

Igor Rocha: Web crawlers, web scrapers e o Direito

18/09/2022

Ao longo dos últimos anos, o estudo das intersecções entre o Direito e as novas tecnologias se transformou em um dos temas mais presentes dentro da academia e da prática jurídica, tendo, inclusive, se tornado comum presenciar operadores do Direito se utilizando de diversos instrumentos tecnológicos em suas rotinas diárias.



Ocorre que, por vezes, os juristas ainda possuem uma

visão um tanto fantasiosa dos instrumentos tecnológicos que utilizam, entendendo-os como uma espécie de conhecimento longínquo ou abstrato. Tal fenômeno pode ser sumarizado pela constante utilização do termo "robzinho" para se referir a uma vasta gama de softwares com características e propósitos bastante distintos entre si.

Diante dessa situação, e com o intuito de oferecer uma pequena contribuição para as discussões sobre o tema, o presente artigo busca apresentar as principais características de duas espécies de softwares comumente classificados como "robzinhos", quais sejam: os *web crawlers* e os *web scrapers*.

Nesse sentido, *web crawlers* são definidos como ferramentas que, sistematicamente, escaneiam e analisam sites hospedados na internet com o intuito de os indexar em uma base de dados, sendo comum à sua utilização pelos principais sistemas de buscas existentes no mundo, como o Google ou o Bing [1].

Os *web scrapers*, por sua vez, são definidos como ferramentas utilizadas para extrair dados de um site que esteja disponível online [2]. Ou seja, são programas que possuem por intuito identificar a existência de determinada página web, localizar dentro dessa página algum conteúdo previamente estipulado pelos programadores e, então, armazenar essas informações específicas em um banco de dados.

Como exemplos de sites que se utilizam de *web scrapers* podemos citar portais que já são amplamente conhecidos por grande parte dos operadores do direito como o "escavador.com" (que faz a raspagem de dados em sites como o Lattes,) e o "jusbrasil.com" (que faz raspagem de dados de diversas fontes, como os Diários de Justiça).

Embora para muitos não seja algo tão claro, tanto os *web crawlers* quanto os *web scrapers* possuem uma importância muito grande para a o direito contemporâneo. Isso ocorre pois, em um cenário no qual a prática jurídica vem, cada vez mais, sendo digitalizada por meio do processo eletrônico, a capacidade de realizar a coleta e organização inteligente de dados se tornou uma habilidade de grande valia para todos os agentes envolvidos no sistema de justiça.

Nesse sentido, é possível apontar, por exemplo, que um advogado poderia melhorar a qualidade das suas peças caso possuísse a habilidade de, em poucos minutos, coletar e organizar todos os julgados proferidos por um tribunal em determinado tema que lhe seja de interesse.

Também não é difícil imaginar que um ente público poderia se beneficiar da capacidade de coletar e organizar dados espalhados em diversos bancos de dados públicos para poder balizar as suas escolhas orçamentárias ou mesmo as suas políticas públicas.

Em ambos os exemplos, tanto os *web crawlers* quanto os *web scrapers* poderiam ser as ferramentas utilizadas para resolver os problemas apresentados.

Ressalta-se que a intenção aqui não é elencar todas as possíveis utilidades que tais ferramentas podem possuir, mas sim, demonstrar como o conhecimento acerca delas, bem como sobre outras temáticas envolvendo a área da análise de dados, pode acabar se tornando um ativo essencial dentro do direito.

Não obstante todas as vantagens apresentadas, é necessário ter em mente que, como já apontado pela doutrina [3], a utilização de tais programas ainda é envolta em diversas polêmicas.

Nesse sentido, com o intuito de inteirar o leitor na temática, apresenta-se abaixo, de forma não exaustiva, 5 dessas discussões que permeiam a utilização de *web crawlers* e *web scrapers* para a coleta e armazenamento de dados jurídicos.

