TALLINN UNIVERSITY OF TECHNOLOGY
School of Information Technologies

Md Ehsanur Rahman 194333IASM

Analyzing Web Traps and User's Data Privacy

Master's thesis

Supervisor: Vladimir Viies, PhD
Associate Professor

Tallinn 2021

TALLINNA TEHNIKAÜLIKOOL
Infotehnoloogia teaduskond

Md Ehsanur Rahman 194333IASM

Veebilõksude ja kasutaja andmete privaatsuse analüüsimine

Magistritöö

Juhendaja: Vladimir Viies, PhD
Associate Professor

Tallinn 2021

Author’s declaration of originality

I hereby certify that I am the sole author of this thesis. All the used materials, references to the literature and the work of others have been referred to. This thesis has not been presented for examination anywhere else.

Author: Md Ehsanur Rahman

09.05.2021

Abstract

Nowadays, User's data privacy protection is a prime concern and still a key challenge in the web administration arrangement region. The User and web specialist co-op regularly share their private information with other web administrations will expand the danger of abuse and revelation of security. This work aims to discuss the theoretical basis of web tracking components, manipulation through GUI, Dark patterns, how it is utilized to track and trap a massive number of victims consistently on the web, exhibiting its sorts and giving instances of sites and organizations using it, how to recognize, avoid, what is the proposals of managing them and making users' awareness on the privacy hazard implied with Web tracking. Finally, two applications will be developed. First is the tracking tool “TrackMe” which will work on the windows platform for popular browsers (Google Chrome, Mozilla Firefox, Opera, and Microsoft edge). This tool will detect dark patterns and protect users by warning them in cases where they notice a dark pattern. Second is the web application “darktracker.xyz” which will work as an awareness builder among users about web tracking and manipulation.

This thesis is written in English and is 82 pages long, including 6 chapters, 28 figures and 7 tables.

List of abbreviations and terms

UI	User Interface
JS	JavaScript
LSO	Local Shared Objects
Iframes	Inline Frames
HTML	Hypertext Markup Language
CSS	Cascading Style Sheets
URL	Uniform Resource Locator
HTTP	Hypertext Transfer Protocol
GDPR	General Data Protection Regulation
CCPA	California Consumer Privacy Act
IP	Internet Protocol
GUI	Graphical User Interface
OS	Operating System
CUI	Character User Interface
UX	User Experience
CD	Compact Disc
HIPAA	Health Insurance Portability and Accountability
COPPA	Children's Online Privacy Protection Act
IAPP	The International Association of Privacy Professionals
CSV	Comma-Separated Values
RegEx	Regular Expression
PC	Personal Computer
DOM	Document Object Model
MVT	Model View Template
SQL	Structured Query Language
VPN	Virtual Private Network

Table of contents

List of figures

List of tables

1 Introduction

The open idea of the Internet and the straightforwardness with which content is shared via web-based media has made it simpler for the continuous event of online manipulation without the attention of its users. Most web administrations (e.g., Google, Facebook, Amazon) gather a lot of individual data from our online actions, including the things we search, the sites we visit, individuals we contact, or the items we purchase. The enormous scope assortment and investigation of personal data comprise the center business of numerous online organizations. Even though it is ordinarily accepted that this data is chiefly utilized for focused publicizing, some new works have uncovered that it is misused for some different purposes, including price segregation, personalization of search items, evaluation of monetary believability, assurance of protection inclusion, government reconnaissance, background examining, or identity fraud. When all is said in done, web tracking permits third-party or first-party sites to know the users' browsing history and browsing setup to these closures. These strategies can be utilized to improve users' experience and to upgrade their browsing on the Internet. Web promoting organizations will not recognize web tracking as a danger, and they even openly safeguard web trapping or tracking based on its importance in the Internet economy.

Many websites use deceptive user interfaces to manipulate users. Mostly it is known as Dark patterns. Dark patterns are user interface design choices that benefit an online service by coercing, steering, or deceiving users into making decisions that they might not make if fully informed and capable of selecting alternatives. In the best-case scenario, dark patterns disturb and baffle users. They can delude and misdirect users, e.g., by causing monetary misfortune, fooling users into surrendering immense measures of individual information, or initiating impulsive and addictive conduct in grown-ups and youngsters. Third-party tracking is a type of tracking or trapping performed by different administrations not quite the same as the one unequivocally visited by the user. Third-party trackers are viewed as a genuine data privacy danger, as they can gather and amass colossal measures of individual data from our browsing movement through a wide range of sites.

The expanding use and intricacy of the Web and its related administrations additionally raise worries about losing the user's privacy. Because of this, it is instinctive that current Web innovation presents severe dangers to user's security, either considering the absence of components that gives protection guarantees to users or on account of the changes related to the private data assembled and leaked to outsiders without user's knowledge or assent. Although numerous users are mindful of the protection hazards implied by internet providers, the strategies and advances utilized for tracking them are significantly less known.

Many tech companies or individuals are working hard to develop the latest technology to make our lives smoother and provide security to users' data. But there is not any proper solution which can stop these bad practices forever.

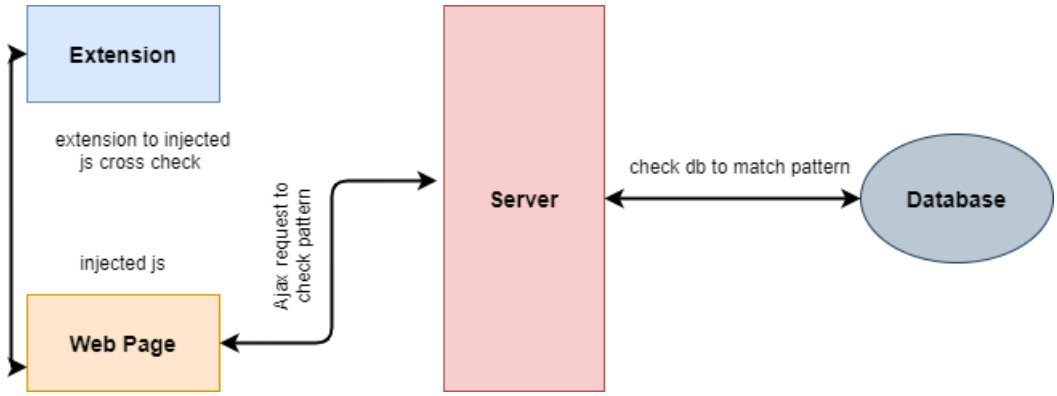

Figure 1. Dark pattern tracker system for popular browsers

Figure 1 shows a simple backend workflow of an extension. An extension will check the web browser through server. There will be one database from where server will check and match pattern. For this project Python and JavaScript will be used mostly. For pattern matching Python RegEx and general text matching logic will also be used.

This audit is the first that surveys numerous web-tracking methods, countermeasures, and legislation to the best of my knowledge.

Aim and Objectives

The objective of this work is to address this issue. This paper mainly focuses on Dark patterns techniques, how it is utilized to track and trap a massive number of victims consistently on the web, exhibiting its sorts and giving instances of sites and organizations using it, how to recognize, avoid, what is the proposals of managing them and making

users' awareness on the privacy hazard implied with Web tracking. This paper also aimed at developing a dark pattern tracker, which will detect or track dark patterns on the website and inform users. Along with this, a website will be designed that will consist of valuable information about dark patterns, the latest information about current tracking technology, graphical representation of dark patterns, and a question asking form to help users to get confirmation about the dark pattern. This way more awareness will be created among web users against those culprits.

The paper is divided into six chapters. The contents of the specific chapters are as follows:

The introduction part describes the background, key features of this work, some basic ideas, and significance of the paper. This part also describes the aim and objective of this work.

The second chapter introduces the base of this paper. It demonstrates what is involved in tracking and trapping users every day without their consent. It describes the current situation about today's users' data privacy. It also includes information type that has been taken by web owners and third-party organizations and the reason behind doing this.

The third chapter elaborates on the process and techniques that organizations use to trap users. This portion provides detailed information about 'Dark Pattern,' its type, and how those patterns trap and manipulates.

The fourth chapter elaborates basic ideas about this work. This part is important because it clarifies the fundamentals of the structure and programming perspective. It is a technical foundation involved in the development of the Dark Tracker system.

The fifth chapter is the final output of this work. It contains the features and workflow of the system. It describes how this whole system functions and runs.

And the last chapter provides the summary of this whole project, limitations, and the future studies.

