Anti-Scraping Techniques

Shubham Sandeep Zujam[1] and Prof. Bhanudas Satam[2]
Student, Department of MCA[1]
Mentor, Department of MCA[2]
Late Bhausaheb Hiray S. S. Trust's Institute of Computer Application, Mumbai, India

Abstract: *Web scraping refers to a software program that mimics human web surfing behavior by pointing to a website and collecting large amounts of data that would otherwise be difficult for a human to extract. A typical program will extract both unstructured and semi-structured data, as well as images, and convert the data into a structured format. The activity is deemed illegal, but the change in legality has not stopped people from doing the same. This paper aims to list challenges and proposes techniques to developing anti-scraping application.*

Keywords: Web scraping

REFERENCES

[1]. Renita Crystal Pereira and Vanitha T, "Web Scraping of Social Networks," International Journal of Innovative Research in Computer and Communication Engineering, pp. 237-240, Vol. 3, 2015.

[2]. Kaushal Parikh, Dilip Singh, Dinesh Yadav and Mansingh Rathod, "Detection of web scraping using machine learning," Open access international journal of Science and Engineering, pp.114-118, Vol. 3, 2018.

[3]. Sameer Padghan, Satish Chigle and Rahul Handoo, "Web Scraping-Data Extraction Using Java Application and Visual Basics Macros," Journal of Advances and Scholarly Researches in Allied Education, pp. 691-695, Vol.15, 2018.

[4]. T. W.Bell, "Internet law," http://www.tomwbell.com/NetLaw/ Ch06/eBay.html, [Online; accessed 15-Feb-2015].

[5]. P. Vixie, "cron - daemon to execute scheduled commands (vixie cron)," Unix and Linux Forums, 2010, [Online; accessed 15- Feb-2015]. [Online]. Available: http://www.unix.com/man-page/ linux/8/cron/

[6]. Wikipedia, "Discrete fourier transform — wikipedia, the free encyclopedia," http://en.wikipedia.org/w/index.php?title=Discrete Fourier transform&oldid=650317831, 2015, [Online; accessed 15-March-2015].

[7]. R. Agrawal, T. Imielinski, and A. Swami, "Mining association ´ rules between sets of items in large databases," in Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data, ser. SIGMOD '93. New York, NY, USA: ACM, 1993, pp. 207–216. [Online]. Available: http://doi.acm.org/10.1145/170035.170072