Objetivo

Monitorar um o site GSM Arena, para verificar quando novos dispositivos foram adicionados.

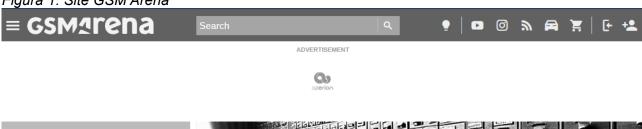
O site

GSM Arena é um site onde contém informações sobre vários tipos de dispositivos móveis de diferentes marcas e modelos.

Tecnologias utilizadas

Scrapy, pandas, Mongo DB, JSON, Requests





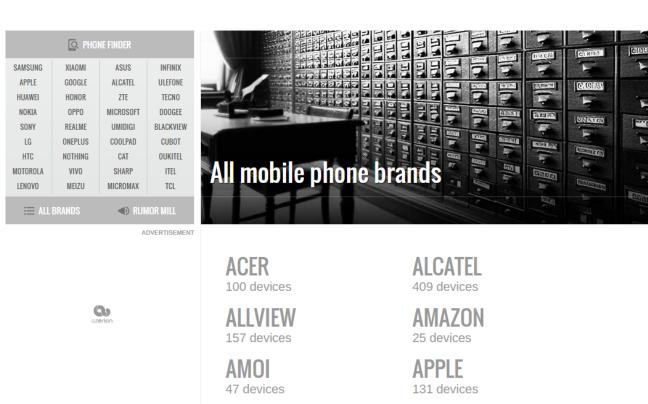


Figura 2: Exemplo de um dispositivo cadastrado no site



Funcionamento

Inicialmente, o bot extrai todas as URLs de todos os produtos presentes no site. Em seguida, uma verificação é feita no banco de dados Mongo DB, para assegurar que aquele produto já não esteja cadastrado lá. Caso o produto ainda não tenha sido registrado, uma requisição será enviada para aquela URL, com objetivo de extrair a tabela de informações desse produto, que será convertida em um arquivo JSON e salvo no banco de dados.