- a) É possível efetuar a raspagem automatizada de dados em ambientes públicos virtuais, em especial tribunais, quando existirem arquivos "robot.txt" [4] em seus códigos fontes?
- b) Quais são os limites da utilização de *web scrapers* para a raspagem de dados públicos em face das disposições da LGPD/GDPR?
- c) Quais são os limites éticos da utilização de *web crawlers* e *web scrapers* na realização de pesquisas acadêmicas e de pesquisas com intuito comercial?
- d) Quais são as consequências de se utilizar dados obtidos por meio de bancos públicos que possuam informações erradas ou inconclusivas?

e) É possível que a raspagem automatizada de dados crie ou reforce a existência de vieses cognitivos e preconceitos presentes na nossa sociedade?

Ressalta-se que a intenção aqui não é de exaurir a temática, mas sim fazer uma breve introdução e rápidos apontamentos para que, no futuro, mais operadores do direito possam se utilizar de tais ferramentas para otimizarem suas rotinas de trabalho.

Desta feita, espera-se que o presente texto tenha sido capaz de esclarecer e desmistificar um pouco esses conceitos que, por vezes, ainda parecem distantes para o jurista.

REFERÊNCIAS

CHANDRIKA, G. Naga *et al.* Web Scrapping for Unstructured Data Over Web. *Em*: BHATEJA, Vikrant; SATAPHATY, Suresh; SATORI, Hassan (org.). **Advances in Intelligent Systems and Computing: Embedded Systems and Artificial Intelligence**. Singapore: Springer, 2020. v. 1076, p. 853-859. *E-book*. Disponível em: <http://www.springer.com/series/11156>.

GOLD, Zachary; LATONERO, Mark. Robots Welcome? Ethical and Legal Considerations for Web Robots Welcome? Ethical and Legal Considerations for Web Crawling and Scraping Crawling and Scraping. **Washington Journal of Law, Technology & Arts**, [s. l.], v. 13, n. 3, p. 277-311, 2018. Disponível em:

<https://digitalcommons.law.uw.edu/wjltahttps://digitalcommons.law.uw.edu/wjlta/vol13/iss3/4>.

KROTOV, Vlad; JOHNSON, Leigh; SILVA, Leiser. Tutorial: Legality and ethics of web scraping. **Communications of the Association for Information Systems**, [s. l.], v. 47, n. 1, p. 539-563, 2020.

[1] GOLD, Zachary; LATONERO, Mark. Robots Welcome? Ethical and Legal Considerations for Web Robots Welcome? Ethical and Legal Considerations for Web Crawling and Scraping Crawling and Scraping. **Washington Journal of Law, Technology & Arts**, [s. l.], v. 13, nº 3, p. 277-311, 2018..

[2] CHANDRIKA, G. Naga et al. Web Scrapping for Unstructured Data Over Web. *Em*: BHATEJA, Vikrant; SATAPHATY, Suresh; SATORI, Hassan (org.). **Advances in Intelligent Systems and Computing: Embedded Systems and Artificial Intelligence**. Singapore: Springer, 2020. v. 1076, p. 853-859. *E-book*. Disponível em: <http://www.springer.com/series/11156>.

[3] KROTOV, Vlad; JOHNSON, Leigh; SILVA, Leiser. Tutorial: Legality and ethics of web scraping. **Communications of the Association for Information Systems**, [s. l.], v. 47, nº 1, p. 539-563, 2020.

[4] "Robot.txt" são arquivos comumente inseridos dentro da raiz dos sites que tem por intuito instruir aos programadores acerca de quais páginas e quais informações poderão ser acessadas



por meio da utilização de *web crawlers* ou *web scrapers*. Ou seja, quando determinado site insere tal arquivo dentro do seu código-fonte ele, basicamente, está informando que tais softwares não são bem vindos.

Fonte: <https://conjur.jumps.com.br/2022-set-18/igor-rocha-web-crawlers-web-scrapers-direito/>