PIONEER
ACADEMIC PUBLISHING LIMITED

Studies in Law and Justice
ISSN 2958-0382
www.pioneerpublisher.com/slj
Volume 3 Number 2 June 2024

# Exploring the Path of Personal Information Infringement Governance by Web Crawlers

**Yan Liu[1], Ping Zheng[1], Yutong Li[1], Mengxuan Li[1] & Xin Zhang[1]**

[1] Tianjin Normal University, China
Correspondence: Yan Liu, Tianjin Normal University, China.

doi:10.56397/SLJ.2024.06.03

**Abstract**

In the era of big data, the explosive growth of data information, network crawler as a new type of data search engine came into being, and gradually become indispensable technology. However, with the continuous penetration of crawler technology, related infringement problems have gradually surfaced. This topic starts from the definition and technical application of network crawlers, further analyzes the specific forms of network crawlers' personal information infringement, and discusses and summarizes the shortcomings of the legal governance measures involved, and finally draws on the experience of overseas and combines the views of relevant scholars, proposes to improve the relevant legal provisions and clarify the division of responsibility; introduces the rule of balanced trial; establishes the criteria for the reasonable use of data by crawlers; creates a "cyclical protection system"; establishes the "data protection system"; and establishes the "data protection system". The proposal to establish a cyclical protection system, establish a "data protection officer", set up an independent regulatory body, form an internal and external linkage mechanism, and then build a scientific governance system for scientific governance.

**Keywords:** web crawler, personal information, governance path

## 1. Introduction

In contemporary society, Internet technology is prevalent, and the transformation of technology makes more and more data tools appear. Web crawler is an important contemporary data acquisition tool, which as a program or script, can simulate manual clicking, access to websites and other computer systems, with high efficiency, comprehensiveness and precision, and thus is often used as a tool to capture data, which can be said to be wherever there is data, there is a web crawler. However, as data plays an increasingly important role, the phenomenon of illegal activities of cyber crawlers has become more frequent, and various cases of data infringement have been disclosed. However, at present, China's legal norms and policies related to network security, data crime are not complete and need to be improved. Therefore, it is of some significance to explore the path of personal information infringement governance for web crawlers. From the theoretical aspect, the network crawler is concerned about the independence of data legal interest; from the practical level, although the technology is neutral, but the use of technology behavior has but there is a difference between right and wrong, the illegal use of the network crawler is

very easy to induce illegal and criminal behavior, increase the convenience of the implementation of criminal behavior. As an emerging thing, the exploration of network crawlers personal information infringement governance path can better promote the balance between technological development and social value, so that the relevant regulation is neither overstepping nor missing, so that the legal rules with a positive attitude to deal with the raging technological trend.

## 2. The Definition and Technical Applications of Web Crawlers

### *2.1 Definition of Web Crawler*

Web crawler is a kind of script or program that dynamically crawls a large amount of Internet data according to certain proposed rules, and through code editing, it can realize the purpose of automatically obtaining target data in batch under the target base station, and its programmable and controllable performance includes but is not limited to information acquisition, information extraction, data storage and so on. Network crawlers are broadly divided into the following three types: First, general-purpose network crawlers, mainly for non-specific websites for search engines to capture a wide range of topics and download them to the local area to establish a mirror backup, and its target resources cover the entire Internet, with a strong application value. Second is the theme web crawler, a crawler program oriented to the needs of a specific theme, which crawls data for a specific website or a certain theme defined in advance, and screens the content when crawling the webpage, thus filtering out the webpages that are not related to the theme. Third, incremental web crawler, due to the data resources on the Internet are in real-time changes, in order to ensure that the crawled is as new as possible, technicians usually combine incremental web crawler, on the basis of having saved part of the data of the target website, take incremental update of the downloaded webpage, only crawl the newly generated webpage and do not repeat to get the unchanged webpage resources, so as to save time and hardware resources.

