais como `<head>`, `<title>`, `<body>`, `<header>`, `<div>`, `<img>`, e muitos outros.

# Introdução – Páginas web

```
<html>
  <head>
    <title>Título da página</title>
  </head>
  <body>
    <h1>Isto é um cabeçalho.</h1>
    <p>Isto é um parágrafo</p>
    <p>Isto é um segundo parágrafo</p>
  </body>
</html>
```

# Introdução – Páginas web

Para conhecer um pouco mais sobre HTML:

<http://tableless.github.io/iniciantes/manual/html/>

# FIM