



TECHNISCHE HOCHSCHULE MITTELHESSEN

THM

**CAMPUS
GIESSEN**

MNI

Mathematik, Naturwissenschaften
und Informatik

Cookie Tracker

CS2366 Datenschutz und Datensicherheit SS21

Jan Geicke, Leon George

Dozent/in: Tobias Reimann

Cookie Tracker



Developed by Leon George and Jan Geicke

Inhaltsverzeichnis

1. Verfolgen im Internet
 - 1.1 Was ist ein Cookie?
 - 1.2 Was ist DNT?
2. Cookie Tracker
 - 2.1. Was ist der Cookie Tracker?
 - 2.2. Funktionsweise des Cookie Trackers
 - 2.3. Benutzeroberfläche des Cookie Trackers
 - 2.3.1. Hauptfenster
 - 2.3.2. Einstellungen
 - 2.3.3. Ergebnisübersicht
3. Für wen ist diese Software gedacht?
4. Wie sicher ist der Cookie Tracker?

1. Verfolgen im Internet

Viele Menschen, die häufig im Internet unterwegs sind, legen sehr viel Wert auf welche Informationen von ihnen gespeichert werden und welche nicht. Dabei spielen sogenannte *Cookies* eine sehr große Rolle.

1.1 Was ist ein Cookie?

Bei Cookies werden Informationen des Nutzers gespeichert und eventuell wieder verwendet. Dabei gibt es verschiedene Arten von Cookies¹:

- *Persistent Cookies*, bei denen beispielsweise Login-Daten gespeichert werden um diese bei dem nächsten Login nachzuladen
- *Session Cookies*, die nur für die Zeit einer "Sitzung" auf einer Internetseite leben und danach gelöscht werden
- *Third-Party Cookies*, die gesetzt werden um gezielt Informationen über den User zu sammeln um diese beispielsweise bei Werbung anzuwenden
- *Flash Cookies*, die immer auf dem Rechner des Users existieren, auch wenn alle anderen Cookies gelöscht werden
- *Zombie Cookies*, sind wie Flash Cookies und werden automatisch nachgeladen wenn sie gelöscht werden

Es ist verständlich warum viele Anwender des Internets lieber auf Cookies verzichten wollen, aber kann man wirklich etwas dagegen tun? Einer der Methoden, Cookies zu vermeiden ist mit den sogenannten *Do Not Track*-Header (DNT).

¹ <https://www.rocketlawyer.com/gb/en/quick-guides/different-types-of-internet-cookies>

1.2 Was ist DNT?

Wie bereits erwähnt dient der DNT-Header dazu, nicht im Internet verfolgt zu werden, nicht nur durch das Setzen von Cookies sondern auch jegliche andere Form von Datensammlung. Der DNT-Header funktioniert so, dass bei einer Request-Anfrage auf einer Webseite (Suchen nach etwas im Internet zum Beispiel) in dem Header-Teil des Requests der DNT-Flag auf 1, 0 oder null gesetzt werden kann.

Sollte der Flag auf 1 stehen, heißt das dass der Benutzer mit DNT im Internet suchen will, 0 heißt dementsprechend dass DNT nicht gesetzt ist. Null bedeutet dass keine spezifischen Angaben gemacht wurden.

```
GET /something/here HTTP/1.1
Host: example.com
DNT: 1
```

2

Da der DNT Header mittlerweile als 'deprecated' gilt, wurde ein ähnliches Konzept entwickelt, was als Ersatz dienen soll: *Global Privacy Control* (SEC-GPC bzw. GPC). Somit gibt es scheinbar eine gute Möglichkeit, ohne jegliche Verfolgung im Internet zu surfen. Aber leider ist das nicht immer der Fall. Es kann sein, dass trotz gesetztem DNT-Header immer noch Cookies gesetzt werden. Hier kommt der *Cookie Tracker* ins Spiel.

