

Polisis: Automated Analysis and Presentation of Privacy Policies Using Machine Learning



Ref. Nr

6.1878

Keywords

online privacy, legal, privacy policies, automation, deep learning, machine learning

Intellectual Property

Software license

PublicationsHarkous et al., "[Polisis: Automated Analysis and Presentation of Privacy Policies Using Deep Learning](#)" in Proceedings Of The 27Th Usenix Security Symposium, 531-548.<https://pribot.org/>

Date

16/10/2020

Description

The almost universal use of the internet by both consumers and third-party service providers has highlighted the importance of appropriate data storage and privacy policies. These privacy policies are long and complex, often filled with technical and legal jargon which makes it even more difficult for consumers to understand. Nowadays, a myriad of companies and their services have become standard for most consumers, meaning they must read and comprehend dozens of different privacy policies.

The efforts made to simplify this process have not been scalable, which represents the key hurdle in this space due to the number of policies concerned, as well as their evolution over time. For example, an initiative from regulators to simplify policy elements into a grid was not taken up by many service providers. A crowdsourcing solution for popular services is also not scalable as it relies on volunteers.

This technology addresses these scalability issues by introducing an automated framework. It is built upon a hierarchy of neural network classifiers which allow scalable, dynamic and multi-dimensional queries on natural language privacy policies. Polisis consists of two applications which support structured and free-form querying, respectively.

Advantages

The application, especially the second embodiment for free-form querying, dubbed Pribot, is already in use.

At the core of the technology is a language model which is built on a set of 130,000 privacy policies. The hierarchical structure of the model allows both high-level aspects and fine-grained details to be classified.

The structured querying application of Polisis can achieve an accuracy of 88.4% when tasked with automatic assignment of privacy icons or categories from policies. Pribot can produce a correct answer in its top-3 results for 82% of free-form test questions. An example of a typical free-form question regarding the privacy policy of a service provider could be: "Do you share my *address* with other companies?", with a response: "We will provide your *location* to third parties."

Applications

- Automated interpretation of privacy policies
- Simplification of user experience