### *2.2 Technical Application of Web Crawler*

First of all, in the field of search engine, network crawlers are most widely used, and they are an important part of search engine systems such as Google, Baidu, Yahoo and so on. Search engine through the network crawler will be hidden in all corners of the Internet space of information for search, summarize and then classified, sorted, and the establishment of the index library, so as to provide users with keyword search function. It can also use algorithms to automatically sort out hot news, and even realize targeted dynamic update push according to user's selection preference. With the continuous innovation of various business models, operators use web crawlers to develop endless application scenarios, and crawler technology is increasingly becoming a powerful tool for news media (such as daily headlines), content platforms (such as Facebook, microblogging) and other large-scale Web service providers to collect news and information, hot ranking, dynamic push.

Secondly, in the field of scientific research, web crawlers can provide researchers with powerful data support, becoming a new tool for high-quality scientific research in the big data environment. For example, in economics research, the efficient use of web crawlers can help economists batch access to literature resources or statistical data for research objects, automatically fill out questionnaires, and can automatically organize the data into the format needed for empirical research, eliminating the high cost of purchasing various databases.

In addition, web crawler technology is also widely used in public opinion analysis, data collection and visualization in weather forecasting, batch detection of web page vulnerabilities, competition for data resources among enterprises, data trading and other fields.

## 3. Web Crawler Data Infringement Patterns

### *3.1 Illegal Collection*

Malicious network crawlers in violation of the "Robots Agreement"[8] to capture large amounts of data, will be captured to the personal data such as special data, crawling the personal information that is not disclosed in accordance with the law. Such as crawling Taobao seller transaction information, from which to obtain the corresponding Taobao user's transaction

[8] The Robots Exclusion Protocol (REP), also known as a robots.txt file, is a text file in the root directory of a website that guides web crawlers (such as search engine spiders) on how to crawl and index the content of that website. This protocol is a non-mandatory, voluntary-based specification that informs crawlers which pages or files can be crawled and which should be excluded.

data, telephone, home address and other sensitive personal information sold to the relevant profit-making organizations, resulting in citizens by the unknown phone calls or text messages intrusion, disturbing the peace and quiet of the individual's life or even deceived and caused by the loss of property.

Utilizing crawler technology to invade the computer information system to obtain the data stored, processed and transmitted in the computer information system, and the seriousness of the situation may even constitute the crime of illegally obtaining the computer information system data.

*3.2 Illegal Provision*

3.2.1 Illegal Sale and Dissemination of Personal Data to Promote Telecommunication Fraud, Obscene and Pornographic Profits and Other Related Crimes

According to the relevant provisions of the law, if the malicious network crawler crawls prohibited in the market circulation of prohibited information, such as obscene articles or containing obscene videos on the network disk, and then dissemination may constitute the crime of dissemination of obscene articles; if the dissemination of obscene articles through the sale of obscene articles to earn other costs or even constitute the crime of dissemination of obscene articles for profit.

3.2.2 Illegal Sale and Dissemination of the Gray Industry such as "Loan Sharks" and "Violent Debt Collection"

In order to obtain benefits, the data company sells the personal information of the citizens crawled by the network crawler technology to the illegal collection agency, and indirectly helps the collection agency to carry out violent collection and other criminal activities, for example, the data company grabs a large amount of various types of data of the borrower's mailboxes and cell phone desktops on the website and "hands over" the borrower's address and location information to the collection agency. The company to collect money back, or other private information of borrowers sold to other platforms as a basis for credit risk control decisions, the implementation of the above criminal activities can not be separated from the crawler technology "with", so illegal crawling will not only cause the leakage of personal information of citizens and even contribute to the development of related gray industry.

*3.3 Illegal Use*

In the market economy, some dominant operators rely on their own technical advantages to use crawlers to maliciously steal the fruits of labor of other enterprises. For example, in the classic case of Dianping v. Baidu, Baidu crawled the user review information on Dianping.com through the technical means of network crawlers and used it for its Baidu Maps and Baidu Know, which violated the law. Know", violating business ethics and the principle of honesty and credit; or to restrict other competitive enterprises to obtain data, suppressing other operators in the relevant field, constituting unfair competition and disrupting the normal order of market competition.

