УДК 004.624

ЭТИЧЕСКИЕ АСПЕКТЫ СБОРА ПЕРСОНАЛЬНЫХ ДАННЫХ В ИНТЕРНЕТЕ: РЕКОМЕНДАЦИИ ДЛЯ БИЗНЕС-ЛИДЕРОВ

Кошкаров А.В.

к.т.н.,

Астраханский государственный университет

Астрахань, Россия

Аннотация

Компании могут использовать различные данные в своем бизнесе с целью принятия более точных и эффективных решений. Часто данные собираются из Интернета с помощью автоматизированных скриптов. Это может нарушать этические и правовые нормы, если происходит сбор, в том числе, и персональных данных. В данной статье рассматриваются этические аспекты хранения и сбора персональных данных в Интернете и даются рекомендации бизнес-лидерам касательно манипуляций с данными в маркетинговых или исследовательских целях.

Ключевые слова: сбор данных, этика данных, персональные данные, наука о данных, данные и общество

ETHICAL ISSUES RELATED TO SCRAPING PERSONAL DATA: RECOMMENDATIONS FOR BUSINESS LEADERS

Koshkarov A.V.

PhD.,

Astrakhan State University,

Astrakhan, Russia

Annotation

Companies can use data in their business to make more accurate and effective decisions. Often, data can be collected from the Internet using automated scripts. This may violate ethical and legal norms, if personal data are collected. This article discusses the ethical aspects of storing and collecting personal data on the Internet and gives recommendations to business leaders about manipulating data for marketing or research purposes.

Keywords: data scraping, data ethics, personal data, business data, data science, data and society

Introduction

Internet is not only a channel of communication between users, businesses and machines, but also it is a warehouse of all kinds of information that can be collected for various purposes by using modern technologies. Very often, user personal data are openly available. This happens either due to users choose to make their data public (e.g. social media), either because of the neglect to the data from the companies. These data can help companies make more accurate decisions and attract more customers by offering them more relevant products and services [4].

Computer technologies allow to collect data from the Internet quickly and automatically with the use of data scraping techniques. On the one hand, such techniques help to gather the visible information from web pages and scan the inner structure of the site to extract the additional information which is not visible on a computer screen. And if any of the users' personal information is not properly protected by the site owners, it can also be scanned, retrieved and processed. On the other hand, many websites provide API (Application Programming Interface) to be able to automatically extract the required information. A data scraper (web robot) can access as many Internet pages as necessary, step by step, can explore content on the pages to find and extract desired data and its structure [1].

There are many software solutions for data scraping, and the technical ease of data collection carries an increasing potential benefit for researchers and businesses. Business leaders can use this technology for marketing purposes. But a cooperation between data hosts and web scrapers can often be beyond the scope of the law, and it should be taken into account at the top level.

The key legal and ethical challenges

Data scraper can accidentally or intentionally collect the personal data during web scraping. According to EU law, personal data are “data which relate to a living individual who can be identified from those data and/or other information which is in the possession of, or is likely to come into the possession of, the data controller” [6, 6]. The open availability of such information could negatively affect not only the person whose personal data has been collected, but also on the company (data host), which allowed the collection of such data. Moreover, the collected personal data may be processed and subsequently published in the public domain.

An example of poor practice is OKCupid case. Researchers from Denmark using data scraping software collected and published a large collection of user data from OkCupid dating website. These data also contained personal information (e.g. age, gender, interests). Despite the fact that the real names have not been published, it was possible to identify the person by other variables [3]. A large stream of criticism has fallen on the researchers and the company. Researchers were sure that they do not break the law, but the ethics issues of ongoing actions moved to the forefront. OkCupid changed data policy after that.

Top manager should know that what is ethical, is not always legal. And what is legal is not always ethical. Companies should protect user data from external access to enhance the credibility of the company from the target audience in terms of ethics [5]. Before user registration, the websites offer to accept the user agreement with the data use policy, and this policy should not contradict with existing legislation.

