**Exe 04: Busque mecanismos de crawler na web. Cite 2 deles.**

Exemplos de Web Crawlers:

* Google Bot
* Amazon Bot
* Bing Bot
* Slurp Bot (Yahoo)
* DuckDuck Bot
* Baidu Spider
* Yandex Bot
* Sogou Spider
* Exabot
* Facebook External Hit

**Exe 06: Descreva uma estratégia (pense individualmente) para coletar dados sobre a COVID-19 no Brasil para buscar denúncias sobre aglomerações ou abusos ocasionadas por serviços públicos (ônibus, órgãos públicos com redução ou local para idosos sentarem, bancos gerando aglomerações em filas, falta de suporte para gestantes e puérperas/lactantes, etc.). Obs: Escolha uma situação dentro do cenário da pandemia descreva o que quer saber (qual informação que você quer saber). Seu cenário deve proporcionar esse entendimento. Em seguida, aponte uma mídia social ou página web (pelo menos) onde você vai coletar dados. Descreva sua estratégia: possui API (sim ou não)? vai usar a API caso exista (sim ou não)? vai construir um scrape? vai buscar por termos (palavras, frases ou hashtags) ou por publicações em algum local? vai restringir por idioma, qual o período? O objetivo é “afiar” um pensamento crítico sobre a coleta de dados pensando em um cenário de análise.**

Para coletar dados relativos a qualquer coisa desse tipo eu criaria uma ferramenta que se conectasse com a API do Twitter ou do Facebook, pois essas duas são as redes sociais mais utilizadas pela população para tal tipo de postagem. Ambas as APIs possuem rotas para recuperação de dados de posts por hashtag ou título de mensagem. Bastaria criar fazer um scrape utilizando a/as palavra/as chaves do assunto da postagem. A restrição de idioma é necessária pois os termos buscados (ou parecidos, dependendo de quão precisa a ferramenta vai ser na busca por palavras) podem existir em outras línguas com significados completamente diferentes e isso pode comprometer o scrape. Também seria necessário limitar o período de busca para começar aproximadamente em fevereiro de 2020, pois em datas diferentes os termos buscados não estarão relacionados ao COVID-19 – de novo, comprometendo a estatística dos dados do scrape.