*3.4 Illegal Intrusion, Destruction of Computer Information Systems*

The use of crawler technology to intrude into national affairs, national construction, cutting-edge scientific and technological fields of computer information system behavior, violation of national information network security. The use of web crawlers to crawl information resulted in increased traffic, slowed system response or even paralysis, interfering with the normal operation of the crawled website and affecting the normal use of citizens. For example, Yang Jieming authorized Zhang Guodong, an employee of the company, to develop a software called "Fast Pigeon Credit System", which used network crawlers to query the information on Shenzhen residence permits for 151,140 times within two hours and saved the queried information in the form of a Mouyun's network cloud disk, causing the Shenzhen residence permit system to be unable to operate normally during that period of time, which greatly affected the operation of the system, normal operation, greatly affecting the use and daily operation of this residence permit system.

*3.5 Automated Decision-Making*

Web crawlers can automatically analyze and evaluate an individual's behavioral habits, hobbies, or economic, health, credit status, etc. through computer programs and make decisions. The rights and interests of individuals may be infringed upon in the process of automated decision-making by personal information processors using the collected personal information, such as what we often call

big data killing.

## 4. The Network Crawler Technology Legal Governance Analysis and Improve the Proposal

### *4.1 China's Existing Governance Measures Briefly Analyzed*

#### 4.1.1 Existing Legal Regulation in China

First of all, China's government has issued a number of regulations involving the protection of personal information, gradually forming a governance pattern of pluralistic and common governance.

Secondly, Article 253, Article 251, Article 286, Article 286 one of the Criminal Law of the People's Republic of China focuses on the protection of personal information from the link of personal information to clarify the elements and legal consequences of the criminal elements and legal consequences of the leakage of personal information through the sale, illegal sale, and failure to fulfill the obligations of network security management and other circumstances, and to make specific provisions for the aggravating circumstances and the circumstances of the unit crime. In 2017, the The Interpretation of the Supreme People's Court and the Supreme People's Procuratorate on Several Issues Concerning the Application of Law in Handling Criminal Cases of Infringing on Citizens' Personal Information provides a clear definition of citizens' personal information and stipulates the aggravating circumstances of the relevant provisions of the criminal law in the method of generalization plus enumeration. Chapter 4 of the Civil Code of the People's Republic of China provides for the right to personal privacy of natural persons, the right to inviolability of personal information, as well as the obligations undertaken by various types of information processors. In the field of economic law, Article 14 and Article 29 of the Protection of Consumer Rights and Interests Law, from the rights and interests enjoyed by individual consumers and the obligations borne by operators to make it clear that consumers' personal information is respected and protected operators should follow the lawful, Article 2 of the Anti-Unfair Competition Law, Article 12, and Article 18 of the Opinion Draft of the General Administration of Market Supervision and Administration of the People's Republic of China (GAMSA) of November 2022, from the point of view of the use of illegal data to carry out unfair competition. The perspective of legitimate competition provides for the protection of personal data, the penalties for the illegal use of personal data, which details the breach of contract and the grabbing of data without justifiable reasons to infringe on the rights and interests of other operators and consumers, and the use of honesty and trust and business ethics for underwriting.

Finally, the introduction of the Personal Information Protection Law of the People's Republic of China marks the achievement of China's goal of constructing a fundamental law for the information society, which incorporates the collection, transportation, deletion and other related aspects of behavior into the scope of legal regulation, and adopts a mixed legal system of administrative and criminal law, which makes the regulation of personal information infringement issues more complete.

