



DSGVO Classifier

Table of contents

- Context/Motivation
- Goal
- Related work
- Approach
 - Dataset creation
 - Dataset refinement
 - Observations on the dataset
 - Naive classifier
 - SVM classifier
- Demo

Context/Motivation



Was ist easy Login?

easy Login ist eine Initiative für **die einfache Anmeldung an den Onlinediensten und BiPRO-Webservices** der Versicherer. Mit nur einem Login gelangen Sie in die Portale aller easy Login-Versicherer.

Ihre Vorteile auf einen Blick:

- easy Login ist für unabhängige Vermittler **kostenfrei**.
- Sparen Sie Zeit:** Mit nur noch einem Zugang für Ihre Versicherer.
- Passwortlisten gehören für Sie durch **einheitliche Zugangsdaten** der Vergangenheit an.

Diese Webseite verwendet Cookies

Wir setzen Cookies für verschiedene Zwecke ein (Statistiken & Nutzungsverhalten, Optimierung und Fehlerbehebung). In unseren [Datenschutzhinweisen](#) finden Sie mehr Infos und eine Opt-out-Möglichkeit.

[Nur notwendige Cookies verwenden](#)[Cookies zulassen](#)[Details zeigen](#)

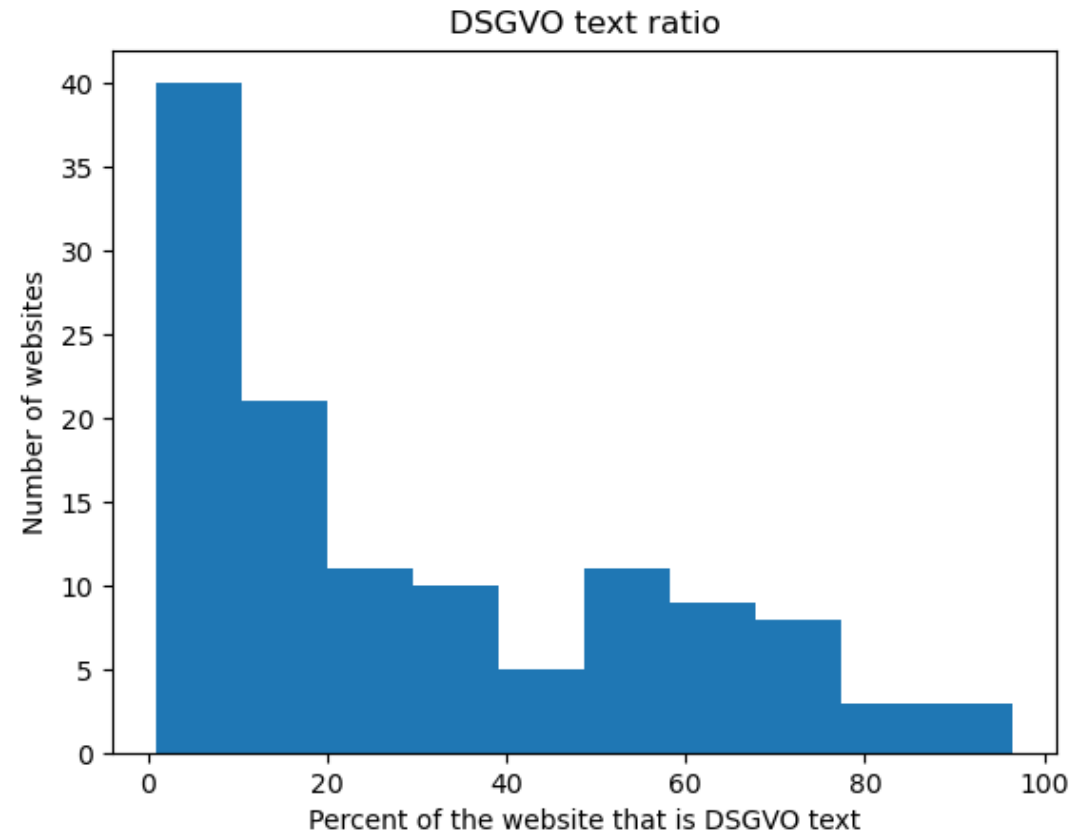
```
<!DOCTYPE html>
<html lang="de" dir="ltr" prefix="content: http://purl.org/rss/1.0/modules/content/ dc: http://purl.org/dc/terms/ foaf: http://xmlns.com/foaf/0.1/ og: http://ogp.me/ns# rdfs: http://www.w3.org/2000/01/rdf-schema# schema: http://schema.org/ sioc: http://rdfs.org/sioc/ns# sioc: http://rdfs.org/sioc/types# skos: http://www.w3.org/2004/02/skos/core# xsd: http://www.w3.org/2001/XMLSchema# " class=" js">
<head>_</head>
<body class="node 5 path frontpage page node type page">
...
  <div id="CybotCookiebotDialog" name="CybotCookiebotDialog" role="dialog" aria-describedby="CybotCookiebotDialogBodyContentTitle" lang="de" dir="ltr" ng-non-bindable style="display: block; opacity: 1; top: 12px; width: 100%; height: auto;">_</div> == $0
  <a href="#main-content" class="visually-hidden focusable skip-link"> Direkt zum Inhalt
  </a>
  <div class="dialog-off-canvas-main-canvas" data-off-canvas-main-canvas>_</div>
  <iframe name="__uspapilocator" tabindex="-1" role="presentation" aria-hidden="true" title="Blank" style="display: none; position: absolute; width: 1px; height: 1px; top: -9999px;">_</iframe>
  <!-- Ende der Zentrierung der Seite -->
  <script type="application/json" data-drupal-selector="drupal-settings-json">_</script>
  <script src="/core/assets/vendor/jquery/jquery.min.js?v=3.6.0" data-cookieconsent="ignore"></script>
  <iframe tabindex="-1" role="presentation" aria-hidden="true" title="Blank" src="https://consentcdn.cookiebot.com/sdk/bc-v4.min.html" style="position: absolute; width: 1px; height: 1px; top: -9999px;">_</iframe>
  <script src="/core/misc/polyfills/nodelist.foreach.js?v=9.4.14" type="text/javascript">_</script>
  <script src="/core/misc/polyfills/element.matches.js?v=9.4.14" type="text/javascript">_</script>
  <script src="/core/misc/polyfills/object.assign.js?v=9.4.14" type="text/javascript">_</script>
  <script src="/core/assets/vendor/css.escape/css.escape.js?v=1.5.1" type="text/javascript">_</script>
  <script src="/core/assets/vendor/once/once.min.js?v=1.0.1" type="text/javascript">_</script>
  <script src="/core/assets/vendor/jquery-once/jquery.once.min.js?v=2.2.3" type="text/javascript"></script>
  <script src="/core/misc/drupalSettingsLoader.js?v=9.4.14" type="text/javascript"></script>
  <script src="/sites/easylogin.de/files/languages/de_kcZiB0WmSNPSbosbhk2lwJAv3fXKIIIf1lBwsJrdvY.js?rt0wm6" type="text/javascript"></script>
  <script src="/core/misc/drupal.js?v=9.4.14" data-cookieconsent="ignore"></script>
```