2 Web Traps and Users Data Privacy

A browser is a product that is utilized to get to the web. A browser allows users to visit sites and do exercises inside them like login, see mixed media, interface starting with one site then onto the next, visit one page from another, print, send and get an email, among numerous different exercises. The most widely recognized browser software titles available are Microsoft Edge, Google's Chrome, Mozilla Firefox, Apple's Safari, and Opera. Browser accessibility relies upon the user PC's working framework (Microsoft Windows, Mac OS, Linux, among others).

When a user types a page address, for example, www.darktracker.xyz, into a browser, that web page ultimately is not put away on a worker prepared to be conveyed. Each page that the user demand is exclusively made because of the user's request.

Sometimes user mix up with web browser and web server where both the terms are different.

Table 1. Web browser vs web server

Web Browser	**Web Server**
Web Browser is software that is utilized to peruse and show pages accessible over the web.	Web server is programming that gives these reports when mentioned by web browsers.
A web browser sends a request to the server end for web-based reports and services.	Web server sees and endorses those requests made by web browsers and sends the report accordingly.
A web browser sends an HTTP Request and gets an HTTP Response.	The web server receives HTTP Request and sends an HTTP Response.
Web browsers store user information in cookies in a local machine.	Web servers give a region to store the site.
Web Browser is installed on the user's machine.	The web server can be installed anyplace; however, it should be on a network or a local computer.
Mozilla Firefox, Google Chrome, Opera, Safari, etc.	Apache Web Server, Apache Tomcat, etc.

Table 1 shows difference between web browser and web server.

During getting to the site, a trade of information happens, and a few traces are left in computing gadgets. This data is stored in cache files, cookies, URL history, typed URL, session information, and searched terms.

Browsers are vital for an active workspace; however, they also act as the data provider to the organizations that use it to make money. This is causing serious concern about users' data privacy these days.

2.1 Importance of User's Data Privacy

Web security is getting to be a developing concern these days for individuals of all ages. Data Privacy is centred around how the information should be gathered, managed, stored, and shared with any third parties, just as consistent with the appropriate security laws (such as General Data Protection Regulation or California Consumer Privacy Act.) Together with Data Security, Data Privacy makes a Data Protection range with protected usable information as an output.

However, Data Privacy is almost the correct taking care of information and the public desire for security, centering around the person as a key figure.

Data privacy also looks at the following issues:

- what type of data acquired
- how data is gathered or stored
- whether or not the information is shared with a third party
- regulatory restrictions, such as GDPR or CCPA

Numerous organizations, for example, Google, Facebook, and Amazon have benefitted abundantly in the "data economy" - gathering user information to augment item or promotion deals. Great work regarding information privacy implies keeping user data secure, not offering it to outsiders without assent, or utilizing information maliciously or carelessly.

Data or information is quite possibly the main resource an organization has. With the ascent of the information economy, organizations discover tremendous worth in gathering, sharing, and utilizing information. Organizations, for example, Google, Facebook, and Amazon, have all assembled realms on the information economy. Straightforwardness in how organizations demand assent, submit to their security strategies and deal with the information they have gathered is fundamental to building trust and responsibility with users and accomplices who anticipate privacy.

The significance of Data Privacy was additionally improved with the presentation of the General Data Protection Regulation.

From the business point of view, securing individual information and emphasizing data privacy can decidedly affect the association.

2.2 Data Privacy vs Data Security

Protecting data and complying with data protection laws, we need both Data Privacy and Data Security. Data privacy and data security are often used interchangeably, but there are distinct differences.

Table 2. Data privacy vs data security

Data Privacy	**Data Security**
Data Privacy centers around the privileges of a user, the motivation behind information processing and collection, privacy inclinations, and how organizations oversee the individual data of data subjects.	Data Security incorporates many principles and various safeguards and measures that an association is taking to keep any outsider from unapproved access to digital data or any intentional or unintentional deletion, modification, or exposure of information.
It centers around how to gather, measure, offer, document, and erase the data as per the law.	It centers around the insurance of data from malicious assaults and prevents the abuse of stolen data.

Table 2 shows the difference between data privacy and data security. It shows data privacy is the process that mainly focuses on how data can be collected and users' rights. In contrast, data security focuses on safeguard that prevents the third party from unauthorized access.

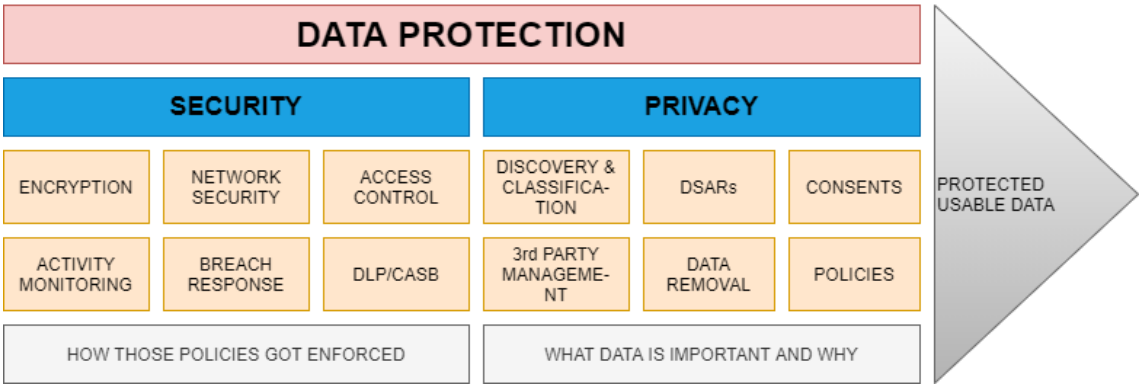

Figure 2. Visual representation of data security and data privacy areas

Figure 2 shows that along with data security, data privacy is a part of the data protection area that help getting protected usable data as an output.

2.3 Tracking Data through Web

While browsing the web, most of us have been tracked, and are not aware of this reality. We all have the experience of searching for an item on the web and afterward having promotions for a similar item chase after us for quite a long time. What might be less clear is what information sites gather, why they do as such, and how sites track us.

2.3.1 Definition of Website Tracking

A few groups utilize the terms website tracking and web tracking conversely, yet they are not by and large the equivalent. Rather than how we have recently characterized web tracking, website tracking is the act of observing how a website changes over the long run.

Web tracking is collecting and sharing data about a person's movement on the web — what they do on the web and how they approach doing it. Web tracking gives organizations a full comprehension of users' inclinations, permitting them to customize their content likewise. The practice is utilized to control numerous online administrations, including computerized promoting and website analytics. Site tracking is incomprehensibly normal; a 2017 study tracked down that 79% of sites use trackers that gather user information.

Organizations utilize a range of software tools, like website trackers and other tracking methods, to see how users collaborate with their sites — a cycle is known as website visitor tracking — and to pursue users around the web to perceive what else users are doing.

Here are a few instances of how sites track users outside of promoting.

- When users look for hotels on Google, it lists hotels in the users' neighbourhood because the web search tool knows where users are based.
- When an e-commerce store shows users a list of suggested items, it understands what users like since it has followed things users took a gander at or purchased beforehand.

This would not happen without website tracking technology.

2.3.2 First-Party vs Third-Party Tracking

One of the principal issues regarding web tracking is in the distinction between first-party and third-party web tracking.

Table 3. First-party tracking vs third-party tracking

First-Party Tracking	**Third-Party Tracking**
When the user visits one website, and first-party tracking is done by that website.	Third-party tracking is done by the site other than the one user are currently browsing.
For example, when user visit e-commerce site, that tracks pages user has visited and the things they looked at.	For example, when users notice ads of other companies such as Facebook buttons on the site they are currently browsing or visit one news site that also loads various third-party cookies that track users' behavior both on that site and on other sites that user visit in the future as well.
To know user preferences, language settings, location, and other information that the user shares.	To help advertisers, marketers, and other organizations promote their business and to make money.
Usually, there is nothing to worry about third-party tracking until they sell users' data to other organizations.	This can be risky because the user does not know anything about this and what data has been taken.

Table 3 shows difference between first-party and third-party tracking.

2.3.3 Information Collected by Websites

Websites gather a huge array of information for a wide range of uses. This incorporates data users give through forms, for instance, email address and credit card data, just as numerous different sorts of data are acquired from tracking innovation. Website collects:

- Browsing movement across various websites. This gives those with access to the data understanding about the individual user's shopping habits, issues they are confronting, interests, and more.
- Data about how the user associates with sites. For instance, what they look for, click on, and how long they visit a page.
- Information about the device that the user used to access websites and a browser that the user browses.
- IP addresses to determine a user's location.

