



# Tìm hiểu tổng quan về Cookie Block

Owner	@hải nam hoàng
Date	@January 9, 2024
Status	

## ▼ 1. Giới thiệu

- CookieBlock là một tiện ích mở rộng của trình duyệt tự động thực thi các tùy chọn chấp thuận GDPR của người dùng đối với cookie của trình duyệt cho bất kỳ trang web nào, bất kể trang web đó có cookie banner hay không.
- Người dùng chỉ định các tùy chọn đồng ý của họ một lần khi tiện ích mở rộng được cài đặt lần đầu tiên và sau đó CookieBlock sẽ cố gắng xóa mọi cookie không phù hợp với chính sách của người dùng khi chúng được tạo.
- Nó phân loại cookie nhanh chóng thành bốn danh mục riêng biệt và xóa những danh mục mà người dùng không đồng ý.
- Điều này nhằm đảm bảo rằng quyền riêng tư của người dùng được bảo tồn. Người ta có thể từ chối bất kỳ loại nào sau đây:
  - Cookie chức năng
  - Cookie phân tích/thống kê
  - Cookie quảng cáo/theo dõi
- Lưu ý rằng CookieBlock không tự xử lý cookie banner. Để xóa các banner gây phiền nhiễu này, nhà phát triển gợi ý nên sử dụng tiện ích mở rộng Consent-O-Matic (1 extension sử dụng tùy chọn cookie của mình người dùng đã chọn từ trước để tự động điền các cookie banner)

## ▼ 2. Hoạt động

- Mỗi khi một sự kiện cookie được đưa ra trong trình duyệt. Nếu sự kiện này cho biết rằng một cookie đã được thêm hoặc cập nhật thì tiện ích mở rộng sẽ lưu trữ cookie trong lịch sử cập nhật cookie cục bộ, sau đó thực hiện phân loại cho cookie đó.
- Danh mục cho mỗi cookie được dự đoán bằng cách sử dụng mô hình rừng cây quyết định được đào tạo thông qua trình phân loại XGBoost và một tập hợp các bước trích xuất tính năng. Đầu tiên, cookie được chuyển thành một vectơ số, sau đó được cung cấp làm đầu vào cho rừng cây. Điều này tạo ra điểm cho mỗi lớp và điểm cao nhất là lớp được gán cho cookie.
- Các phân loại cookie có sẵn là:
  - **Strictly Necessary (Thực sự cần thiết)**
  - **Functionality (Chức năng)**
  - **Analytics (Phân tích)**
  - **Advertising/Tracking (Quảng cáo/theo dõi)**
- Độ chi tiết được cố tình giữ ở mức thấp để đưa ra quyết định đơn giản nhất có thể cho người dùng. Lưu ý rằng không thể từ chối các cookie "cực kỳ cần thiết" vì đây là loại cookie bắt buộc để trang web hoạt động. Không có chúng, các dịch vụ thiết yếu như đăng nhập sẽ ngừng hoạt động.

## ▼ Known Issues

- Bộ phân loại không hoàn toàn chính xác. Có thể xảy ra trường hợp một số chức năng nhất định trên một số trang web bị hỏng do các cookie thiết yếu bị phân loại sai. Điều này khó giải quyết nếu không thu thập thêm dữ liệu cookie để đào tạo. Như vậy, cách tiếp cận này có giới hạn của nó.
- Để giải quyết những vấn đề này, nhà phát triển duy trì một danh sách các danh mục cookie đã biết. Đây là tệp JSON lưu trữ nhãn cookie cho các mã nhận dạng cookie đã biết. Nếu có cookie trong tệp này, dự đoán bị bỏ qua và lớp đã biết được áp dụng.
- Bằng cách báo cáo các trang web bị hỏng, người dùng có thể giúp nhà phát triển cập nhật danh sách các trường hợp ngoại lệ về cookie. Điều này làm cho tiện ích mở rộng trở nên dễ sử dụng hơn cho mọi người đồng thời vẫn giữ được mức độ riêng tư cao.

## ▼ 3. Triển khai

Để đạt được kết quả cuối cùng là tiện ích mở rộng Cookie Block hoàn chỉnh, nhà phát triển đã triển khai theo 4 thành phần được đánh giá riêng biệt

- 1. **\*\*The Cookie Consent Webcrawlers\*\***
- 2. **\*\*The Feature Extraction and XGBoost Classifier\*\***
- 3. **\*\*The GDPR Violation Detection Scripts\*\***
- 4. **\*\*The Browser Extension: CookieBlock\*\***

## ▼ 4. Credits & License

### Credits

- CookieBlock logo designed by Charmaine Coates.
- Czech translation provided by Karel Kubicek.
- Japanese translation provided by Shitennouji.
- Spanish translation provided by @6ig6oy.
- Automated localization in store pages performed using DeepL.

### Libraries

CookieBlock includes code from the following libraries and projects:

- **Difflib JS**
  - Ported by Xueqiao Xu xueqiaoxu@gmail.com
  - Licensed under: PSF LICENSE FOR PYTHON 2.7.2
- **Levenshtein**
  - Copyright (c) Gianni Chiappetta
  - Unlicensed: <https://unlicense.org/>
- **LZ-String**
  - Copyright (c) 2013 pieroxy

- Released under the MIT License
- **Consent-O-Matic**
  - CSS and HTML code used as basis for the interface.
  - Copyright (c) 2020 Janus Bager Kristensen, Rolf Bagge, CAVI
  - Released under the MIT License

## Thesis

This repository was created as part of the master thesis "**Analyzing Cookies Compliance with the GDPR**", which can be found at:

<https://www.research-collection.ethz.ch/handle/20.500.11850/477333>

as well as the paper "**Automating Cookie Consent and GDPR Violation Detection**", which can be found at:

<https://karelkubicek.github.io/post/cookieblock.html>

### Thesis Supervision and co-authors:

- Karel Kubicek
- Dr. Carlos Cotrini
- Prof. Dr. David Basin
- Information Security Group at ETH Zürich

See also the following repositories for other components that were developed as part of the thesis:

- [OpenWPM-based Consent Crawler](#)
- [Cookie Consent Classifier](#)
- [Violation Detection](#)
- [Prototype Consent Crawler](#)
- [Collected Data](#)

## License

**Copyright © 2021-2022 Dino Bollinger, Department of Computer Science at  
ETH Zürich, Information Security Group**