Context/Motivation

Diese Webseite verwendet Cookies Wir setzen Cookies für verschiedene Zwecke ein (Statistiken & Nutzungsverhalten, Optimierung und Fehlerbehebung). In unseren Datenschutzhinweisen finden Sie mehr Infos und eine Opt-out-Möglichkeit. undefined Nur notwendige Cookies verwenden Cookies zulassen Details zeigen OK Nur notwendige Cookies verwenden Auswahl erlauben Cookies zulassen Notwendig Präferenzen Statistiken Marketing Details zeigen Cookie-Erklärung [#IABV2SETTINGS#] Über Cookies Notwendig (1) Präferenzen (0) Statistiken (0) Marketing (0) Nicht klassifiziert (0) Notwendige Cookies helfen dabei, eine Webseite nutzbar zu machen, indem sie Grundfunktionen wie Seitennavigation und Zugriff auf sichere Bereiche der Webseite ermöglichen. Die Webseite kann ohne diese Cookies nicht richtig funktionieren. Name Anbieter Zweck Ablauf Typ CookieConsent Cookiebot Speichert den Zustimmungsstatus des Benutzers für Cookies auf der aktuellen Domäne. 1 Jahr HTTP Präferenz-Cookies ermöglichen einer Webseite sich an Informationen zu erinnern, die die Art beeinflussen, wie sich eine Webseite verhält oder aussieht, wie z. B. Ihre bevorzugte Sprache oder die Region in der Sie sich befinden. Wir nutzen diese Cookie-Typen nicht. Statistik-Cookies helfen Webseiten-Besitzern zu verstehen, wie Besucher mit Webseiten interagieren, indem Informationen anonym gesammelt und gemeldet werden. Wir nutzen diese Cookie-Typen nicht. Marketing-Cookies werden verwendet, um Besuchern auf Webseiten zu folgen. Die Absicht ist, Anzeigen zu zeigen, die relevant und ansprechend für den einzelnen Benutzer sind und daher wertvoller für Publisher und werbetreibende Drittparteien sind. Wir nutzen diese Cookie-Typen nicht. Nicht klassifizierte Cookies sind Cookies, die wir gerade versuchen zu klassifizieren, zusammen mit Anbietern von individuellen Cookies. Wir nutzen diese Cookie-Typen nicht. [#IABV2_LABEL_PURPOSES#] [#IABV2_LABEL_FEATURES#] [#IABV2_LABEL_PARTNERS#] [#IABV2_BODY_PURPOSES#] [#IABV2_BODY_FEATURES#] [#IABV2_BODY_PARTNERS#] Cookies sind kleine Textdateien, die von Webseiten verwendet werden, um die Benutzererfahrung effizienter zu gestalten. Laut Gesetz können wir Cookies auf Ihrem Gerät speichern, wenn diese für den Betrieb dieser Seite unbedingt notwendig sind. Für alle anderen Cookie-Typen benötigen wir Ihre Erlaubnis. Diese Seite verwendet unterschiedliche Cookie-Typen. Einige Cookies werden von Drittparteien platziert, die auf unseren Seiten erscheinen. Sie können Ihre Einwilligung jederzeit von der Cookie-Erklärung auf unserer Website ändern oder widerrufen. Erfahren Sie in unseren Datenschutzhinweisen mehr darüber, wer wir sind, wie Sie uns kontaktieren können und wie wir personenbezogene Daten verarbeiten Die Cookie-Erklärung wurde das letzte Mal am 12/11/22 von Cookiebot aktualisiert