Not every website gathers all the above information. Some do not collect any information whatsoever. It will all rely upon the service the site is giving and how the site is adapted.

It is additionally worth recalling that sites are not the only way that organizations gather information about users. Organizations additionally collect data from cell phone applications, emails, and smart speakers.

2.3.4 Reasons Behind Tracking User’s Data

In the present computerized age, user's information is amazingly significant. Organizations need to find out as much about users as possible. This information permits them to tailor their applications to user's inclinations, and it will empower promoters to target users with messages users are bound to draw in with. The primary reasons sites track users are to acquire bits of knowledge about how their users utilize their website, give a customized online encounter, and adapt the user by showing them focused on adverts.

Website Performance: Numerous site functions will not work without tracking. For instance, sites track users to keep them signed into their site as they peruse various pages, and e-commerce suppliers screen users to save items in the shopping cart.

Different functions would not function too without tracking. Consider when a user watches recording on YouTube, and the algorithm suggests recordings to see straightaway. This element works because YouTube understands what users have watched previously and utilized this data to show different clips users may discover fascinating.

Advertising: At the point when sites guarantee to "improve user experience" with cookies, they are normally discussing focused on advertisements. Third-party promoting cookies permit sites to show user advertisements that are projected to coordinate with user interests.

These promotions are depicted as "relevant," "personalized," or "targeted" because the advertisement provider is utilizing user information to show user advertisements that are bound to appeal to users. That is why users will regularly see advertisements that straightforwardly identify with users' new web searches or to recordings users have recently watched on YouTube. Publicists trust that users will be more able to tap on advertisements that line up with user-perceived inclinations. Also, when users click on their promotions, advertisement providers get cash.

Website Analytics: Web investigation permits site administrators to figure out how user utilize their website, what users like, notice, and dislike, where their visitors are coming from, and who their visitors are.

This data can help the site proprietors settle on business choices and advance the site dependent on how visitors use it.

For instance, an e-shop owner may see that numerous visitors possibly scroll to one product type when visiting the site. If so, they could change how they utilize interlinking between categories to make other products more recognizable. This could expand the quantity of pages users’ access.

Usability Testing: Usability testing is an interaction where genuine users evaluate an item, for example, an application or site, to survey its qualities and shortcomings. In the

common ease of use testing exercise, the user will be approached to explore a site or complete a couple of undertakings on an application. Because of user input — challenges user encountered, things that confused user, or highlights that appeared well and good — the result can be changed and improved.

Specialists use web tracking technology to lead remote usability testing as a group, acquiring huge loads of important information that they can use to tailor their items toward their real visitors' inclinations.

2.4 Data Privacy Statistics

Over the year data privacy has been a major concern all over the world. Many of us do not know how organizations are collecting our personal data, how they are handling those data, and who should be responsible for this. This facts and figures are still unknown for many.

Table 4. Data privacy statistics

Research	**Data**
Pew Research Center	79% of respondents said they are really or to some degree worried about how organizations are utilizing the information they gather about them. In correlation, 64% say they have a similar degree of worry about government information collection [18].
Salesforce research	46% of clients feel they have let completely go over their information [19].
Cisco Consumer Privacy Survey 2019	24% of respondents track down the individual user responsible for ensuring information security [20].
Pew Research Center	81% of respondents feel they have next to zero power over the information gathered [18].
Cisco Consumer Privacy Survey 2019	43% of everything respondents do not accept that they can adequately protect their information today [20].
U.N.	18% of nations have no information protection law executed [21].

IAPP	58% of European organizations announced GDPR consistency as a first concern, while just 11% of U.S. respondents chose it as number one [22].
Egress survey	93% of US IT leaders said they had, in any event, found a way a few ways to conform to protection guidelines like CCPA or the E. U's. GDPR [23].
eMarketer	35% of U.S. organizations studied said that they would not be CCPA agreeable by January 1, 2020, because they feel it is too costly to consider achieving consistency [24].

Table 4 shows a short list of facts about data privacy over the year and organizations name that worked on making this survey.

2.5 Global Privacy Regulations for Web Tracking

Privacy-awareness will be quite possibly the main turn of events, squeezing the governments to execute information protection laws, directing how organizations will deal with user's information, and what esteems should include endeavouring on the market.

Luckily, officials have identified the significance of having information protection guidelines and the need to consider organizations liable for end-user information.

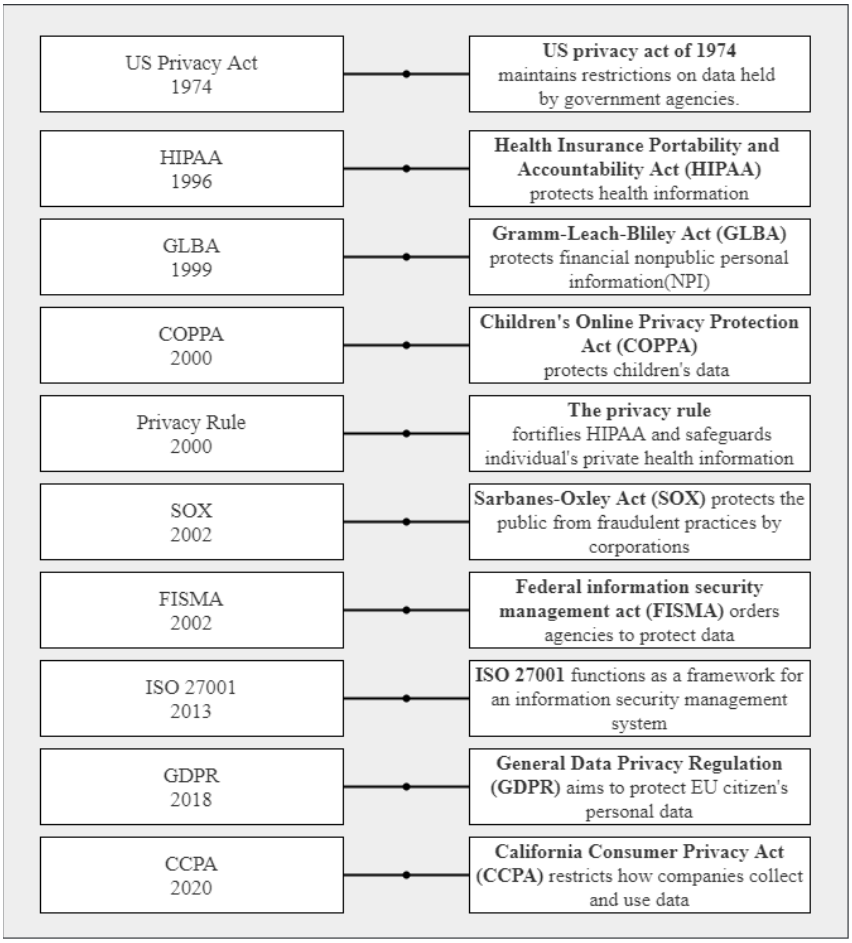

Figure 3. Data privacy law and act timeline

Figure 3 shows the data privacy law and the timeline that was introduced. From the figure it shows there are so many data privacy laws such as US Health Insurance Portability and Accountability (HIPAA), the Children's Online Privacy Protection Act (COPPA), General Data Privacy Regulation (GDPR), or California Consumer Privacy Act (CCPA) have been built to provide user's data privacy. But most of them are now outdated. Among them, GDPR and CCPA are the most ground-breaking and comprehensive data privacy laws.

2.5.1 Website Tracking and GDPR

General Data Protection Regulation denoted the main genuine aim to properly control the extreme abuse of individual information and fine both information processors and information regulators.

GDPR has enabled information subjects to recover command over their privacy.

Before collecting users' data, businesses should receive opt-in consent from consumers. GDPR characterizes individual information as any snippet of data that identifies with a recognizable individual. This incorporates online identifiers, for example, cookies, IP addresses, and digital fingerprints.

This means that websites must receive permission from the user while using this technology to track users. But they can track users using the technology if it is necessary for websites to function, for instance, cookies used to keep users logged in while they browse websites.

Also, sites cannot utilize pre-marked boxes or banners that say something like, "By utilizing our website, you consent to the utilization of treats." Instead, sites should make clients find a positive way to collect information, for example, checking a formerly unticked box.

There must be a way to withdraw users' consent and delete their previously stored information if they have provided consent before.

In Europe, because of GDPR, companies started to use cookie banners for the last couple of years, and that banner consists of information about why this cookie is used for.

