According to “Tutorial: Legality and Ethics of Web Scraping”, big web data is the data available on the web in different forms. There is three characters are vast volume in Zettabytes, vast data repository available on the web come in different format reply on technological and certain standards, it is not static or velocity as well. Data needed to be collected and organized by using technologies and tools refer to as Web Scraping. Web Scraping processes are website analysis, which is underlying structure of a website; website crawling, which is developing and running a script browses the website and retrieves the needed data; data organization. Also, Web Scraping requires understanding web technological, at least one programming language like Python. Tools like import.io, Dexi.io, Octaparse, ParseHub, OutwiHub, FMiner, 80Legs. For Web Scraping, researchers need to use of web data is legal and ethical. The article mentions legal frameworks include, illegal access and use of data, breach of contract, copyright, trespass to chattels, trade secrets. It is important for researchers to be aware of the legal and ethical issues. Next, the article mentions most Web Scraping Literature two phrases go only a few years back, and Total number of journal and conference papers explicitly dedicated to Web Scraping found in several popular databases. These papers have two categories are technical and non-technical. As technical papers, discussing technical and tools of Web Scraping like IXPath, others are about technical solutions related to data acquisition from the web.

Another broad but smaller class of web scraping papers has to do with legal and ethical web scraping. Ethics and law are distinct but complementary in that legal frameworks often cover the most obvious ethical issues surrounding a particular practice, which can be codified in relatively clear legal form. A review of the literature and cases also reveals that more subtle ethical issues surrounding Web Scraping are also frequently addressed in the legal literature, but rarely explicitly discussed. Various laws prohibit any illegal or fraudulent use of data obtained through web scraping like CFAA, “Terms of Use”, “Terms of Service”. Next, Scraping and republishing data owned by the site owner and expressly copyrighted can lead to "copyright infringement" cases, especially when a party uses the scraped data for financial gain. Also, if web scraping overloads or damages a website or web server, the person responsible for the damage can be held liable under trespassing on movable property. Web Scraping should not be used as a deliberate surveillance mechanism that intends to reveal trade secrets of a competing organization. While existing law and legal theory have been applied to web scraping in courts and in the legal literature, ethical issues of web scraping have been addressed to a limited extent. The first ethical consideration to address is whether a website contains a robots.txt file that prohibits automated web crawling. Next, research projects that rely on data collected from a website may inadvertently compromise the privacy of individuals involved in the activities offered by the website, and even if personal privacy is not violated, the problem is that the website's customers may not have consented to any third party's use of their data. Information from web scraping activities can lead to discriminatory practices, inferences based on pre-existing biases, and biased labeling. Automated web scraping can inadvertently reveal confidential information about the operations of the organization owning the crawled website. Also, organizations and government officials often make strategic decisions based on data gathered through web scraping campaigns. Due to the accuracy of web data, this can lead to wrong decisions.