Direkt zum Inhalt 0921 75758 555 info@easy-login.de Wartung & Störung Mein Account Login neuen Account beantragen Account aktivieren Account löschen Handynummer ändern Login-Verfahren ändern QR-Code zurücksetzen Account erweitern Fernwartung Main navigation Home Easy client Digitales Zertifikat Totp Über uns FAQ Main navigation Home Easy client Digitales Zertifikat Totp Über uns FAQ 0921 75758 555 info@easy-login.de Störungsmeldungen Mein Account neuen Account beantragen Account aktivieren löschen Handynummer ändern Login-Verfahren ändern QR-Code zurücksetzen Account erweitern Fernwartung Impressum Nutzungsbedingungen Datenschutz Presse Login mit mobileTAN / TOTP Benutzerkennung Passwort Passwort Benutzerkennung Passwort vergessen Main navigation Home Easy client Digitales Zertifikat Totp Über uns FAQ 0921 75758 555 info@easy-login.de Störungsmeldungen Mein Account neuen Account beantragen Account aktivieren Account löschen Handynummer ändern Login-Verfahren ändern QR-Code zurücksetzen Account erweitern Fernwartung Impressum Nutzungsbedingungen Datenschutz Presse Login mit mobileTAN / TOTP Benutzerkennung Passwort Passwort vergessen Login Login mit mobileTAN / TOTP Benutzerkennung Passwort Passwort vergessen Was ist easy Login? Im folgenden Video zeigen wir Ihnen die Funktionsweise von easy Login und erklären Ihnen, wie Sie das easy Login-Portal nutzen können. Zum Aktivieren des Videos müssen Sie auf den untenstehenden Link klicken. Wir weisen darauf hin, dass nach der Aktivierung Daten an den jeweiligen Anbieter übermittelt werden. Video aktivieren Video aktivieren Teilnehmende Versicherer Folgende Gesellschaften erreichen Sie über easy Login oder über die kostenfreie Webanwendung easy Client . Eine detaillierte Übersicht finden Sie hier: list Detailsansicht öffnen Einfach und sicher in die Extranets easy Login ist eine Initiative für die einfache Anmeldung an den Onlinediensten und BiPRO-Webservices der Versicherer. Mit nur einem Login gelangen Sie in die Portale aller easy Login-Versicherer. Ihre Vorteile auf einen Blick: easy Login ist für unabhängige Vermittler kostenfrei . Sparen Sie Zeit: Mit nur noch einem Zugang für Ihre Versicherer. Passwortlisten gehören für Sie durch einheitliche Zugangsdaten der Vergangenheit an. Post digital abholen? Ist mit nur wenigen Klicks in unserer kostenfreien Webanwendung easy Client möglich. Oder nutzen Sie ein Maklerverwaltungsprogramm ? Kein Problem! Über 25 Anbieter setzen auf easy Login. Mit dem digitalen Zertifikat von easy Login erhalten Sie Dokumente in Ihrem MVP nun noch einfacher. Wir garantieren höchste Sicherheitsstandards durch BSI-Zertifizierung unserer Dienstleister. easy Login in Zahlen 53 Versicherer 20.181 aktivierte Accounts 34 Softwarehersteller Stimmen zu easy Login „Wir sind froh, dass durch die Zusammenarbeit von easy Login und TGIC ein einheitliches Single Sign-On im Markt geschaffen werden konnte. Die sicherheitszertifizierte Technologie der TGIC gepaart mit dem Erfahrungsschatz von easy Login stellt für uns eine vielversprechende und zukunftsorientierte Lösung dar. Durch die Teilnahme bei easy Login möchten wir diesen Weg begleiten und unseren Vertriebspartnern die Möglichkeit geben, die Chancen der Digitalisierung noch stärker für sich zu nutzen.“ Dr. Rolf Wiswesser / Vorstandsmitglied der Allianz Versicherungs-AG „easy Login ist eine Pflicht für den modernen und zukunftsorientierten Versicherungsmakler! Wir nutzen easy Login von Beginn an und sparen so wertvolle interne Ressourcen, die wir unseren Kunden zukommen lassen können und somit nicht in unnötige technische Administration stecken. Als Schnittstelle zwischen MVP-Herstellern, Versicherungsunternehmen, Versicherungsmaklern und weiteren Marktteilnehmern spielt easy Login eine wichtige Schlüsselrolle im aktuellen digitalen Wandlungsprozess der Branche.“ Holm Dömel / Gesellschafter-Geschäftsführer der Beckert & Dömel Versicherungsmakler GmbH „Wir sind davon überzeugt, dass die Verknüpfung von easy Login mit der TGIC-Technologie Maklern einen großen Mehrwert bietet. Sie profitieren von Service-Vorteilen bei gleichzeitig hoher Sicherheit. Der einheitliche Login und die durch BiPRO verschlankten Arbeitsprozesse sorgen dafür, dass ihnen viel mehr Zeit für Kundengespräche bleibt.