2.5.2 Website Tracking and CCPA

After introducing the GDPR, in 2020, the US Congress has introduced similar laws: California Consumer Privacy Act (CCPA). The CCPA is the first US privacy law of a similar magnitude (considering California is the fifth-largest economy in the world).

CCPA became fully effective, created new obligations for businesses in California, and empowered Californians to control their data.

CCPA does not expect sites to get opt-in consent from consumers to gather their information (except if the consumer is under 16). In any case, sites that collect data should illuminate the consumer about the classes regarding the information it will gather and why it gathers this data at the mark of collection.

This regulation gives consumers the option to get to the data the business has gathered and the option to have their data erased. At last, organizations that offer individual

information to third parties should permit the consumer to quit the offer of this information.

This implies that sites can track customers or consumers; however, they should be upfront regarding why they do as such. They should likewise have an approach to sort out the information they gather because of this tracking to erase it if they receive a request. Sites can utilize a cookie banner to help them stay agreeable with CCPA.

Table 5. GDPR vs CCPA

	GDPR	**CCPA**
Who It Applies To	It applies to any organization that processes personal data of European citizens and residents, even if the organization is outside of European Union.	It applies to profit entities that process data of residents of California
Basis for Consent	Opt in	Opt out
Penalties	4% of annual turnover or 20 million EUR, whichever is greater	$2,500 per record each unintentional violation
Enforcement	From May 25th 2018	January 1st 2020

Table 5 shows the difference between GDPR and CCPA where it describes the key points of these laws.

3 Methods Used for Tracking and Manipulation

The web of today is altogether different. Any time users go on the web, somebody is following them. This is exactly how the web functions, particularly in our advanced information-driven society. The site that users visit every day can record and store a wide range of data about them. Since we realize these websites track everything users might do, an inquiry emerges: how do these websites follow users around the web?

3.1 Methods Used by Websites

There are numerous techniques that are utilized to track users. A portion of these strategies is hard to counter. We should probably investigate the most broadly utilized web trackers and website tracking systems.

Cookies: Cookies are little files that sites send to user gadget that the websites at that point use to recall certain data about users and monitor them — like what is in users shopping basket on a web-based business webpage or login data. These pop-up cookie sees everywhere on the web are benevolent and expected to advance straightforwardness about user's online protection [1]. However, recent research has shown that the vast greater part of the present web cookies is utilized to track users and gather their information with the expectation of serving them focused on advertisements.

When a site has dropped a cookie on a user's PC, the cookie supplier can keep getting to it. This is how sites can utilize cookies to follow users from one website to another website or from one webpage to another webpage.

How long a cookie can follow a user relies upon the sort of cookie.

Session cookies are just stored on the user's PC for the length of their session; they vanish when the user shuts the browser. A first-party persistent cookie stores users more extended term inclinations, for example, the time region user live in and login certifications. As a rule, first-party cookies are essential for site usefulness, or if nothing else, supportive in improving user experience on the site. These cookies can be set to remain on a PC for quite a long time into the future except if they erase them.

Third-party cookies are made by parties other than the site users are browsing. These are the tracking cookies that track users as users browse, starting with one site then onto the next. Website analytics and publicizing are two of the essential uses for tracking cookies — indeed, one recent investigation tracked down that 99% of all cookies are tracking and advertisement cookies [2].

Web Beacons and Tracking Pixels: A web beacon point is a little label set on a site or an email to follow how the user associates with the substance. The beacon is generally a reasonable 1-pixel by 1-pixel clear picture set inside the website page's code.

At the point when the browser lands on a webpage with a web beacon, it requests to download the picture. The requests will contain details that can track the user, including time, data about the browser, or the IP address.

This permits web administrators to track users as they explore a site. On the other hand, when utilized in email advertising, they can give the organization that sent the email data, including when the user opens the email and how often a user opens the email.

Browser fingerprinting: Browser fingerprinting is a moderately new procedure that permits sites to recognize exceptional visitors by means of their internet browsers. At the point when the user interfaces with a site, the user's browser transfers much information that the site uses to streamline the user's experience. This information can incorporate the user's operating system (OS), history, time region, gadget model, screen resolution, and a wide range of other data.

Since the chances are little that any two users will have the very same browser information, this arrangement of parameters turns into the user's "browser fingerprint." Once a site gathers browser unique fingerprint, it can unquestionably expect that any future associations with a similar fingerprint as the user are coming from that user. Browser fingerprinting allows sites to track user's behaviour each time you visit.

Canvas fingerprinting: HTML5 — the most recent form of the HTML coding language that engineers use to construct sites — incorporates a "canvas" component. Initially utilized for animations on a website, it has currently been co-picked into an amazing fingerprinting tool.

Canvas fingerprinting results from sites figuring out how a browser reacts to graphical guidelines. At the point when users visit a site that uses canvas fingerprinting, that site teaches the browser to draw a secret picture. The browser delivers the picture somewhat not the same as another person would, giving data about the user's one-of-a-kind computerized unique mark because of varieties in every individual's gadget equipment, settings, and graphics card. At the point when combined with other tracking data, canvas fingerprinting is profoundly precise.

3.2 Manipulation Using GUI

GUI is a Graphical Interface that is a visual portrayal of correspondence introduced to the user for simple collaboration with the machine [3]. GUI implies Graphical User Interface. It is the regular UI that incorporates Graphical portrayal like icons and buttons, and communication can be performed by cooperating with these icons instead of standard command-based or text-based communication.

It eases the use of electronic devices and makes users' life smoother. Some organizations are using this to manipulate or trick users. They are taking benefits by using something called Dark Patterns. In the rest of this chapter, Dark Patterns will be illustrated broadly: how it is used to manipulate and how it can be detected and avoided.

3.2.1 Dark Patterns

A dark pattern is a term first coined by Harry Brignull in 2010 [4]. They are a feature of UI plan in which configuration is created to nudge users towards an activity they presumably would not have taken something else, towards activities that advantage the ones doing the nudging. Dark patterns existed in the actual world well before the internet came along: '90s children will recollect the mail-request music club Columbia House's incredible deal to purchase 12 CDs for only one penny (including handling and shipping), which at that point naturally selected them into a CD-a-month club that was practically difficult to cancel.

Yet, the web has made dark patterns a lot more powerful. Sites can refine their strategies utilizing the detailed feedback their users give, streamlining their control or manipulation at a scale that the actual world would never in its most extravagant fantasies achieve.

Brignull (2010) further clarifies that when we consider "bad design," we think about the maker being lazy or messy, however, without bad expectations. Dark patterns, then again, are not mistakes. They are deliberately created with a strong comprehension of human psychology, and users' interests are not the prime concern.

As indicated by Brownlee (Fast Co. Plan), a dark pattern is a deceptive or, in any case beguiling UI/UX choice that attempts to abuse human psychology to get users to do things they would truly prefer not to do [5].

3.2.2 Types of Dark Patterns

Brignull and Darlo have done tremendous work in this matter. Their work opens a new concern about Dark patterns. But their categories need an update.

In this paper, I have utilized the classifications recorded in the scholarly writing Dark Patterns at Scale: Findings from a Crawl of 11K Shopping Websites [6]. As the title of the paper recommends, analysts examined 53,000 pages on 11,000 shopping sites. At that point, they utilized this information to distinguish fifteen sorts of dark patterns, which were coordinated into seven categories.

Sneaking: Sneaking happens when a site adds additional expenses and charges without telling the client. Generally, the site retains these extra expenses until the latest moment possible, in the expectations that the client will not see.

Sneak into basket: As users buy something, the site sneaks an extra thing into the user's shopping basket. This is accomplished utilizing opt-in/opt-out choices that are checked of course. Now and then, when taking the thing off the user's cart, the user needs to go to a particular page to eliminate that thing.