Uncontrolled data gathering from a website may carry several threats. It is a leak of personal data, which may be published in the public domain. In this case, the data owners can accuse the company of violating their privacy, and the company may incur reputational risk of losing their investment value and public trust (especially if it is a big company). In addition, data scrapers can act as a parasite, gathering the necessary information from the site, slowing the speed of the site, and even damage the site, if the data collection script is not working properly [2]. It is important for top managers to consider the issues of ethics and business interests together. Proper interaction between data owners and data scrapers can be beneficial symbiosis, which will benefit both sides of the process. By using API technology, data owner can restrict the list of available data from the site, thereby protecting the rights of their users.

The key principles for data ethics

There are several key principles [7] that organisations could put into practice when addressing the legal and ethical challenges related to scraping personal data from the web.

1. A respectful attitude towards user data is the number one priority. User data should not become a source of open access, the data should be adequately protected, and their use should not go beyond the user agreement.

2. It is necessary to seek to justify the expected conditions of security and privacy with the actual ones. Users expect the protection of their data by default, and these expectations must be justified.

3. Always keep to the law, but it is not a sufficient condition. Technologies are developing rapidly and the law does not have time to adapt to them. Business leaders have to also act in accordance with their own ethical standards.

4. The specialists who process the data must be qualified and to act in accordance with professional standards. Incorrect results of the analysis or improper use of the data by non-professionals may have a negative effect on the company and society in general.

5. Methods and techniques of providing the data must be clear, transparent, verifiable and accountable. There should be no contradictions and breaches in the data exchange.

6. New developments and internal projects must be subject to ethical review. New projects (especially the new research programs) within the company should follow the general ethical standards. This will help to reduce risks and contribute to the public trust.

7. Data security policy, process control methods, ethical standards should be known to all key members of the organisation and need to be periodically reviewed according to the new conditions and/or changes in legislation. Many organisational issues (including ethical issues) can not be solved alone, it needs a team in which all members understand the relevant issues.

Conclusion

Ethical issues in collecting data for research and business purposes should not be ignored. Internet users (especially users of social networks) may neglect the rules of publishing their data (including personal data), but this does not mean that the collection of such data is ethically correct. Companies storing their users' data on their site and companies that collect data for various purposes should take all necessary actions to protect and disclose information from third parties. This can affect the reputation and level of trust in the company that give value to the company's brand.

References

1. Glez-Peña D. et al. Web scraping technologies in an API world // Briefings in bioinformatics. – 2013. – Vol. 15. – №. 5. – pp. 788-797.
2. Hirschey J. K. Symbiotic relationships: Pragmatic acceptance of data scraping // Berkeley Technology Law Journal. – 2014. – Vol. 29. – pp. 897-927.

3. Kirkegaard E. O. W., Bjerrekær J. D. The OKCupid dataset: A very large public dataset of dating site users // Open Differential Psychology. – 2016. – Vol. 46.
4. Koshkarov A.V., Koshkarova T.A. Data science in e-commerce: benefits for society // Innovatsii i perspektivy sovremennoi nauki. Fiziko-matematicheskie nauki. Tehnicheskie nauki: sbornik trudov molodih uchenyh [Innovations and perspectives of modern science. Physics and mathematics. Technical sciences: a collection of works of young scientists] [Electronic resource]. – Astrakhan: «Astrakhanskiy universitet» Publ., 2018. – pp. 106-109.
5. Sipior J. C., Ward B. T., Rongione N. M. Ethics of collecting and using consumer internet data // Information Systems Management. – 2004. – Vol. 21. – N. 1. – pp. 58-66.
6. The Guide to Data Protection [Electronic resource] // Information Commitioner's office. 2018. URL:https://ico.org.uk/media/for-organisations/guide-to-data-protection-2-10.pdf (accessed: 01.03.2018).
7. Tiell, S., Metcalf, J. Universal principles of data ethics: 12 guidelines for developing ethics code // Accenture. – 2016. – 12 p.