“ Roland Schäfer / Hauptabteilungsleiter und Geschäftsführer der Stuttgarter Versicherungsgruppe „Dank easy Login konnten wir viele Versicherer innerhalb kürzester Zeit auf eine einheitliche Authentifizierung per Zertifikat umstellen und so die digitalen Prozesse unserer Maklerkunden signifikant verbessern. easy Login hat uns und unsere Kunden hierbei tatkräftig unterstützt und begleitet - vielen Dank dafür!“ Markus Heussen / Geschäftsführer der b-tix GmbH „Bewährtes nicht lassen - zeitgleich Neues nicht verpassen! easy Login gehört dazu und erleichtert den digitalen Umstieg!“ Michael Richthammer / Geschäftsführer der RICHTHAMMER Versicherungsmakler GmbH & Co. KG „Wir glauben, die Arbeitsabläufe unserer Vermittler sollten einfach und gleichzeitig sicher sein. Dabei spielen eine sichere Authentifizierung und BiPRO-Normen eine entscheidende Rolle. Als Gründungsmitglied des Single Sign-On e.V. und des BiPRO e.V. begleiten wir die Initiativen seit dem ersten Tag.“ Dr. Gerrit Böhm / Mitglied der Vorstände der VOLKSWOHL BUND Versicherungen „Mit easy Login können die Nutzer unserer BiPRO-BOX das System noch schneller und unkomplizierter verwenden als bisher. Die zertifikatsbasierte Anmeldung macht einen aufwendigen Anmeldeprozess vollkommen überflüssig. Auf diese Weise lassen sich die Dokumente der angeschlossenen Versicherungsgesellschaften im Handumdrehen abholen. Einfacher und schneller geht's nicht!“ Emanuel Römer / Geschäftsführer der Mr-Money Service GmbH „easy Login leistet einen großen Beitrag, die Vorteile und Mehrwerte der BiPRO-Welt praxistauglich nutzbar zu machen. Das digitale Zertifikat auf easy Login schafft die perfekte Kombination aus Sicherheit und Praxisorientierung und ermöglicht dem Anwender aus Keasy heraus leistungsfähige und wartungsarme Maschine-zu-Maschine Kommunikation mit vielen Versicherern. Ich hoffe, dass sich noch viele weitere Gesellschaften bei easy Login beteiligen - nur so werden die digitalen Angebote auch wirklich genutzt.“ Robert Schmidt / Geschäftsführer der vfm Gruppe (Keasy Maklerverwaltungsprogramm) „Durch die einheitliche Authentifizierung mittels easy Login wird eine sichere und komfortable Anmeldung bei den BiPRO-Schnittstellen der Versicherungsgesellschaften möglich. Das hilft den Maklern, den Herstellern und den Gesellschaften zugleich.“ Sasha Justmann / Geschäftsführer der zeitsprung GmbH & Co. KG „easy Login ist ein wesentlicher Baustein bei der Digitalisierung der Versicherungsbranche. Daher freuen wir uns, dass wir als Softwareanbieter den an unsere Versicherungsplattform SMART INSUR angebundenen Partnern mit easy Login eine sichere, einfache und BiPRO-konforme Authentifizierung anbieten können.“ Marcus Rex / Vorstand der Smart InsurTech AG prev next easy Login-News Hier finden Sie eine Auswahl aktueller News rund um easy Login. 14. Oktober 2022 easy Client: Erklärvideo In diesem Video zeigen wir Ihnen die Funktionsweise des neuen easy Client und erklären Ihnen, wie Sie diesen nutzen können, um Ihren Arbeitsalltag zu erleichtern. Zum Erklärvideo 7. September 2022 Neu: der easy Client als webbasierte Anwendung Ab sofort ist der easy Client für die Abholung Ihrer digitalen Post, den Einstieg in die Versichererportale oder das Rechnen von Angeboten von jedem Endgerät und über jeden gängigen Browser nutzbar. zur Infoseite 22. Juli 2022 TOTP-App jetzt für Android Sie möchten TOTP für den Login auf Ihrem Android-Handy erzeugen? Das geht ab sofort mit der easy Login TOTP-App. Weitere Infos zu nutzbaren Apps finden Sie über den nachfolgenden Button. Infos zu TOTP-Apps 20. Juli 2022 DKM 2022: Wir laden Sie ein Sie haben noch keine Eintrittskarte für die diesjährige DKM? Sichern Sie sich hier Ihr Messticket! Besuchen Sie uns in der Halle 4 am Stand A-13, wir freuen uns auf Sie. Jetzt kostenlos anmelden 31. Mai 2022 Es geht wieder los: Der Jungmakler Award Er ist DER Nachwuchswettbewerb für alle unabhängigen Vermittlerinnen und Vermittler der Finanz- und Versicherungsbranche und oft der "Startpunkt" für eine lange und erfolgreiche berufliche Reise. Jetzt bewerben prev next easy Login GmbH Wittelsbacherring 49 · 95444 Bayreuth Telefon +49 921 75758-555 Telefax +49 921 75758-554 info@easy-login.de Footer menu Impressum Nutzungsbedingungen Datenschutz Presse