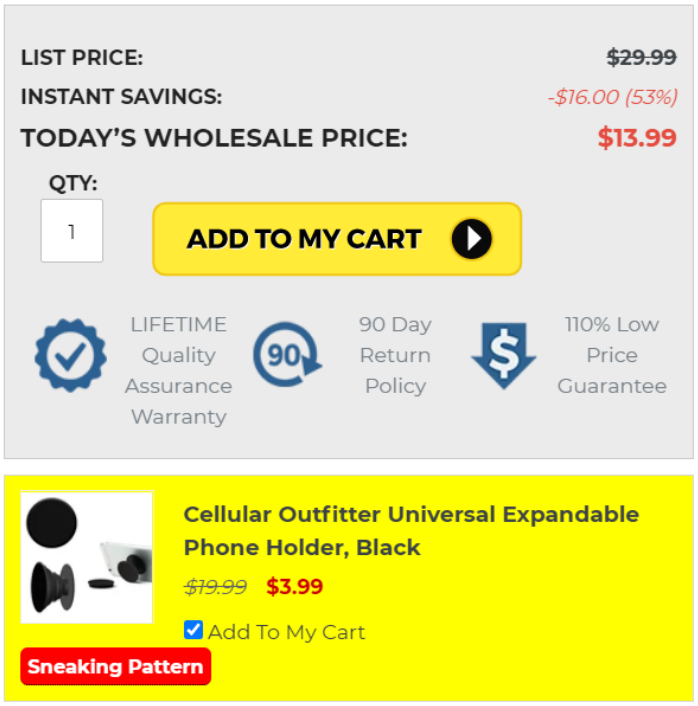

Figure 4. Sneak into basket

In this model in figure 4, the site has sneaked an extra Cellular Outfitter Universal Expandable Phone Holder, Black that costs $3.99. All together for the user to eliminate the thing, the user needs to go to his cart page, just to eliminate the thing.

Hidden Costs: This happens when the user goes through various steps in a checkout stage and, when he gets as far as possible, he finds new and unforeseen expenses were added to the total sum (care and handling cost, fees, and conveyance costs. Charges are not in every case awful; however, they become a more pressing issue when they are uncovered past the point of no return.

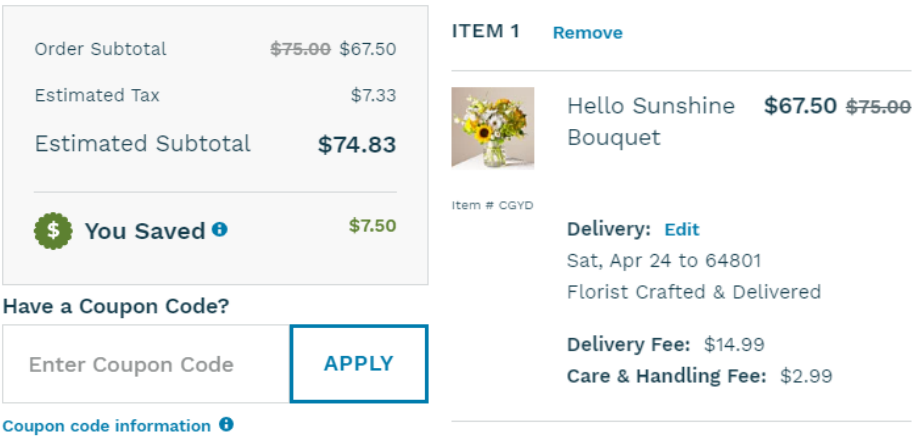

Figure 5. Hidden Cost

While many E-retailers are attempting to be straightforward about it, I, as of late, experienced a current example which is shown in figure 5. In this example, the user got going by seeing that his total due is $ 67.50; however, when he arrived at the last look, it turned out that there are hidden charges, for example, the conveyance charge and care-handling with an expense that cost $14.99 and $2.99 individually.

Hidden Subscription: Numerous applications, particularly mobile games, will rope users in with a "free" name prior to hitting users' potential gain the head with a secret membership. While Google play uncovers this data, it is not in every case enough to prevent small children from spending a big amount of money for a parent.

However, covered-up memberships barely stop there. More underhanded sites will mask a membership as a one-time buy. That way, the customer does not understand that they have bought in until it is past the point of no return.

Fake Urgency: Urgency is not characteristically terrible; however, it tends to be incredibly deceptive when misused. It turns into an issue when sites make fake desperation to deceive individuals.

'Urgency' indicates the group of dark patterns that force a deadline on a deal or sale, in this way speeding up user buys and decision making. Urgency dark patterns misuse the shortage inclination in users—making discounts and offer more attractive than they would some way or another be and flagging that inaction would bring about missing out on likely potential saving. These dark patterns make an intense 'dread of passing up a great opportunity' impact, especially when joined with the Social Proof and Scarcity dark patterns.

We noticed two kinds of the Urgency dark pattern: Countdown Timers and Limited-time Messages on different sites across their cart, checkout pages, and item [6].

Fake Countdown Timer: The 'Countdown Timer' dark pattern is a powerful indicator of a time limit, counting down until the deadline terminates. Countdown timers can be distressing; however, counterfeit countdown timers are tricky.

In many cases, for example, the one beneath countdown timers adds urgency to a deal.

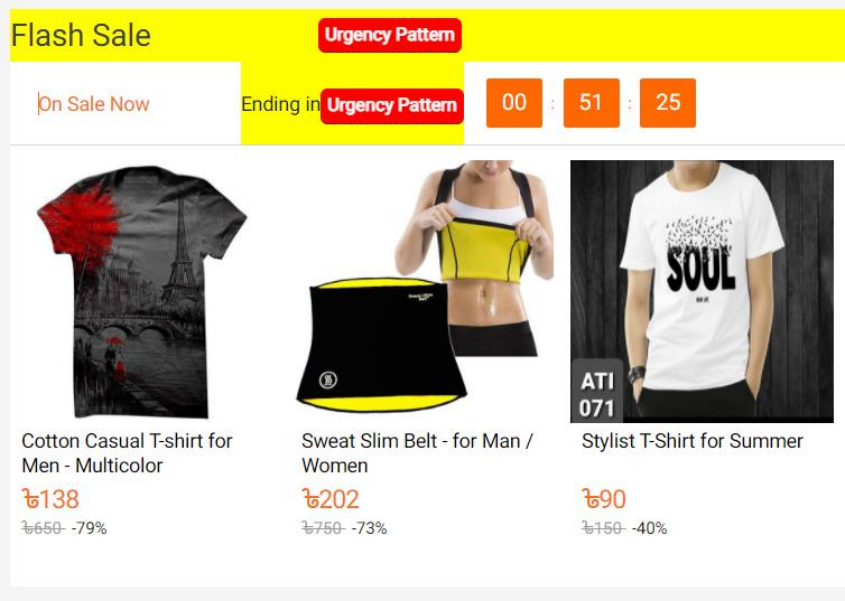

Figure 6. Fake countdown timer

Some measurement only depicts some eCommerce sites utilize counterfeit countdown timers. That implies that a surprising 1.2% of eCommerce sites either reset their clocks or proceed with the deal after it hits zero. Figure 6 shows one of the website name daraz.com.bd using this flash sale count down timer.

Limited-time Messages: Dissimilar to Countdown Timers, the 'Limited-time Message' dark pattern is a static urgency message without an end time. By not expressing the end time, sites retain data from users and consequently distort the idea of the offer.

Deals with an uncertain deadline are tricky because they create a misguided feeling of urgency. Notwithstanding, that does not mean you need to remove all limited time offers.

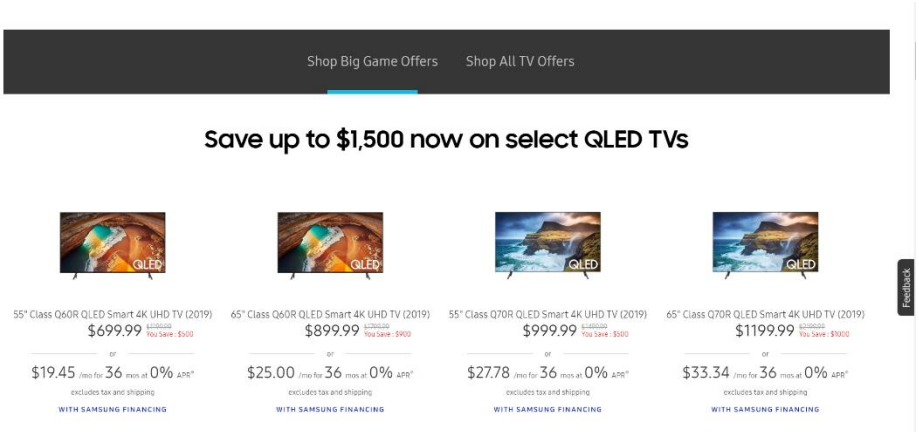

Figure 7. Limited time message

Previously Samsung used to follow these rules. They offer the product on sale but without a deadline. As it is shows in figure 7.

