



Na Mídia

26/01/2022 | [O Estado de S. Paulo](#)

Web scraping: legal ou ilegal?

Cibele Fonseca | Eduardo Magrani | Tatiana Campello



Eduardo Magrani. FOTO: DIVULGAÇÃO

É conhecida a existência da variedade e do grande volume dos dados disponíveis na internet bem como a dificuldade quanto à manualmente manipulá-los. Devido a esse fato, existe a necessidade quanto a recorrer às tecnologias para fins de automatizar essa manipulação. O web scraping é uma dessas técnicas que torna possível a extração e a manipulação dos dados para fins de diferentes análises.

São exemplos dessas diferentes análises: comparação dos preços de produtos e serviços de diversos segmentos, entendimento do perfil de clientes para apresentar-lhes produtos e serviços, entre diversos outros usos.

O web scraping é composto pelas fases de: website analysis (para entender como os dados são armazenados no site da internet), website crawling (desenvolvimento e execução de scripts também chamados de robôs, bot ou crawlers)

que analisam os sites e coletam os dados – são normalmente desenvolvidos nas linguagens R ou Python), e data organization (limpeza, pré-processamento e organização dos dados). Importante mencionar que a depender da situação, nem todas essas fases são automatizadas e podem requerer envolvimento humano.

Destacam-se alguns benefícios da utilização do web scraping, tais como; os dados podem ser coletados nas fontes que os originam tornando possível avaliar sua veracidade, menos tempo e esforço para a coleta de dados em sites da internet e a coleta de dados para serem utilizados na geração de novos negócios. Referente às desvantagens, deve-se reconhecer a complexidade da coleta dos dados dos sites da internet, como os casos em que os dados nos sites não estejam estruturados e devido a isso existirá um árduo trabalho para fins de coletá-los.

Pelo fato do web scraping ter papel fundamental para a pesquisa, coleta e tratamento de informações é possível vislumbrar que sua utilização tende a crescer uma vez que a maioria dos dados têm sido postados em sites da internet. Cidadãos comuns, tem usado esses sites para promoverem suas vidas pessoais, gostos, tendências, entre outros, e as empresas têm seguido a mesma linha.

Devido aos usos acima citados, destaca-se o fato das empresas de comércio eletrônico utilizarem as informações que as envolvem para fins de prospectar produtos e serviços de acordo com cada perfil.

As práticas de web crawling e web scraping podem ser construídas com base em diferentes estratégias de coleta de dados e tecnologias. Dependendo da situação, os dados e a forma como estão sendo coletados podem estar em uma zona cinzenta legal, onde pode ser difícil atestar se a informação coletada se enquadra no conceito de “publicamente disponível”. Portanto, apesar de ser prática corrente em diversas empresas, o web scraping deve ser executado com cautela para não violar leis existentes, principalmente por envolver em diversos casos a coleta de informações de terceiros, inclusive de concorrentes.

Ainda não há uma regulamentação específica para serviços e práticas de web scraping, nem são essas práticas expressamente consideradas ilegais, no entanto, deve-se evitar a coleta indiscriminada e o uso de materiais protegidos e de dados pessoais, sem que haja a permissão adequada com base nos Termos de Uso da plataforma ou em um contrato específico.

Além disso, a coleta indiscriminada de materiais protegidos e de dados pessoais pode violar, por exemplo, direitos dos titulares de dados pessoais e o princípio da necessidade da Lei Geral de Proteção de Dados Pessoais (LGPD). É importante frisar que o fato de as informações estarem publicamente disponíveis e acessíveis não elimina a necessidade de cumprimento geral da LGPD.

Em adição, além de questões relacionadas à privacidade e proteção de dados, a prática pode esbarrar em limitações de propriedade intelectual. Mesmo quando os dados coletados são considerados públicos, ou quando não claramente entendidos como tal, ainda é necessário cautela: (i) de acordo com o art. 29, I da Lei de Direitos Autorais (n. 9.610 / 98), caso os dados coletados refiram-se a uma obra protegida por direitos autorais, destinada a ser utilizada mantendo parcial ou totalmente sua integridade, será necessária autorização prévia e específica para cópia total ou parcial do material, observadas as exceções legais; e (ii) de acordo com o artigo 195, XI e XII da Lei de Propriedade Industrial (n. 9279/96 – “LPI”), se a informação recolhida não for considerada pública ou foi recolhida envolvendo meios ilícitos, podem ser interpretados como uma modalidade de crime baseada na concorrência desleal.

Portanto, apesar de não ser uma prática expressamente ilegal, deve ser conduzida lícitamente e de forma responsável de maneira a evitar violações a leis e a direitos como, por exemplo, os mencionados relacionados à privacidade e à proteção de dados ou em relação à propriedade intelectual.

***Cibele Fonseca, diretora de Tecnologia da Informação no Demarest Advogados**

***Eduardo Magrani, sócio das áreas de Privacidade, Tecnologia e Cibersegurança e de Propriedade Intelectual e Inovação no Demarest Advogados**

***Tatiana Campello, sócia das áreas de Privacidade, Tecnologia e Cibersegurança e de Propriedade Intelectual e Inovação no Demarest Advogados**