² <https://www.geeksforgeeks.org/http-headers-dnt/>

2. Cookie Tracker

Nun wird die Definition, Zweck und Funktionsweise des Cookie Trackers erläutert.

2.1 Was ist der Cookie Tracker?

Der Cookie Tracker besitzt einen Zweck: Das Nachfolgen und Analysieren von Cookies mit und ohne DNT-Header bzw. GPC-Header. Dabei soll zum Einen festgestellt werden, was für Cookies gesetzt werden und zum Zweiten ob diese den DNT/GPC-Header respektieren oder nicht. Im Anschluss an den fertigen Durchlauf werden dann die Ergebnisse in einem Chart dargestellt, um besser nachzuvollziehen, was geschehen ist. Wie dieser Durchlauf genau passiert, wird im Anschluss erklärt.

2.2 Funktionsweise des Cookie Trackers

Bei einem Durchlauf des Cookie Trackers finden mehrere Breiten- und Tiefensuchen statt. Ein Durchlauf startet indem eine URL angegeben wird, oder ein JSON Ergebnisobjekt geladen wird. Ist einer dieser Bedingungen (oder beide) erfüllt wird durch Betätigen des 'start'-Knopfs der Prozess gestartet.

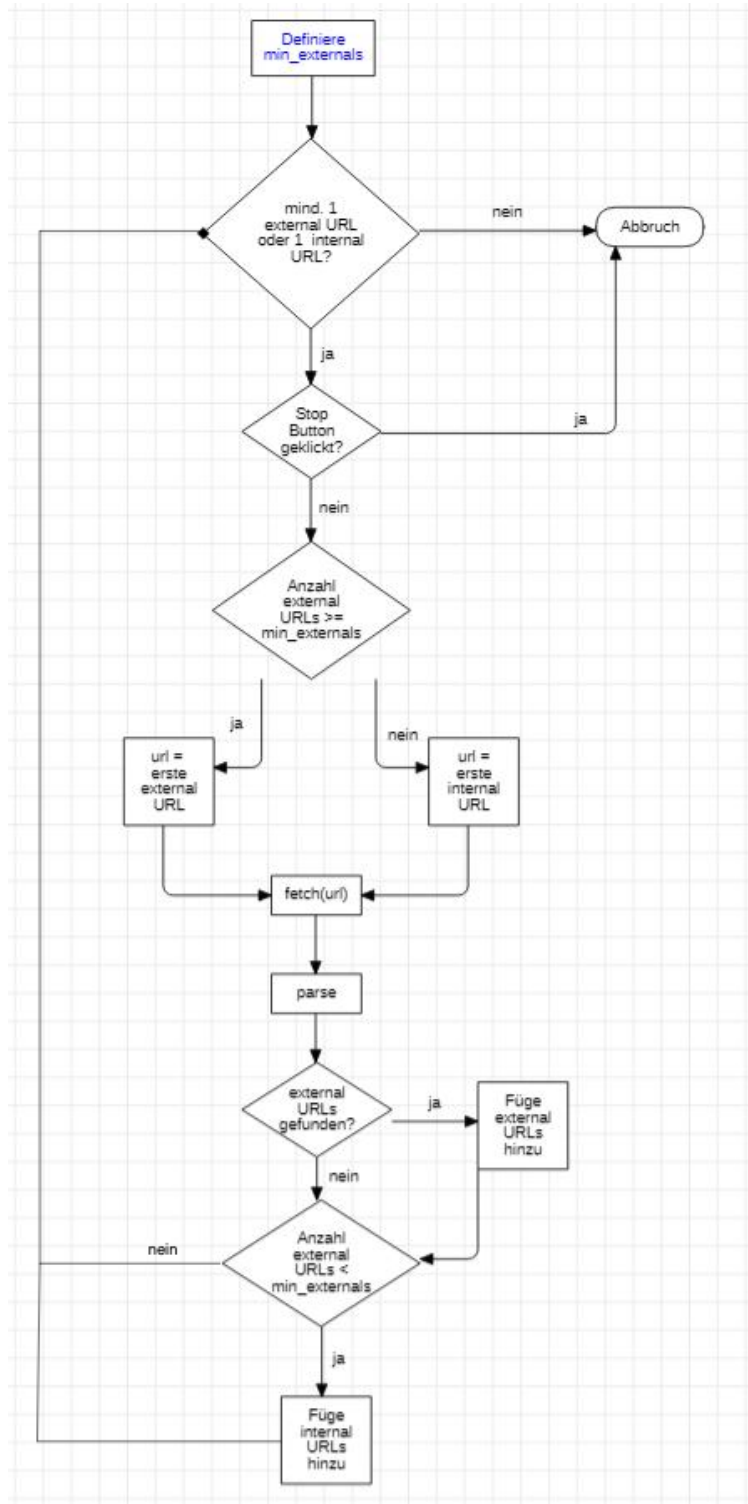
Zunächst wird die Startseite der angegeben Webseite nach externen Ressourcen durchsucht. Werden keine gefunden, wird die Webseite dementsprechend so lange durchsucht, bis eine externe Ressource gefunden wurde. Genauer wird dann von einer Breitensuche auf einer Tiefensuche der Webseite umgestellt, bis eine externe Ressource gefunden wird. Damit dieser Prozess nicht unendlich weiterläuft, wird die Anzahl gefundener Ressourcen begrenzt. Es wird zum Beispiel 5 als Maximalwert gesetzt, dann wird nur so lange untersucht, bis 5 externe Ressourcen gefunden wurden und der Prozess hört auf.