Context/Motivation



Shows how much of the readable text is DSGVO related on websites that have some DSGVO content.
Based on the 122 german websites in the dataset that contain DSGVO text.

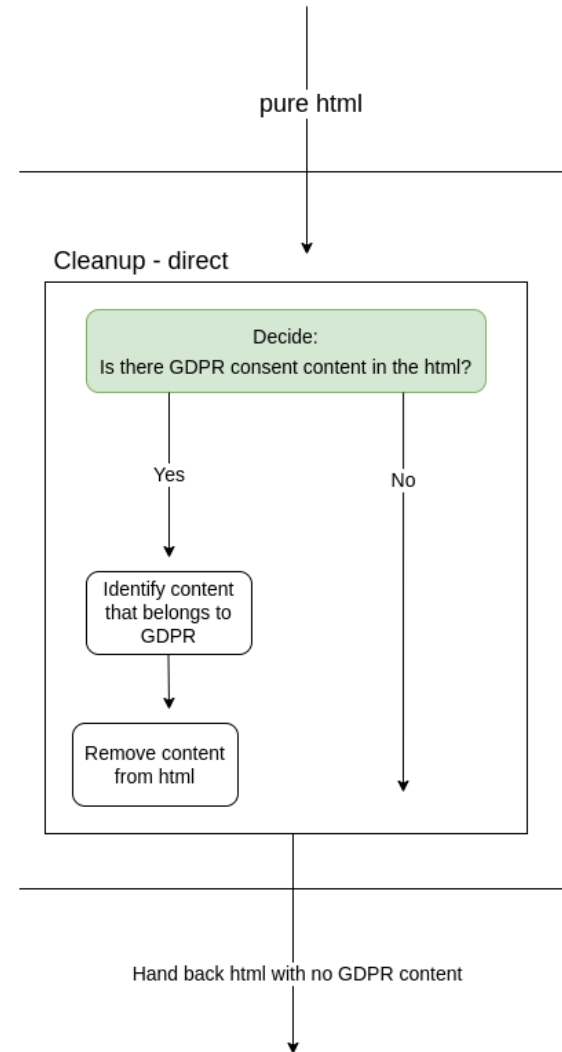
Context/Motivation

Storing and indexing these DSGVO texts leads to

- wasted storage (although not too much overall)
- blurred search results (probably hard to find websites where DSGVO text dominates extracted text)

Goal

1. Curate a GDPR dataset
2. Create a module that cleans HTML from nodes that belong to DSGVO.



Related Work

- Firefox Plugins that block/refuse consent pop-ups:
 - gdpr-not-consent by ciscar
 - Never-Consent by MathRobin

Lot of sites have made their own system, it's complicated to manage them all. But by auto-refusing most of the majors CSP providers we try to do the most we can.

I'm working on a solution for sites wich are not using a platform, if you have any ideas, please, submit it.

Related Work


- Master thesis and paper: "Analyzing Cookies Compliance with the GDPR" and "Automating Cookie Consent and GDPR Violation Detection"
 - Browser extension that blocks cookies based on user preference
 - Dataset built only of websites with consent elements from major CSPs

Dataset Creation

- Start from list of 750 relevant websites
- Manually inspect the first 350 and extract DSGVO HTML
- Parse HTML to plain text

Dataset Creation

Chrome is being controlled by automated test software.

 easy Login
einfach . einmal . einloggen

Was ist easy Login?

easy Login ist eine Initiative für **die einfache Anmeldung an den Onlinediensten und BIPRO-Webservices** der Versicherer. Mit nur einem Login gelangen Sie in die Portale aller easy Login-Versicherer.

Ihre Vorteile auf einen Blick:

- easy Login ist für unabhängige Vermittler **kostenfrei**.
- **Sparen Sie Zeit:** Mit nur noch einem Zugang für Ihre Versicherer.
- Passwortlisten gehören für Sie durch **einheitliche Zugangsdaten** der Vergangenheit an.
- **Post digital** abholen? Ist mit nur wenigen Klicks in unserer kostenfreien Webanwendung [easy Client](#) möglich.

Oder nutzen Sie ein...

Diese Webseite verwendet Cookies

Wir setzen Cookies für verschiedene Zwecke ein (Statistiken & Nutzungsverhalten, Optimierung und Fehlerbehebung). In unseren [Datenschutzhinweisen](#) finden Sie mehr Infos und eine Opt-out-Möglichkeit.

Nur notwendige Cookies verwenden

Cookies zulassen Details zeigen

... age-node-type-page

div#CybotCookiebotDialog

...