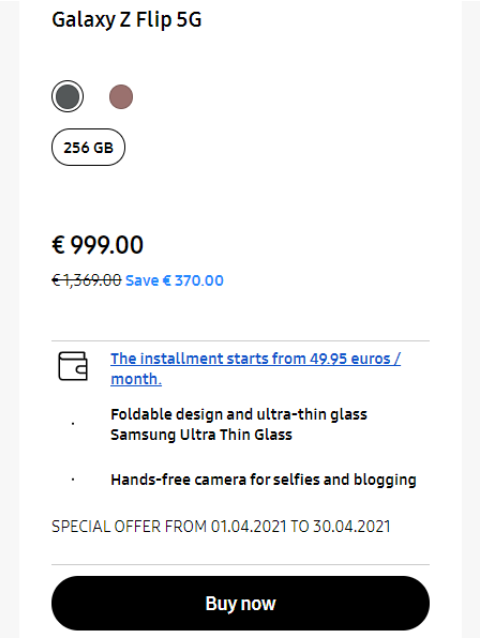

Figure 8. Limited time message (Samsung)

But they have changed the strategy. Samsung now eliminates the dark pattern by advising users precisely when the deal will end. Figure 8 shows Samsung now using the starting and ending time to make users clear.

Misdirection: Misdirection is the most subtle kind of dark pattern since it exists in a grey area. Organizations utilize such a strategy to redirect the consideration of the user from something significant, confounding that person with indistinct decisions or advancing different brands of a similar organization.

Misdirection is made when the user's consideration is guided to a particular spot not to see something different. (Brignull, 2010).

Confirmshaming: Users would regularly see it in websites wherein a little page pops up on users' display, requesting signing up for their services. Yet, rather than having "no, thank you" as a no alternative, the choice to decline ends up being phrased to disgrace the user. It utilizes interesting, enthusiastic language to cause the user to feel awful about not settling on a decision.

By focusing on the users' feelings, organizations increment their odds of poking users into acting not to the user's advantage but rather as per the targets of the organization.

Confirmshaming is a very common practice in email lists and popup notices. But it is not limited to these sources anymore.

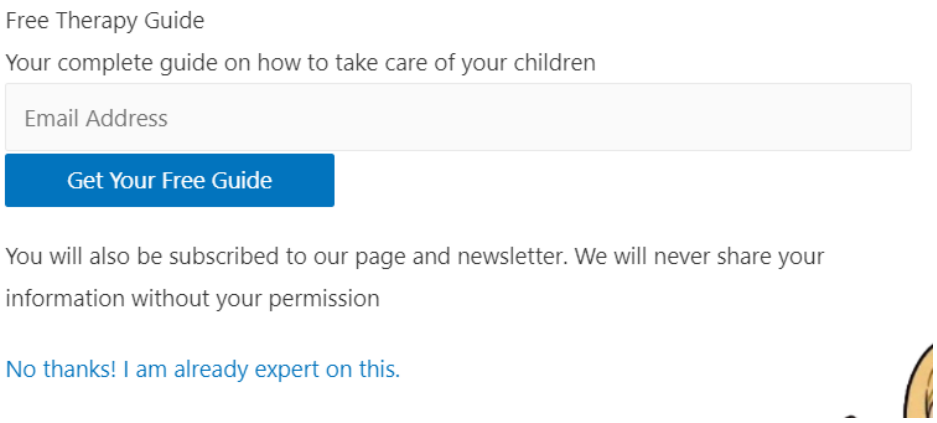

Figure 9. Confirmshaming

In the figure 9 above, rather than having a straightforward no as a choice, they are disgracing the user by the words that are on the no choice. This can cause a few users to feel blame or disgrace simultaneously.

Visual Interference: A decent designer will ensure significant components stick out. A malicious designer will make significant yet non-ideal components harder to see.

"Visual interference" depicts the utilization of visuals to deceive the user. This classification is huge, however, normal outside of eCommerce sites.

Visual interference is not restricted to emails. But emails utilize visual interference constantly. Rather than putting the unsubscribe button or link someplace noticeable, they conceal it away in the expectation that the user will not notice it.

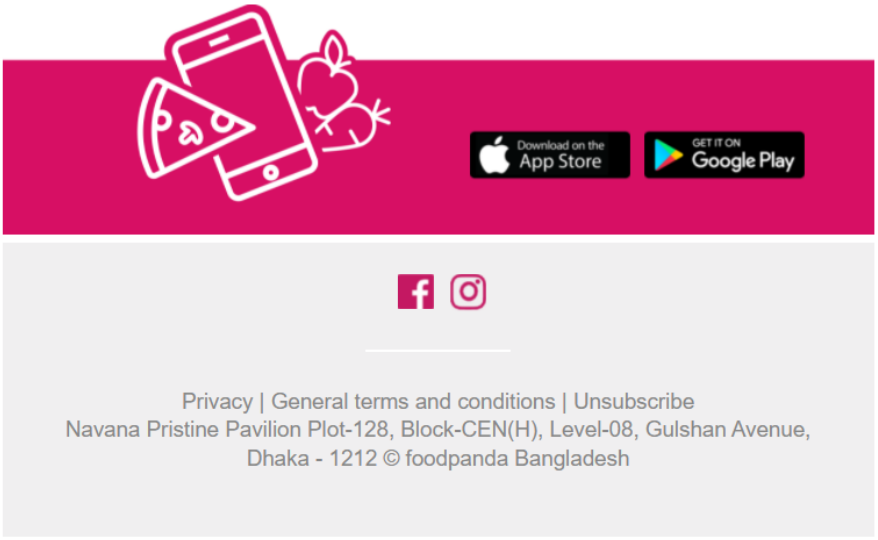

Figure 10. Visual interference

The figure 10 above foodpanda uses unsubscribe links somewhere in their email, expecting user will not notice which can be considered as visual interference.

Trick Questions: Indeed, even forms are not protected from dark patterns. They are particularly helpless against manipulative design. Advertising emails utilize this strategy constantly. After the user register to get to something on the web, he is asked whether he needs to be set on a mailing list. This specific methodology is genuinely standard yet is not tremendously successful because users need to make an express move to select in. Chances are they will be in a rush, and an extent of users will not see this content.

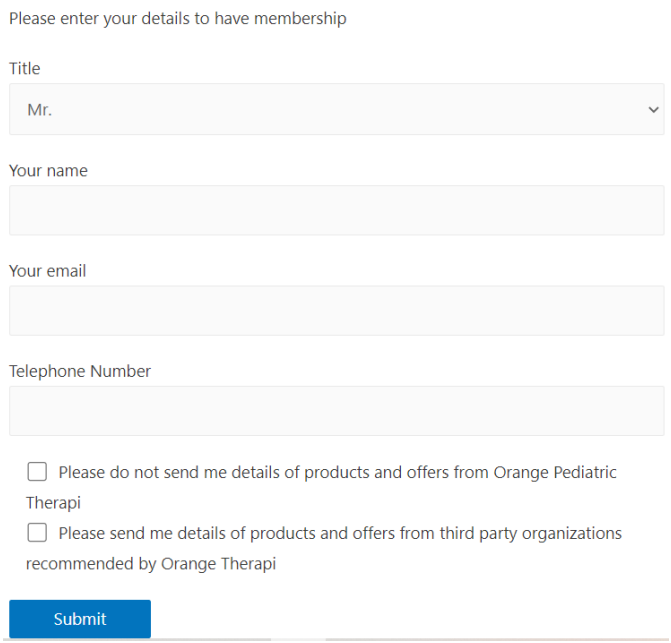

Figure 11. Trick questions

Figure 11 shows one website using this pattern. Looking at one box selects of the newsletter while checking the other picks in. Except if the user notices cautiously, they may check some unacceptable boxes. The user does not need messages, promotion letters, or any emails related to that product, yet they may get them due to how the content is composed either directly or by a third party.

Pressured Selling: The 'Pressured Selling' dark pattern alludes to defaults or frequently high-pressure strategies that steer users into buying a more costly form of an item or buying related items [6]. The Pressured Selling dark pattern misuses a wide range of psychological inclinations, for example, the anchoring effect, the default impact, the anchoring impact, and the shortage bias to drive user buying attitude.

False Scarcity: Many organizations are increasing their products' perceived value and desirability by showing availability as high or limited. This is now a common practice, and it considers as Scarcity dark patterns. Scarcity or shortage, like urgency, is not in every case awful. Nonetheless, it turns into an issue when sites lie about how scarce a thing truly is.

Low Stock Message: The 'Low-stock Message' dark pattern signs to users about limited amounts of an item.

Some of the time, organizations come up short on a thing that is exactly how business functions. In any case, it is somewhat questionable when their whole stock is running low, especially if it is not the special occasion in the calendar.

It is difficult to say which one is a deceptive message and which one is genuine. Possibly the organization truly has low stock, or perhaps they are deceiving drive deals. It is difficult to tell, which is the reason countless such sites were flagged by the research group.

