

## Descrição das fontes de dados - Tarefa 5

Os sites que serão utilizados na realização do crawling são: **Americanas**, **Shoptime** e **Magazine Luiza**. Esses sites foram escolhidos por terem uma interface parecida e elementos HTML que são fáceis de serem selecionados. Mais especificamente, a seção de **smartphones “iPhone”** foi selecionada em cada um desses sites. Abaixo encontra-se o link dessas seções:

### **Americanas:**

<https://www.americanas.com.br/categoria/celulares-e-smartphones/smartphone/iphone/f/sistema-operacional-iphone%20ios/g/condicao-novo?limit=24&offset=0>

### **Shoptime:**

<https://www.shoptime.com.br/categoria/celulares-e-smartphones/smartphone/iphone/g/tipo-de-produto-lphone/pagina-1?ordenacao=relevance&origem=blanca>

### **Magazine Luiza:**

<https://www.magazineluiza.com.br/iphone/celulares-e-smartphones/s/te/teip?page=1>

Analisando a página de alguns produtos, selecionamos apenas os dados que julgamos mais importantes para o desenvolvimento do projeto em questão. A partir desses dados, criamos a modelagem de dados: as entidades, os atributos de cada entidade e o relacionamento entre entidades (consulte o DER para mais informações). Diferentes conjuntos de atributos foram selecionados de cada website, porém a maioria dos atributos são comuns as três fontes. Por exemplo, o atributo *modelo-cod* (codigo do modelo) só esta presente nos sites Americanas e Shoptime. Já o atributo *voltagem* só é encontrado no site Magazine Luiza.

As entidades identificadas nas fontes de dados são: iPhone e Avaliação. Havia uma potencial entidade que não pode ser considerada: a Duvida. Desconsideramos essa entidade pois seus dados são carregados dinamicamente com javascript, assim, não é possível acessa-los com a ferramenta de crawling utilizada (scrapy).