Styles

Computed

Layout

Event Listeners

Filter

element.style { display: block; opacity: 1; top: 689px; width: 100%; height: auto; }

#CybotCookiebotDialog

<html>

Selenium stores the total HTML

If present I copy the HTML that belongs to DSGVO and store it

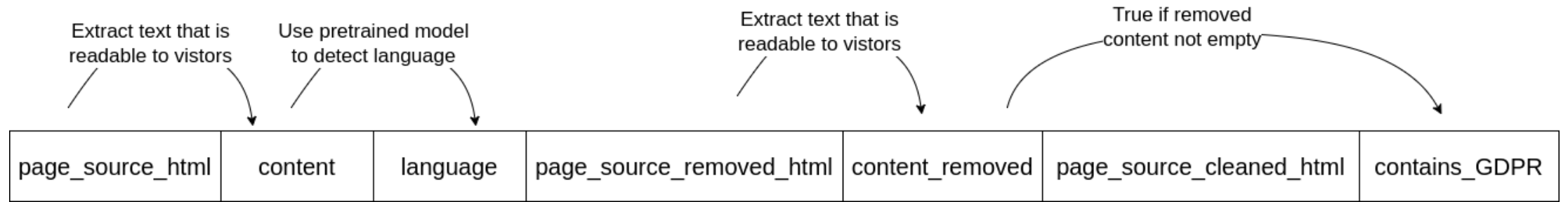
and store the HTML with the DSGVO element removed

page_source_html

page_source_removed_html

page_source_cleaned_html

Dataset Creation



Stats on the resulting dataset:

- Total entries: 350
- German entries: 255
- GDPR entries: 154
- German GDPR entries: 131

Dataset Refinement

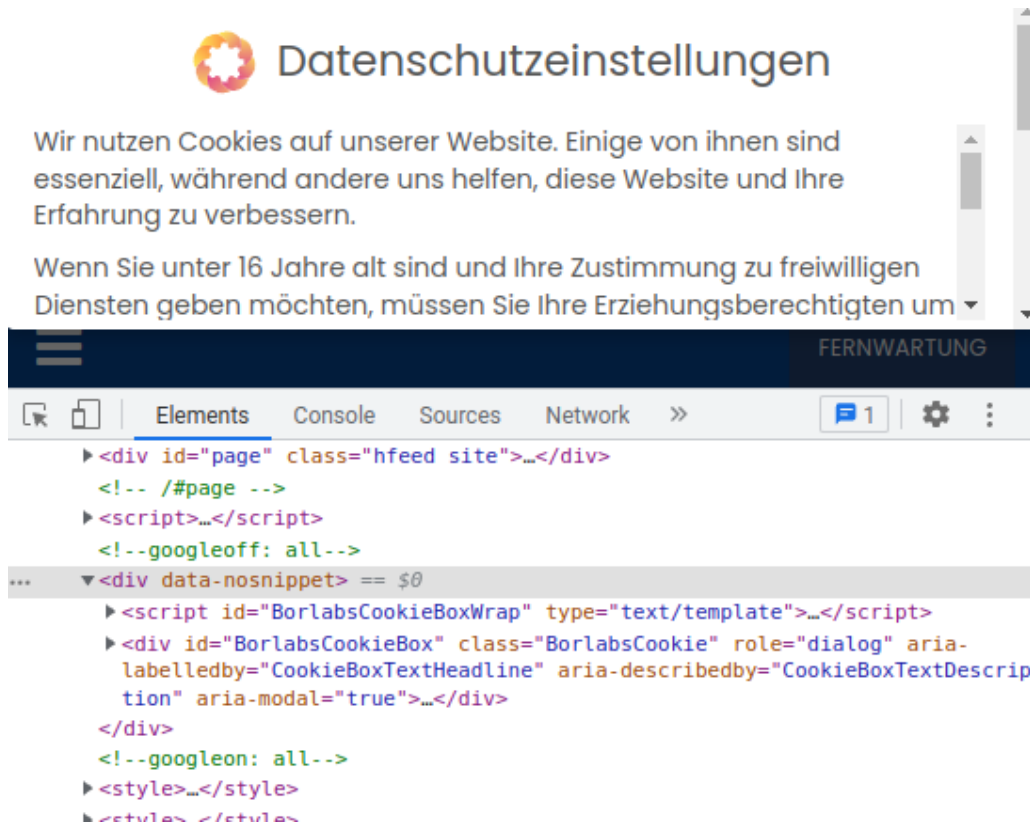
- Only interested in German websites
- Remove websites where Selenium could not fetch page source html
- For some websites I manually stored GDPR html that is not present in downloaded html of website
- Reiteration to detect these by checking that words of extracted GDPR text are findable in page source html and by looking at websites that were classified wrong by the first fitted SVM

Stats on the refined dataset:

- Total entries: 246
- GDPR entries: 122

Dataset Creation Observations

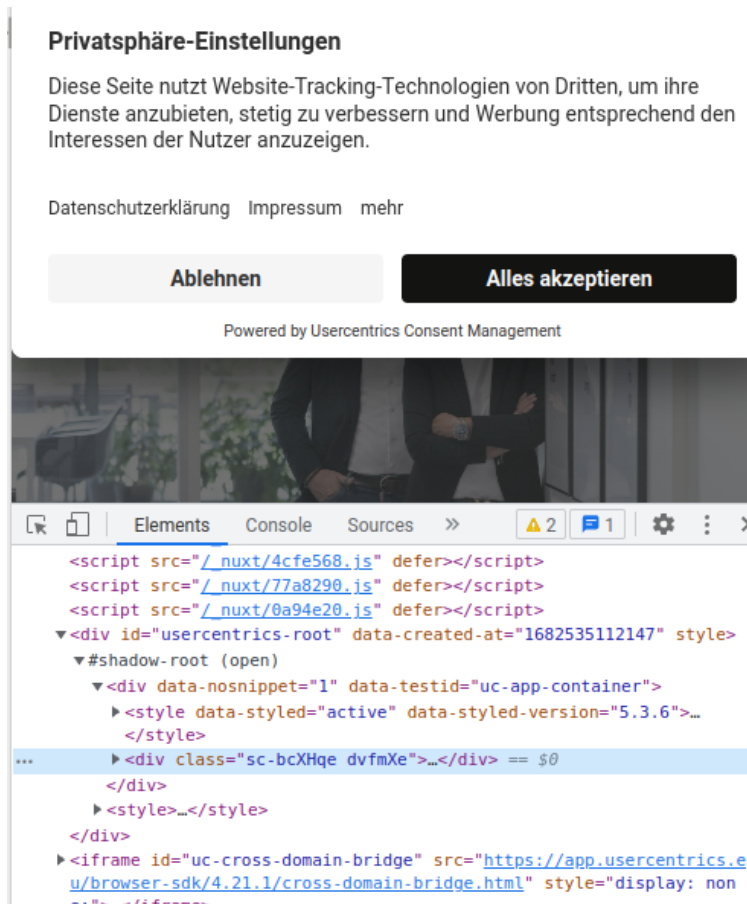
- googleoff, googleon



No official documentation by google, but is said to advice google to not index certain parts of a website.