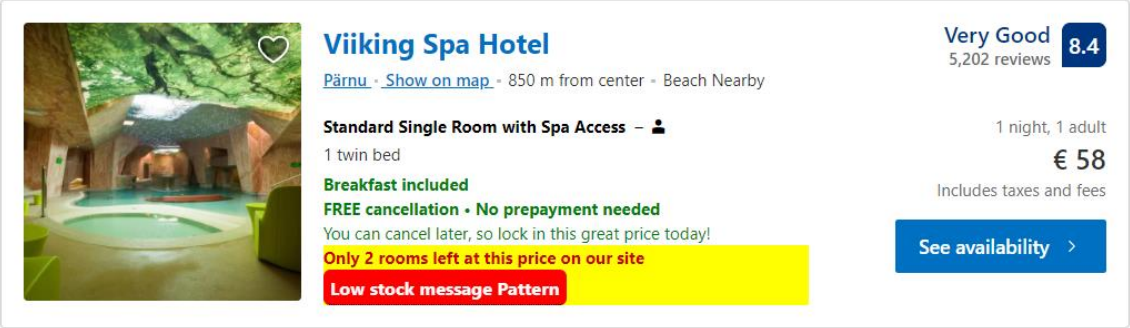

Figure 12. Low stock message

Figure 12 shows that booking.com websites are using similar kind of pattern to manipulate users. I have manually checked booking.com website for seven days and this specific page but surprisingly I have seen the exact output from it. But still, it is difficult to say that they are doing this deliberately.

High Demand Message: The 'High-demand Message' dark pattern indicates that the product is in high demand and sold out very soon.

High-Demand messages work similarly, yet they are more manipulative. "Low stock" shows that things are so famous, the organization is coming up short. "High-demand" signifies the same thing, yet it is significantly more uncertain.

Forced Action: Forced Action is another category of dark patterns originally proposed by Gray [7] that is a more extreme version of misdirection. This is not only misrepresenting an option but also forces the user to choose it.

Forced Enrollment: Many websites trick users into bringing them into their email list before using and accessing their basic functionality. Since this system utilizes forces the user to act against their own benefits, it considers a dark pattern.

This kind of dark pattern forces the user into pursuing promoting correspondence or create an account to give up their data. By utilizing Forced Enrollment, web owners and other organizations gathered more data about their users.

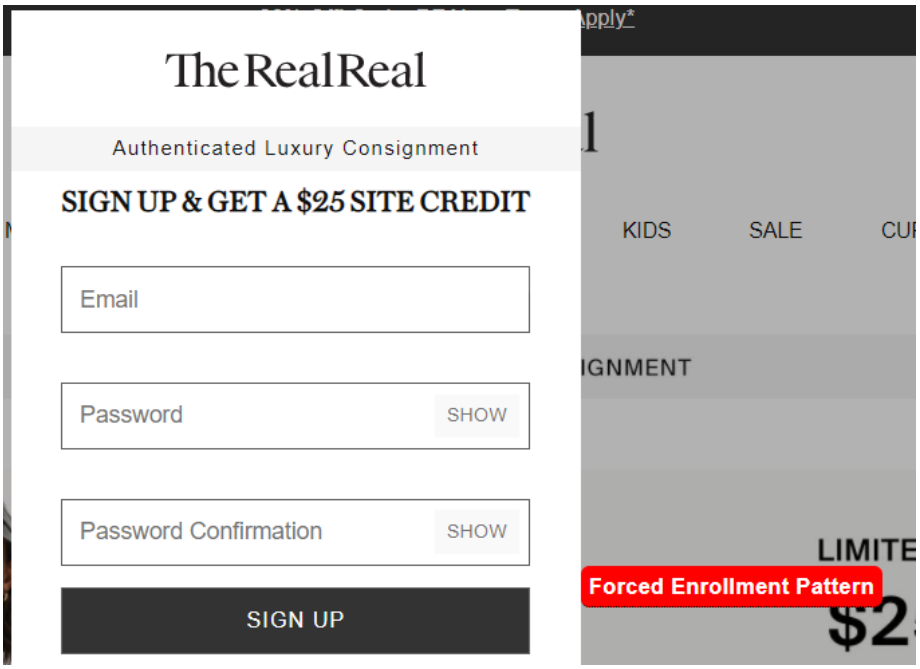

Figure 13. Force enrolment

Figure 13 shows TheRealReal (therealreal.com) using the same technique. They are using this pattern to force users to sign up to their system and then use their services. They will not even let users browse their websites without this.

4 Dark Pattern Detection and User Awareness Strategy

Technology is growing rapidly every day. Everything is converted online day by day. Online shopping, sharing personal details, hobbies, wishes, emotions in social media, inserting credit card information in e-commerce sites, and so many. Besides these advantages, this is threatening our privacy as well. In today's world, data privacy becomes a huge concern. Users are being manipulated by tricks that web designers use to build websites to make profits that they should not. A dark pattern is a term they use to do so. Preventing these dark patterns on websites is still unknown.

4.1 State of the Art

Lots of works, research have already been done on the field of users' data privacy. Developers and researchers are coming forward and providing various ideas, methodologies, and tools to alert users about their privacy. They are trying hard to stop this manipulation and tracking. They are working day and night to find out solution which can make users' information secure.

Summary of some research paper are given below:

A Survey on Web Tracking: Mechanisms, Implications, and Defenses [8]. Here they survey the existing literature on the strategies utilized by web administrations to follow the users online just as their purposes, suggestions, and possible user protections. They have introduced five primary groups of techniques for user tracking, which depend on sessions, client storage, client cache, fingerprinting, and different methodologies. Finally, they present future trends in client following and show that they can potentially deliver critical threats to the users' privacy. The range and variety of tracking strategies are exceptionally huge and broad. Accordingly, it is difficult to try not to be followed by any means. While utilizing the private browsing mode and an AdBlock could be a straightforward method to diminish the tracking rate for some users, this solution also needs the user's consideration; the user avoids visiting various services while being signed into one of them.

Effectively Protect Your Privacy: Enabling Flexible Privacy Control on Web Tracking [9]. This paper presents a novel framework to block third-party tracking by using

Word2Vec. Authors will likely make a more adaptable and well-created ruleset that can assist users with securing their privacy as per their requirements. Rather than impeding all-access tracking, they choose to give more consideration to the sites that have a solid possibility to gather the users' security. Word2Vec characterize the websites. The outcome shows that the error rate drops from 71% to 24% while using this system. They strongly believe it carries fresh blood into the field of web security by giving the new third-party tracking tool as well as a novel perspective about how to obstruct third-party tracking. But this paper researched 380 web pages which are not enough for machine learning. Machine learning requires a lot of data. It needs to work on more web pages so that the outcome of the research turns more accurate. Here they still have a 24% error rate. This error rate should be reduced. The algorithm should be improved to achieve a satisfactory judgment.

Evaluation of Third Party Tracking on the Web [10]. In this paper, their major goal is to measure the tracking components and the presence of trackers in websites so that they can identify the risk and dangerousness to privacy. They propose a natural privacy scoring model to measure the utilization of tracking strategies and recognize how users' activities are tracked when they are browsing websites. To get rid of this, they have made developed a Firefox add-on. This add-on collects the web browsing history of volunteers alongside the detected tracking components and computer the scores of the visited web pages [10]. There are two fundamental restrictions or limitations in this paper. The principal impediment is identified with the scoring model, which is basic and normal. This model is an endeavour to quantify the wild utilization of tracking methods on the Web. The second limitation is methodological. As clarified over, the utilization of their Firefox extension by Tunisian users applies an inclination to the study.

Increasing User's Privacy Control Through Flexible Web Bug Detection [11]. Here they have developed a Web Bug Detector. Web Bug is combined to a web page through images with some special properties. Users do not know about these images because they are hidden from them and hosted on a third-party site. Users will get to know about browsing tracking mechanisms when it detects and informs. This mechanism invisibly stores users' data while they visit multiple web pages. It means users will have better command over their privacy while permitting personalization. After applying the Detector to a real workload, they found about 5.37% of user's requests are tracked by third-party sites. Here, users do not have the ability to block Web Bug. But it is better if the users have this power.

Data obtained by the Detector is not validated because there is no other similar tool that can also obtain data and can be compared by each other. They did not build the proper interface that provides full functionalities.