#### 4.1.2 Practical Measures

The Ministry of Public Security, based first on information security and data security, has actively participated in the formulation of relevant policies and regulations, promoted the introduction of the Personal Information Protection Law and the Data Security Law, and, in conjunction with the Central Internet Information Office and other departments, has formulated and issued the "Methods for Determining the Illegal and Unlawful Collection and Use of Personal Information by APPs," the "Scope of Necessary Personal Information Provisions for Common Types of Mobile Internet Applications," and other provisions, which clearly define standards for the determination of personal information. The government has also taken measures to improve the protection of personal information, industry norms and corporate responsibility, and carried out administrative law enforcement work in parallel, inspecting 55,000 Internet enterprises, handling more than 13,000 administrative cases, punishing a number of cell phone APP operating companies that collect citizens' personal information in excess of the scope, patching more than 3,000 security holes, and supervising Internet head enterprises to improve the system of collecting, storing, and using users' data. Secondly, relying on the "Clean Net" special operation, a long-term cooperation mechanism has been established with the Central Internet Information Office, the Supreme People's Court, the Supreme People's

Procuratorate, the Ministry of Industry and Information Technology and other relevant units, to promote cooperation in a variety of areas, including strict punishment of crimes, focusing on remedial measures, strengthening supervision of the industry, standardizing the handling of cases in accordance with the law, and carrying out publicity and education, so that a mechanism has been established for the protection of citizens' personal information and data security. The law also recognizes the need for synergy in the protection of citizens' personal information and data security, and establishes a working pattern of primary, comprehensive, and systematic governance. Finally, focusing on the difficulties and blind spots of crimes and offences against citizens' personal information, it has made a heavy-handed and ruthless effort to organize and carry out special rectification. In response to the problem of telecommunication fraud caused by courier information leakage, the Ministry of Public Security, in conjunction with the Central Internet Information Office and the State Postal Bureau, jointly carried out a six-month special operation to regulate the leakage of personal information in the field of postal courier, during which a total of 206 cases of stealing and trafficking in courier information were detected, and 844 suspects were apprehended, of whom 240 were internal personnel of courier companies. In response to the problem of "AI face-swapping" leading to mass fraud, public security organs launched a special battle, solving 79 related cases and arresting 515 suspects. In response to the problem of harassing phone calls for renovation and loans, the public security authorities have joined forces with the industry and commerce departments to carry out special remediation, punishing a number of financial, renovation, property and real estate companies for illegally trading in citizens' personal information, and following the lines to break up a number of criminal gangs for telephone promotion.

*4.2 Exploration of New Governance Approaches*

4.2.1 Strictly Distinguish Between Personal Data, Personal Information and Privacy, and Clarify the Relevant Legal Concepts and Infringement Responsibilities

China's Civil Code does not strictly distinguish between personal data and personal information, and provides for personal information and privacy together, leading to the confusion of their rights. Although the Civil Code explicitly protects personal information and the right to privacy in the Personal Rights Section, in the Property Rights Section, it simply provides for the protection of data, and puts it together with virtual property on the Internet, without seeing the property attributes of personal information. At the same time, many rules in the Civil Code are ambiguous. For example, there are no specific provisions on what kind of responsibility should be borne for the infringement of personal information and how to bear it, and what protection obligations should be observed by information processors in different scenarios for different categories of information; for example, Article 107, "If a natural person finds that an information processor has violated the provisions of laws, administrative regulations or the agreement of the two parties to deal with the personal information, the natural person shall have the right to request the information processor to handle the personal information in violation of the laws, administrative regulations or the agreement of the two parties. In the case of personal information, the right to request the information processor to delete the information in a timely manner. In this provision, the decision of whether to delete or not is in the hands of the information processor, and the stolen party is in a weak position. Therefore, the author believes that our county officials legislative department should improve the relevant provisions. Externally, it is necessary to clarify the intersection between the right to personal information and the right to privacy; internally, it is necessary to clarify the various rights and interests contained in the right to personal information and the specific responsibilities of the infringer, and to set up the right to personal information and the right to property on the basis of the original right to personality, and to explore the multiple interests that exist in personal information. At the same time, to further promote the balance of rights and obligations, reference can be made to the "red flag flying principle" and the "safe harbor principle".