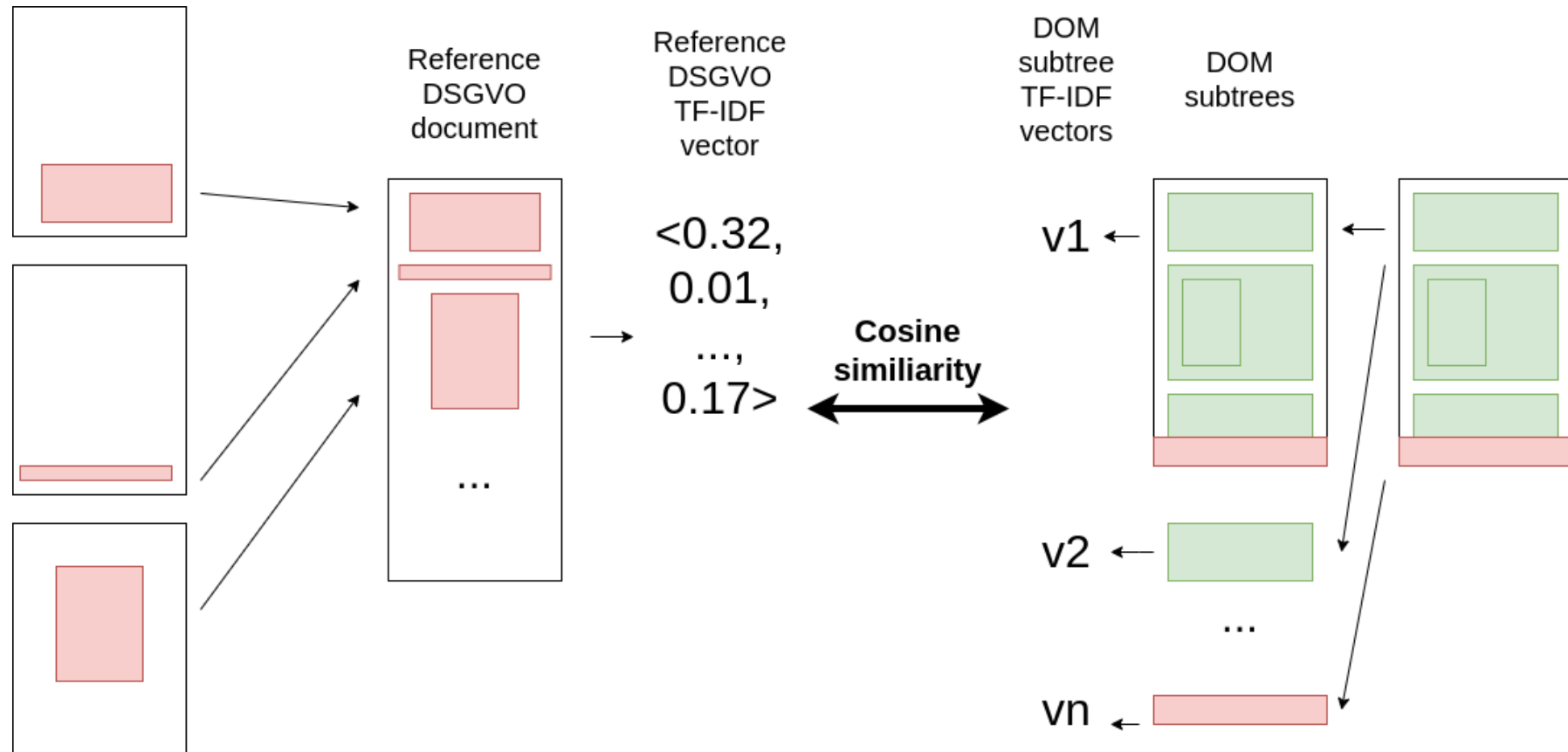
Dataset Creation Observations

- #shadow-root

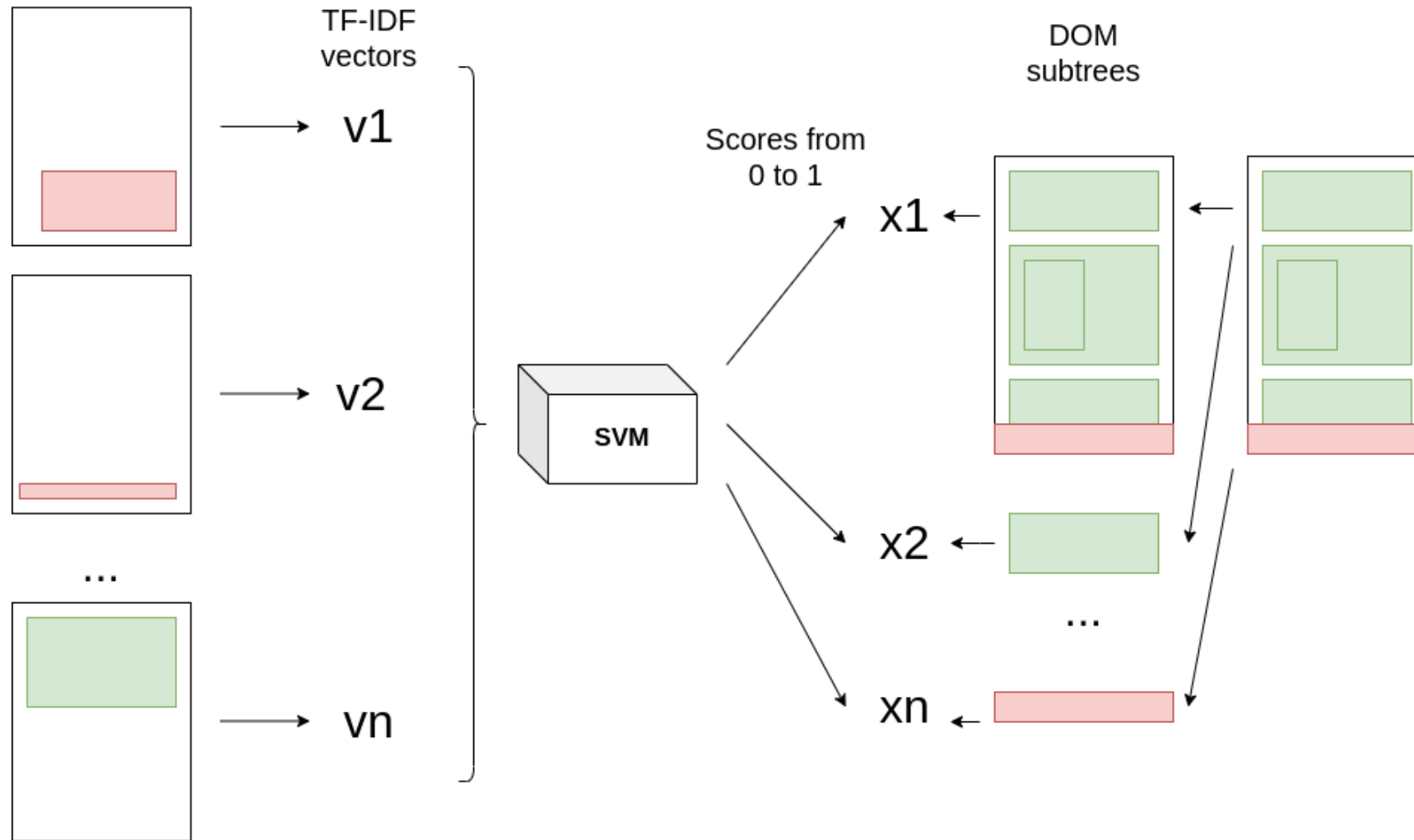


Mit Web Components können Sie im **Shadow-DOM** eigene wiederverwendbare Bausteine entwerfen, deren HTML-Elemente gekapselt sind und deren CSS nur in diesem Gültigkeitsbereich angewandt wird. Diese werden ins normale [DOM](#) der Webseite eingehängt und sind für Debugger und Skripte nicht direkt sicht- und ansprechbar.

TF-IDF + Cosine Similarity Model



TF-IDF + SVM Model



Demo

Next steps/scope of project

- 350/450 hours in
- Java implementation of SVM-based DSGVO cleaner
- Writing of report
- My goal is to finish by 7th of June

Future projects

- googleoff-googleon checker for the crawler
- Accept/Reject button classifier + clicker for Selenium session or browser extension
 - Maybe contribute to Never-Consent
- Crawl/Build large GDPR/DSGVO dataset with help of classifier (for more comprehensive analysis compared to focusing only on CSPs)

- https://wiki.selfhtml.org/wiki/HTML/Web_Components/Shadow_DOM, 03.05.2023
- <https://github.com/MathRobin/Never-Consent>, 03.05.2023
- <https://github.com/csicar/gdpr-not-consent>, 03.05.2023
- <https://github.com/dibollinger/CookieBlock-Violation-Detection>, 03.05.2023