Selected Papers of #AoIR2019:
The 20th Annual Conference of the
Association of Internet Researchers
Brisbane, Australia / 2-5 October 2019

WORK IN PROGRESS: THE EUROPEAN "RIGHT TO BE FORGOTTEN" – LEGAL AND TECHNICAL CHALLENGES OF SEARCH ENGINES COMPLYING WITH THE RIGHT TO ERASURE

Jan Rensinghoff, Tobias Gostomzyk
Technische Universität Dortmund, Germany

Florian M. Farke, Markus Dürmuth
Ruhr University Bochum, Germany

Introduction

The new European *right to erasure* (‘*right to be forgotten’)* in Art. 17 of the *General Data Protection Regulation* (GDPR) of the *European Union* (EU) grants EU citizens the right to demand the erasure of their personal data from any data processor. This obligation applies, for example, to webshops, social networks, and news sites. Even search engines can be responsible for processing personal data. Search engines are in a special position because they allow their users to quickly search a large part of the World Wide Web for a specific topic or name within a fraction of a second. At the same time, they also potentially endanger people’s rights because of their power to determine which search results they deliver and which not.

Furthermore, search engines allow family members, neighbors, or employers to search the Web for personal data about almost everyone quickly. A recent study shows that three out of four German Internet users already searched for their name via search engines and every fifth even once a month[1]. The *right to be forgotten* therefore is the legislative answer to the uncontrolled spread of personal data through the Internet – at least on the EU level. This right aims to help people regain control over the distribution of personal data.

[1] Bitkom Research Survey, available in German. Retrieved February 20, 2019 from https://www.bitkom.org/Presse/Presseinformation/Grosse-Mehrheit-der-Internetnutzer-googelt-sich-selbst.html

Suggested Citation (APA): Rensinghoff, J. & Farke, F(2019, October 2-5). Work in Progress: The European "Right To Be Forgotten" – Legal And Technical Challenges of Search Engines complying With The Right To Erasure. Paper presented at AoIR 2019: The 20th Annual Conference of the Association of Internet Researchers. Brisbane, Australia: AoIR. Retrieved from http://spir.aoir.org.

The "Right to Erasure" and Its Legal Framework

Even with the *right to erasure*, it is not necessary to delete all objected personal data. There are no static criteria when personal data should remain on the Internet, and when the data controller has to delete it. However, the decision to remove data does not only affects the data subject. It also can influence the general public, which has a right to distribute and receive data without interference. It also can have an impact on the data source because of their interest in being found by search engines. Many websites rely on the revenue through advertisement and therefore on visitors brought to them by search engines. Hence, there is a need to examine any claim to remove search results and to balance conflicting rights.

Google was the first search engine operator that was required to take care of this problem after a court decision (Google vs. AEPD, 2014). Thus, they set up an *"Advisory Council to Google on the Right to be Forgotten"* in 2014. This advisory council established a list of criteria determining when to delete personal data and when to keep it online. Furthermore, *Google* started to publish an annual transparency report[2] providing insights into their decision-making process regarding the *right to erasure*.

Catalog of Criteria Concerning the Right to be Forgotten on Search Engines

Google's advisory council and their transparency report reveal a handful of criteria that are – at least potentially – involved in the decision-making process:

1. Source of the data
2. The topicality of the data
3. The importance of the data for the general public
4. The sensitivity of the data for the affected data subject
5. Whether the affected data subject deliberately published the data

All of these criteria can either promote or hinder the appearance of personal data in search results. They could be useful for machine-aided reasoning regarding the right to be forgotten.

Machine-Aided Reasoning on the Right to be Forgotten

Search engines are gateways to the Web for many people. Hence, they are probably more often than other Internet services confronted with deciding whether an individual's privacy outweighs the public's interest to lawful access to information or another legal asset. Unlike *Google*, not every search engine operator may have the resources to review every single request to delete search results appropriately. This lack of resources makes decisions unpredictable and random. However, it is also not entirely clear how *Google's* review process works.

[2] Google transparency report. Retrieved February 20, 2019 from https://transparencyreport.google.com/

None of the operators of the most used Web search engines (i. e., *Google, Bing, Yahoo!, Baidu*[3]) can or want to disclose how they conduct the reviews precisely. Their transparency reports contain some insights but are far from being a manual for a review process. We try to reverse engineer the process, from receiving deletion requests, through analyzing and assessing them, to deriving a decision. To reconstruct the process, we gather and analyze the publicly available information about the deletion request review process of search engines, e. g., from their transparency reports. We identify subtasks in the process that can be (partially) automated and those that require human intervention. These tasks range from checking the identity of the requester to assessing the relevance of the search results for the public. To automate some of the subtasks, we use machine learning techniques that we trained with examples published by search engines. Finally, we plan to model the whole process to estimate the chances of success to remove search results of a given person.

The approach we are working on is to help search engine operators and individuals to assess and decide whether search results may have to be deleted or not. Our idea is to have a system that gives suggestions on how to decide, but a human still needs to make the final decision. However, we think this will make this process more comprehensible and transparent.

Related Work

Researching the automation of legal issues is not entirely new. The idea to model legal reasoning dates back to the 1970s when scholars began to develop systems for giving advice, do legal analysis, and to construct arguments (Rissland, 2003).

Backes et al. (2015) introduced a framework for automated reasoning on privacy case law. The framework consists of formal descriptions and algorithms for reasoning tasks like the extraction of norms or deducing whether an action is legal or not. It requires the translation of all case information needed for the reasoning into a set of formal rules. They designed their framework to be agnostic from the underlying legal system but focus on US privacy regulations like HIPAA or COPPA and therefore on case law.

Focusing on the right to be forgotten of the GDPR, Tiwari et al. (2018) implemented and extended the reasoning framework of Backes et al. (2015). They introduced a *similarity measure* to determine the similarity of cases and to allow to decide new but slightly different cases in an automated way. Furthermore, they implemented the framework using first-order logic to evaluate its run time performance.

[3] Worldwide desktop market share of leading search engines from January 2010 to October 2018. Retrieved February 20, 2019 from https://gs.statcounter.com/search-engine-market-share/desktop/europe

References

Google Spain SL and Google Inc. v Agencia Española de Protección de Datos (AEPD) and Mario Costeja Gonzálezm, Case C-131/12. (Court of Justice of the European Union 2014). Retrieved February 20, 2019 from https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX:62012CJ0131.

Rissland, E., Ashley, K., & Loui, R. (2003). "AI and Law: A Fruitful Synergy". *Artificial Intelligence*, 150 (1-2), 1-15. Retrieved February 20, 2019 from https://doi.org/10.1016/S0004-3702(03)00122-X.

Backes, M., Bendun, F., Hoffmann, J., & Marnau, N. (2015). "PriCL: Creating a Precedent, a Framework for Reasoning about Privacy Case Law". In *International Conference on Principles of Security and Trust*, 2015. Retrieved February 20, 2019 from https://doi.org/10.1007/978-3-662-46666-7_18.

Tiwari, A., Bendun, F., & Hammer, C. (2015). "A Formal Logic Framework for the Automation of the Right to be Forgotten". In *Security and Privacy in Communication Networks*. Retrieved February 20, 2019 from https://www.researchgate.net/publication/328096597_A_Formal_Logic_Framework_for_the_Automation_of_the_Right_to_be_Forgotten.