Um einen direkten Vergleich zu ermöglichen von Cookies die gesetzt werden mit DNT und die ohne, finden intern zwei Abläufe statt: Zuerst das Untersuchen der Webseite ohne DNT-Header gesetzt und dann eine zweite Untersuchung mit DNT-Header.

Dadurch könne wichtige Informationen gewonnen werden, ob der DNT-Header gesetzt wurde oder nicht.

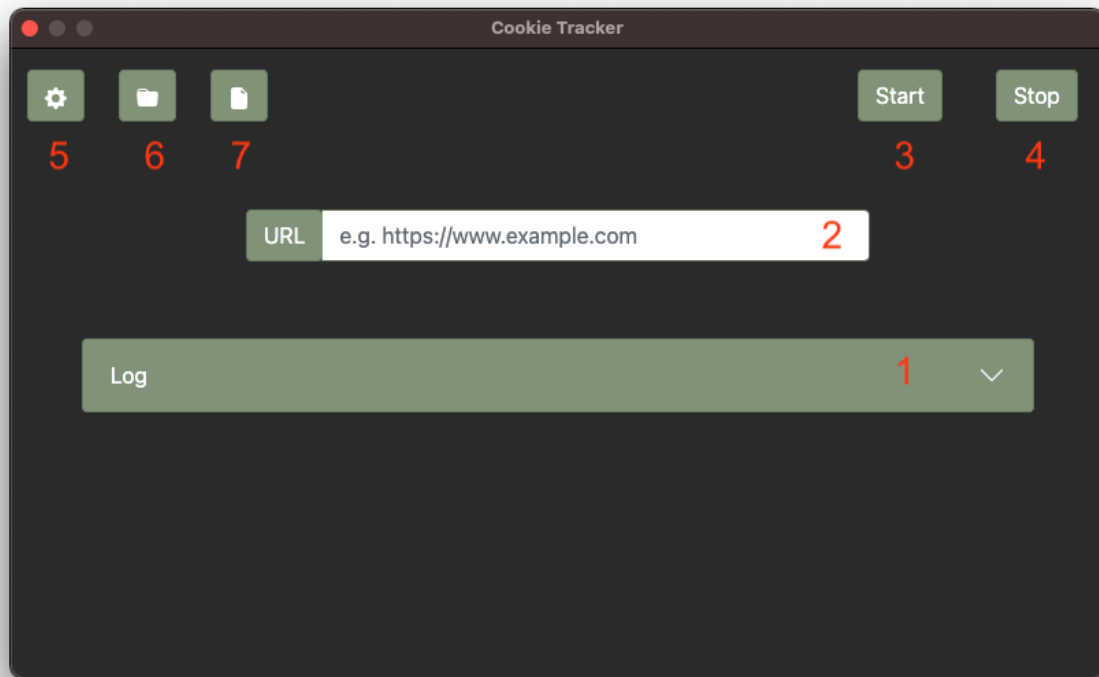
Am Ende wird das Ergebnis anhand eines Pie-Charts dargestellt, indem die Anteile der Webseiten mit persistenten Cookies zu sehen sind, sowie die Anteile, die den DNT-Header nicht respektiert haben.

