Bloomberg Law®

bloombergbna.com

Reproduced with permission. Published October 21, 2019. Copyright © 2019 The Bureau of National Affairs, Inc. 800-372-1033. For further use, please visit http://www.bna.com/copyright-permission-request/

INSIGHT: Guilt-Free Web Scraping? Not So Fast

BY HARRY RUBIN AND KAROLINA EBEL

Web scraping returned to center stage in the Sept. 9 Ninth Circuit decision that affirmed a preliminary injunction in favor of hiQ Labs, Inc., holding that LinkedIn cannot prevent hiQ, a web scraping company, from harvesting data from publicly available LinkedIn profiles.

Overall, the Ninth Circuit's decision should not be taken as a green light to scraping. The decision was not a full review of the merits and scraping cases are very fact specific. The safest way to avoid scraping is to use technology, which identifies scrapers, blocks them, and alerts the website owner.

"Web scraping'' is the collection of data from computer servers through specialized software or "robots.'' Such software simulates human web browsing to collect information from scraped websites. Collected data is either used by the scraper for internal purposes, to provide its services and products, or sold in one form or another to the scraper's clients.

hiQ used robots to gather information about employee skills and sold the information to its customers, such as eBay, Capital One and GoDaddy. hiQ also scraped information about client employees in order to assess which employees are most likely to leave their job.

After LinkedIn served hiQ with a cease-and-desist letter, hiQ sought a preliminary injunction for LinkedIn's tortious interference with hiQ's contracts. LinkedIn used a claim under the Computer Fraud and Abuse Act (CFAA) as a defense. The CFAA prohibits "intentionally accessing a computer without authorization, or exceeding authorized access, and thereby obtaining information from any protected computer.''

Circuit Splits Significantly, circuit courts have been split in interpreting "unauthorized access.'' The Second, Fourth and Ninth Circuits held that the CFAA prohibits unauthorized access by means of hacking. This means that a scraper would not violate the CFAA, so long as the *access* to information was authorized.

The First, Fifth and Eleventh Circuits, by contrast, held that even if scrapers are authorized to access and use information, they may violate the CFAA, if they *use* the information in an unauthorized manner, as is the case when scrapers violate the scrapee's website's terms of use.

LinkedIn's key argument was that after it had sent a cease-and-desist letter to hiQ, hiQ was no longer authorized to scrape any data from LinkedIn profiles. The court disagreed. It interpreted "accessing a computer without authorization'' as the action of circumventing a target website's technological access barriers, such as usernames or passwords.

Other key legal theories potentially apply to web scraping claims: the tort of trespass to chattels, breach of contract, and copyright infringement.

Trespass to chattels was successfully used in *eBay Inc. v. Bidder's Edge Inc.*, 100 F. Supp. 2d 1058 (N.D. Cal. 2000). However, since then, courts have been reluctant to accept this theory without proof of tangible damage to, or interference with, the proper function of the target website resulting from scraping.

The breach of contract theory is applicable where scraping violates the contractual "terms of use'' of a website. Such terms are generally upheld, so long as they do not contain onerous or unusual provisions.

Copyright law protects creative expression and may, therefore, protect the manner in which information is arranged on a website. However, web scraping is often only a collection of data, rather than a collection of data arranged in an original manner. A mere collection of data not arranged in any creative manner cannot be protected under copyright law, because it will not meet the originality requisite for copyright protection.

Moreover, with public websites like LinkedIn, into which users input their information, the owner of the website often does not own the scraped data in the first place. Therefore, copyright likely does not effectively protect computer servers from scraping.

Overall, the Ninth Circuit's decision should not be taken as a green light to scraping. However, the decision is merely a grant of a preliminary injunction and was not a full review of the merits. Moreover, all scraping decisions are both fact-intensive and specific, rarely representing scenarios identical to one another.

A Roadmap for Scrapers and Scrapees Nevertheless, several general themes have emerged that provide a useful and practical roadmap for scrapers and scrapees alike.

Scrapers can eliminate their exposure by entering into, and strictly abiding by, a license agreement with the targeted website owner. If this option is not available, then scrapers should be careful not to damage, slow down, or interfere with the scraped website to avoid tort claims.

Scrapers should also ensure that they do not violate the terms of use to which they assented. Even if a website user does not manifestly assent to a website's terms, courts have generally upheld terms if the user was under actual or constructive notice and is deemed to have consented to terms that are not objectively unreasonable (*Nicosia v. Amazon.com Inc.*)

Instead of relying on the courts as a first line of defense, the safest way to avoid scraping is to use technology, which identifies scrapers, blocks them, and alerts the website owner.

Website owners can avoid the burden of proving that a scraper had actual or constructive notice of the terms, if they ensure that their website's terms of use specifically prohibit scraping and that all users must affirmatively assent to the terms before accessing any information.

Website owners can also avoid making the information public. Understandably, however, this might not be a viable solution for many businesses, such as LinkedIn, whose users rely on the information being public.

In the meantime, interested parties should closely monitor the *hiQ Labs* litigation for a decision on the merits and to see whether a future Supreme Court decision will ultimately resolve the current circuit split.

This column does not necessarily reflect the opinion of The Bureau of National Affairs, Inc. or its owners.

Author Information

Harry Rubin is a partner and chair of the Technology and IP Transactions Group at Kramer Levin Naftalis & Frankel LLP in New York. He is a globally recognized authority on IP-driven transactions and the development and implementation of global IP protection, monetization and commercialization strategies.

Karolina Ebel is an associate in the Technology and IP Transactions Group at Kramer Levin Naftalis & Frankel LLP in New York. She works on corporate matters, assisting in the representation of clients in mergers, acquisitions and other strategic transactions, financing transactions, and corporate governance issues.

COPYRIGHT © 2019 BY THE BUREAU OF NATIONAL AFFAIRS, INC.