4.2.2 Establishment of Criteria for the Reasonable Use of Data by Reptiles and Creation of a Cyclical Data Protection System

China's criminal law and "personal information protection law" although there has been a data application stage of the existing regulations, but

through comparison can be found, the criminal law field is mainly to combat the illegal flow of personal information and the sale of behavior, the scope of protection is relatively small, in order to solve the practice of endless cases, can only rely on the continuous expansion of the judicial interpretation, there are still some of the illegal handling of personal information with a serious social harm is not by the Criminal law adjustment. Although the Personal Information Protection Law is more complete than the Criminal Law, the criminal liability provisions therein are only declaratory, and they only initially delineate the scope of infringement of personal information to enter into the public eye of criminal offense evaluation, and cannot be applied independently. The criminal law should still prevail in the pursuit of responsibility for personal information. At the same time, due to the differences in the scope of regulation between the two, the judicial process can not be properly connected, so that the application of crimes in judicial practice tends to be "pocketed," and the relevant judicial trials are mostly based on the conviction of pocketed crimes with a lower standard of proof, and the judge's discretion is too large and is prone to "different judgments in the same case" phenomenon. Phenomenon. In the author's view, the existing legislation should take timely remedial measures, the specific aspects of the application of crawler technology to make clear provisions for different aspects of the crawler's data infringement may lead to the harm and risk of qualification, clear crawler in different stages of personal data infringement of the specific legal interests infringed upon as well as should be applied to the specific crime, to promote the network crawl infringement determination "De-pocketization" creates clear standards for the courts to follow in order to promote the protection of legitimate data rights and interests and to avoid malicious lawsuits.

4.2.3 Breaking away from the Decentralized Hierarchical Management System and Establishing an Independent Supervisory Body

China has long been a decentralized hierarchical regulatory system, with many agencies and departments involved in the protection of personal information, and a lack of necessary communication and coordination among the various regulatory agencies in the operation of the supervision program. The relevant regulatory agencies usually adopt fines or administrative interviews to restrain their behavior, and the means of interviews may easily be superficial, just going through the motions without actual binding force. At the same time, under the model of "empowerment + responsibility", the binding force of private law on personal information is increasingly strengthened, failing to realize that in the age of data, in most cases, personal information does not exist independently of the individual, but is classified as a collection of groups with specific identifying characteristics. Therefore, the author believes that the government should set up an independent regulator to coordinate and plan the relevant regulatory work, clarify the authority and responsibility of each authority, break the rigid hierarchical regulatory system, promote the implementation of control through unified management, avoid the abusive consumption of relevant resources, and create an order that promotes the smooth advancement of the freedom of information and the protection of information.

4.2.4 Create an Internal and External Linkage Mechanism Based on the Overseas "Data Protection Officer System"[8]

The protection of personal information is multifaceted, and the government, enterprises and individuals are indispensable to the promotion of personal information security. Enterprises, as the root cause of the problem and its solution, should be fully concerned about what they do. In today's era, when the Internet is in full swing and data has become an important commercial asset, some enterprises, driven by profit, violate the "Robots" agreement by illegally obtaining, storing, using, or even trading personal data in order to make huge profits or engage in unfair competition. In reality, the risk of leakage and misuse of personal information can no longer be prevented by relying only on corporate self-regulation. In this regard, the author believes that we should increase the external supervision in addition to legal discipline, drawing on overseas legislative experience to set up a "data protection officer system" to promote China's major enterprises to establish a first-class capabilities, status and

[8] A Data Protection Officer (DPO) is a formal internal corporate position responsible for ensuring that the way an organization processes personal data complies with relevant data protection regulations, such as the European Union's General Data Protection Regulation (GDPR).

authority, clear rights and responsibilities of the "person in charge of personal information" department, as well as the establishment of a government-oriented "data protection officer system" to promote the security of personal information. At the same time, the government should set up a "chief data protection officer", and strengthen the independent responsibility of each enterprise for personal information protection by establishing a monitoring and preventive mechanism that combines internal self-correction of the company and external supervision by the government.