Den ganzen Prozess sieht man nochmal unten in der Grafik. Der Begriff *min externals* bedeutet hier die mindeste Anzahl an externen Ressourcen, die gefunden werden sollen.



2.3 Benutzeroberfläche des Cookie Trackers

2.3.1 Hauptfenster



1. Das Ergebnisfeld, was auch als Log dient. Zeigt die gefundenen URLs an.
2. Das Inputfeld für die URL, die untersucht werden soll
3. Der Startknopf, der den Prozess startet
4. Der Stopknopf, der den Prozess anhält beziehungsweise endet
5. Einstellungen die u.A. enthalten ob DNT oder GPC verwendet werden soll und Weiteres
6. Lädt eine Datei in das Ergebnisfeld, nur JSON Dateien
7. Speichert das Ergebnisobjekt in eine JSON Datei

2.3.2 Einstellungen

Es gibt einige mögliche Einstellungen, die angepasst werden können. Diese haben direkten Einfluss darauf, wie der Cookie Tracker genau seinen Suchvorgang durchführt.

Cookie Tracker Settings

User Agent

Select the type of user agent used by the crawler.

Standard

Custom

Custom User Agent

Header

Select the headers used in the requests by the crawler.

DNT-Header

GPC-Header

Crawl Behaviour

Set the behaviour of the crawler.

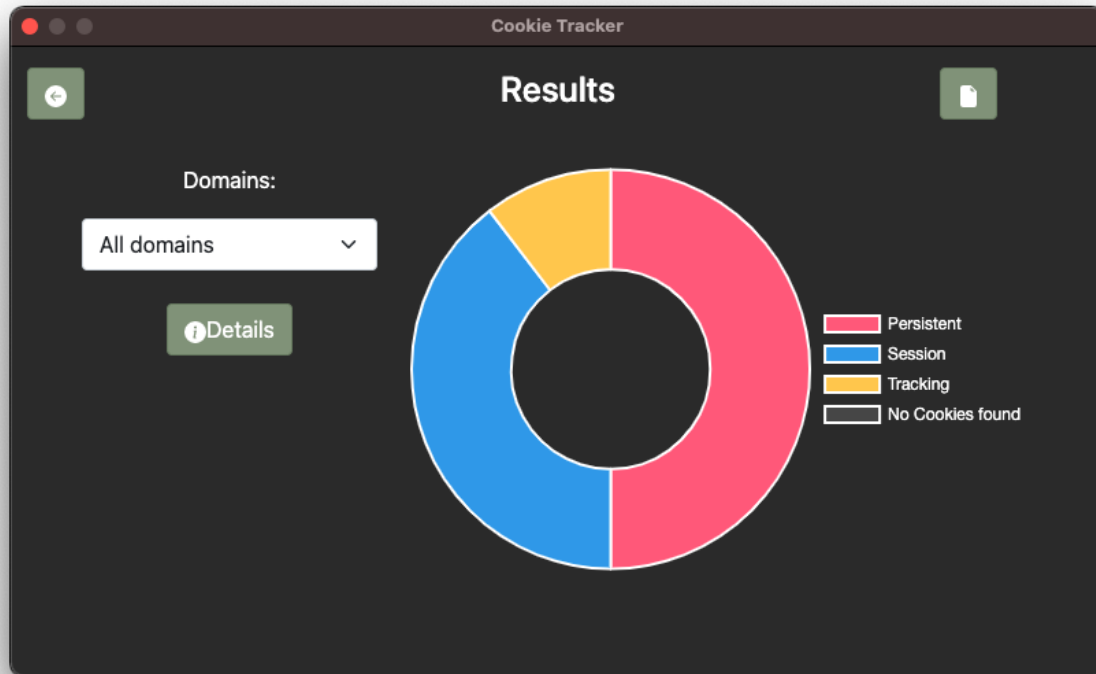
Breadth

Single Page

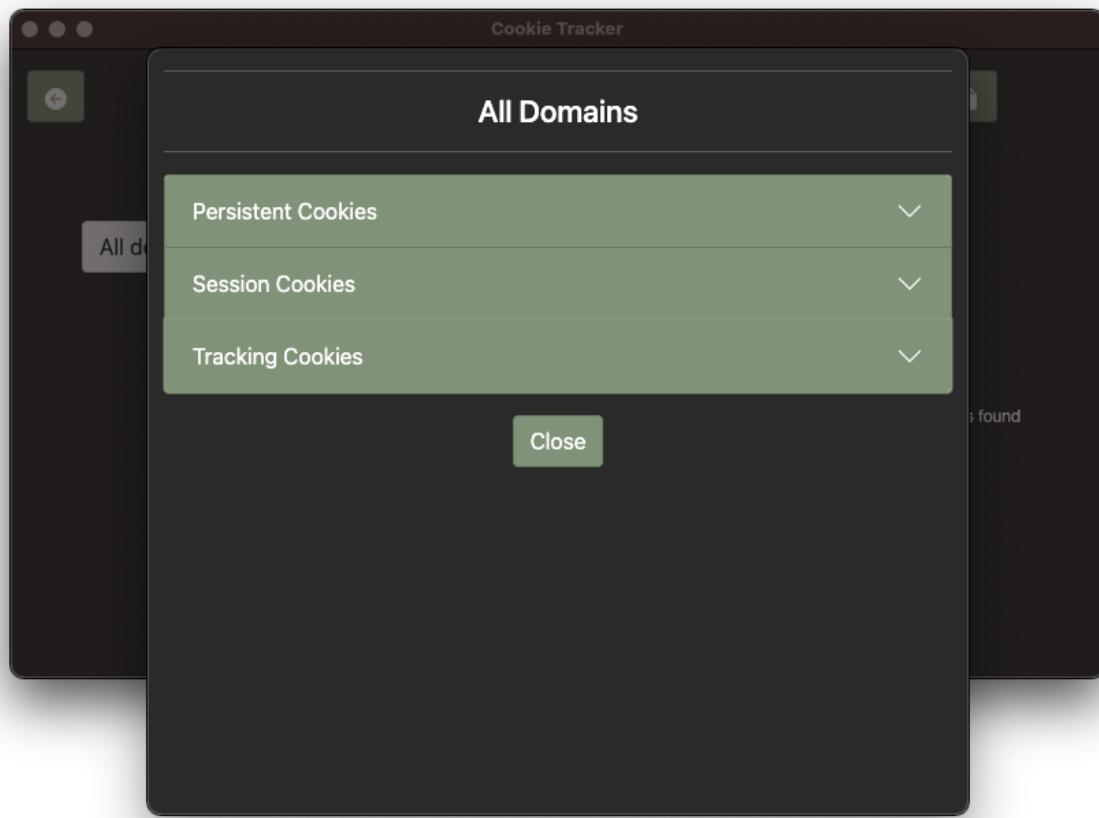
Save

In den Bilder oben ist das “Settings”-Fenster zu sehen. Zunächst hat man die Wahl, einen User Agent selber auszusuchen. Dieser kann dann in dem Input-Feld eingegeben werden und der Cookie Tracker verwenden ihn dann bei seiner Suche. Wenn nichts eingegeben wird, wird einfach der standard User Agent verwendet. Als Nächstes kann zwischen DNT und GPC Header ausgewählt werden. Der Standard Header ist hierbei DNT. Als letztes hat man noch die Möglichkeit zu entscheiden, ob eine einzelne Seite untersucht werden soll oder ob eine Breitensuche stattfinden soll.

2.3.3 Ergebnisübersicht



Oben sehen wir das Ergebnis eines erfolgreichen Tracker-Durchlaufs. In dem Pie Chart wird angezeigt, welche Persistente, Session und Tracking Cookies gefunden worden sind. Welche Cookies bei welcher Domäne gesetzt wurden kann man unter "All domains" einsehen. Dort wird eine Liste der untersuchten URLs angezeigt. Wählt man eine spezifische Domäne aus dieser Liste aus, dann passt sich der Pie Chart so an, dass man nur die Cookies von dieser Domäne sieht. Auch von diesem Fenster kann man sich das Ergebnisobjekt abspeichern indem auf den Speicherknopf oben rechts geklickt wird. Wenn man wieder zurück zum Hauptfenster möchte, dann kann man das durch betätigen des Zurück-Knopfs oben links tun. Möchte man mehr Informationen über die Cookies haben, kann man noch auf den "Details"-Button klicken:



Dort sieht man nochmal alle Cookies aufgelistet. Klickt man auf einer der Cookies, dann öffnet sich ein Dropdown Fenster in der alle Informationen gesammelt sind.

All Domains			