4.2.5 Clarify the Multiple Values of Personal Information and Introduce Rules for Judging Interests

Under the background of big data network, the understanding of personal information should be from the perspective of dynamization and scenario, multiple subjects, multiple scenarios, requiring a balance of different interests. However, China's existing legal regulations do not fully realize the complex nature and multiple values of personal information. For example, in the field of unfair competition, the relevant legal regulations emphasize the consideration of the interests of the operator, but the lack of measurement of the interests of the subject of consumers, in the specific trial process, most of the value of the calculation only take into account the illegal operators to make improper profits and other operators to bring economic losses, but ignored the same as the main body of the consumers of this group suffered economic losses and the leakage of personal information property value. In this regard, the author believes that, based on the era of big data, the diversified value of personal information should be re-measured, and the rule of judging interest measurement should be introduced, so that when focusing on combating market misconduct, conscious attention should be paid to the rights and interests of consumers as well as the value of personal information infringed upon, and at the same time, the specific provisions of the law should be improved to make up for the shortcomings brought about by the lag in the law itself.

**5. Conclusion**

Data is one of the most valuable resources in modern times, and its development and utilization cannot be separated from the support of advanced technology. However, technology is a double-edged sword, which can cause some problems while promoting social progress. In the era of big data, we have recognized that the network crawler technology through large-scale mining and use of network data to effectively promote the flow of data, but also lead to an increase in data security risks. As a technology, web crawlers provide an important guarantee for the rapid circulation of data and promote the development of the data industry. However, the malicious use of network crawlers repeatedly touches the bottom line of the law and constitutes a criminal offense. Exploring the governance of network crawlers is a very meaningful topic, which is not only a positive response to the law to combat cybercrime and protect data security, but also a necessary move to combat new types of crime. The study of web crawler governance is not to kill the technology, but to better utilize the technology to help the development of data.

**References**

Chen Zhuo. (2023). On the Omission and Improvement of Criminal Protection of Citizens' Personal Information. *Journal of Xi'an University of Science and Technology, Philosophy and Social Science, 40*(05), 18-23.

Feng Yuxuan, Wang Zhen. (2024). Research on the Construction of Integrated Governance Path of Administrative, Civil and Criminal of Web Crawlers. *Journal of Xi'an Petroleum University (Social Science Edition)*, (01).

Jiang JL, Zhang DD. (2024). Determination of criminal liability for overstepping the boundaries of web crawler technology use. DOI:10.13878/j.cnki.yjxk.20240228.002.

Li Zhengyan. (2024). Criminal Laws and Regulations on Web Crawler Behavior in the Era of Big Data. *Market Weekly, 37*(01), 153-158.

LIAO Tong, ZHANG Jigang. (2023). Judicial Dilemma and Optimization Path of Personal Information Protection in Digital Era. *Journal of Zhangjiakou Institute of Vocational Technology, 36*(03), 4-7.

LIU Yanhong, YANG Zhiqiong. (2020). Research on the criminalization standard and path of network crawlers. *People's Procuratorate*, (15).

Liu Yanhong. (2019). Research on the criminal regulation of network crawler behavior: taking the perspective of the crime of

infringing citizens' personal information. *Politics and Law*, (11).

Wang Mengting, Sun Beibei, Li Mengyuan. (2023). What to do when enterprise data is infringed by web crawlers — on the infringement response of web crawler technology. Shanghai: Pi Xing Fa Tan.

Xiao Zhi-Ying. (2023). Study on Civil Law Protection of Personal Data. Hebei University.

Yi Wentao. (2023). Research on Anti-Unfair Competition Laws and Regulations of Commercial Data Crawling Behavior. Nanchang University.

Zhaodi Xue. (2023). Research on the Improvement of Administrative Supervision System of Personal Information Protection. Hubei University.

Zhu Qimin. (2023). Research on the Legal Regulation of Web Crawlers Crawling Enterprise Data. Beijing: People's Public Security University of China.