Persistent Cookies			
Session Cookies			
Tracking Cookies			
Domain	Name	Value	Expires
www.youtube.com	CONSENT	PENDING+819	01-01-2038 01:00
www.youtube.com	YSC	ajPFdm6gLF4	
www.youtube.com	VISITOR_INFO1_LIVE		01-11-2018 12:33
www.google.com	CONSENT	PENDING+491	01-01-2038 01:00
www.google.de	CONSENT	PENDING+266	01-01-2038 01:00

3. Für wen ist die Software gedacht?

Diese Software ist für die Menschen gedacht, die mehr Wissen wollen über wie mit ihren Daten umgegangen wird. Für viele Menschen ist es extrem wichtig, dass ihre Daten in allen Fällen möglichst sicher sind und genau für diese Menschen ist diese Software entwickelt worden. Auch Webseitenbetreiber können diese Software besonders nützlich finden, wenn sie zum Beispiel ihre eigene Seite untersuchen wollen und sehen wollen ob die Cookies die verwendet werden datenschutzkonform sind. Im Endeffekt ist diese Software aber für jeden Interessierten gedacht, der mehr darüber wissen möchte, wie sich die gesetzten Cookies auf einer Seite verhalten. Das heißt dass jeder aus dieser Software einen Nutzen finden könnte.

4. Wie sicher ist der Cookie Tracker?

Natürlich fragt man sich auch, wie sicher der Cookie Tracker selber überhaupt ist und ob irgendwelche personenbezogene Daten bei einem Tracker-Durchlauf verwendet oder verarbeitet werden. Das ist bei dem Cookie-Tracker nicht der Fall. Nur die IP-Adresse des Benutzers wird genutzt, aber sonst werden keine persönlichen Informationen verwendet wenn eine URL untersucht wird.

Abschließend also kann der Cookie Tracker guten Gewissens verwendet werden, ohne sich darüber Gedanken zu machen, ob jegliche persönliche Informationen verwendet werden.