Privacy risk assessment for Web tracking: A user-oriented approach toward privacy risk assessment for Web tracking [12]. Here, they worked on building a dynamic privacy scoring model that comes up with a solution. This model evaluates the risk to the users' privacy associated with tracking components. This model is a vigorously unique model to estimate the risk of privacy for users while browsing multiple web pages. They concentrated on two parameters, i.e., the seriousness and the relative importance of tracking components and their interactions. The commitment of this work is in its user-oriented methodology toward users' consciousness of privacy risk included when the browser the Web. Though this scoring model is proposed to evaluate the risk to users' privacy while they browse web pages, this has a limitation that tracker tools are not yet considered as in the privacy scoring model. Continuous works are tending to this issue and are extending the trial study by focusing on a bigger number of users.

The Web is watching you: A comprehensive review of web-tracking techniques and countermeasures [13]. This paper dissects and examines the current procedures for web-tracking just as strategies for its recognition and investigation, and countermeasures to prevent web following. They have considered the prevalence of various sorts of web tracking and fingerprinting. These works have discovered that web tracking is an exceptionally common practice and that not just it depends on its less harmful structures as web analytics yet, in addition, progressed fingerprinting techniques, for example, textual style testing or canvas fingerprinting.

This paper consists of analyzing web tracking techniques. They are aware of users not being manipulated by web tricks. They did not make any tools that will detect those traps automatically.

Privacy-aware Web Services Selection and Composition [14]. In this paper, the author initially examines the privacy protection issue as per characterize four types of shared information inside web service composition and afterward present web service composition and privacy models, this model permit user and web service provider to characterize user's privacy preference and web service's privacy strategy specification.

An algorithm is introduced to check the approach consistency on the dependency edge and select a bunch of strategy consistent web services with compositing a good privacy assurance web service composition. Analyses exhibit the applicability of the models and algorithm.

Privacy Scoring and Users' Awareness for Web Tracking [15]. Here, the author proposes a way to deal with improving users' awareness based on scoring the privacy risk and based on distinguishing the parameters just as the relationship between parameters that establish a high risk to users' privacy. They propose an intuitive scoring model to quantify to which degree users' protection is in danger when perusing the Web. They present a Firefox add-on that measures and alarms users about Web tracking. Two versions of this extra (desktop context and mobile android devices) were conveyed to two sets of Tunisian volunteers (students and analysts). The author then examines the measurable aftereffects of the gathered datasets and shows that the clients' conduct, the trackers, and tracking parts can build the risk of security violations on the Web.

Third-Party Web Tracking: Policy and Technology [16]. This paper studies the current approach debate encompassing third-party web tracking and clarifies the important innovation. They analyze the security ramifications of the contrary case where a first-party site approves a third-party site to find out about its users. It additionally presents the FourthParty web estimation stage and studies they have conducted with it. Their point is to inform analysts with a fundamental foundation and devices for contributing to public understanding and strategy debates about web tracking.

Web Tracking Site Detection Based on Temporal Link Analysis [17]. In this paper, the authors propose Web tracking webpage recognition and blacklist age-dependent on temporal link examination. Their proposal investigates traffic at the network passage with the goal that it can screen all tracking sites in the authoritative network. The proposed algorithm develops a chart among sites and their visited time to describe each site. Then, the framework characterizes suspicious destinations utilizing machine-learning algorithms. They affirm that 62-73% of blacklisted sites are recognized by the proposed framework, and 96% of unlisted sites are obscure or suspicious tracking sites.

Dark Patterns at Scale. Proceedings of the ACM on Human-Computer Interaction [6]. The author presents computerized procedures that empower specialists to recognize dark

patterns on an enormous set of sites. Utilizing these procedures, they study shopping sites, which frequently utilize dark patterns to impact users into making a bigger number of buys or revealing more data than they would something else. Examining a huge number of pages from different shopping sites, they find numerous dark patterns occasions. They inspect these dark patterns for misleading practices and find 183 sites that take part in such practices.

Finally, they build up a scientific classification of dark pattern attributes that depicts the fundamental impact of the dark patterns and their possible harm on user decision-making. Considering their discoveries, it makes suggestions for partners, including analysts and controllers, to consider, mitigate, and limit the utilization of these patterns. This proposed framework considers text-based dark patterns and, hence, leaves out those that are characteristically visual. Large numbers of the dark patterns they record are gotten from the current dark patterns writing. However, a portion of these exists in a gray area. In those cases, they are deciding if a dark pattern is intentionally deceptive. They just scrolled product pages and checkout pages, passing up dark patterns generally present in different pages, for example, the landing page and product search. Some dark patterns also show up after purchase.

4.2 Detection Strategy

Dark patterns are interface plans that push users towards behaviour that is against their interests. Since users are regularly ignorant that these malicious patterns impact them, research needs to recognize approaches to secure web users against them. One chance is to caution users when they experience uses of dark patterns on the web. This work moves toward this procedure, in which I intend to identify the utilization of patterns on pages naturally. Apparently, there is not a lot of work directed in the broad field of pattern detection on pages. One outcome is finished by Mathur et al., who utilized automatic crawling to distinguish dark patterns on shopping sites. Be that as it may, their attention was on making dark patterns, scientific categorization, and surveying how normal dark patterns are on the analyzed shopping sites [6]. Another work is finished by Philip et al., who utilized automatic detection of malicious patterns within the domain of cookie banners [1]. I will likely form an overall system to identify dark patterns on arbitrary web pages.

4.2.1 General Awareness Strategy

Users as consumers do not have all that amount of power over how organizations are storing their information and how well they are keeping it hidden. There are various steps users can take that can improve the protection of data. A decent initial step is to familiarize themselves with the privacy tools that are accessible. Users can prevent web tracking by themselves. Here are few ways users can follow to stay safe from being tracked.

Table 6. Methods to avoid being tracked

Methods	**Recommendation**
Change browser settings	Most browsers have basic protections against tracking and use this by default, but some browsers do not use those protections by default. The user should configure those settings before browsing.
Clear cookies regularly	Regularly clearing cookies and deleting browser history can temporarily help users from not being tracked. But this is not a permanent solution.
Private browsing mode	Every browser has a private browsing mode, and that mode does not store browsing history and cookies. Users can browse in private mode when they are working with some confidential data.
Ad blocker	Tracking is also done by ads. This can be avoided if the users start using the adblocker extension. This will block ads.
Use VPN	Although some countries banned using VPN, they are sometimes helpful. Using VPN software, browsing data will be encrypted.
Use anti-tracking software	Software like Avast AntiTracker find and block trackers. It also blocks cookies and show which websites are tracking users.
Use multi-factor authentication	Multi-factor authentication provides additional security layers. This will help users from not being hacked. It is better not to use SMS-based MFA. Many companies nowadays offer multi-factor authentication to provide additional security to the users.

Table 6 describes some methods that will help users to avoid being tracked. Many other strategies can also be followed, such as switching to a secure browser, limiting social media interaction, block cookies, etc.

Table 7. Dark pattern detection strategy

Pattern Type	**Case**	**Recommendation**
Sneak into Basket	When users add products to their cart page, they see some extra products on the cart page.	If they encounter some different product, they should check all the products one by one and find out those additional products added automatically by the system. If users can not remove or uncheck it, they should avoid checking out and that website also.
Hidden Costs	When a user encounters an extra charge on the cart page.	If the user sees some extra charge applied with the total amount he did not use, he should check the list one by one and calculate each amount by himself to make it clear. If he finds some imbalance in the total amount, he should check every part of the cart page. Users sometimes do not focus on the total amount, and that is where the website owners take advantage.
Hidden Subscription	When asked for user identity to use the application.	Some applications use this strategy to bring users into their list. Before using their services and providing personal data to them, check the terms and conditions. Check the subscription process and make sure it is free of charge for one time or for all time. If it states that it will not be free of charge or unsubscribe procedures are not clear, then

		the user should avoid using this application.
Misdirection	When a user has no idea about the websites, he is using. The user does not find everything in a normal way. When the website's actions buttons are not designed properly, and other designs look suspicious.	Look for reviews about that application. If the websites have bad functionality, users can be able to see in the reviews. Usually, bad websites will have bad reviews. If users are not completely sure and confident about the functionalities offered by the application, they should avoid using it.
Confrimshaming	When users see odds of poking in the email subscription stage or popup notices.	Always read the text carefully. If it states that some odds are poking, try to figure out the actual meaning of the sentence. The sentence might be insulting but if it is not causing harm to a user, then being insulted is better.
Visual Interference	When a user does not find his desire options or buttons. He finds it hard to see.	Go through every sentence of the websites. Sometimes these buttons or options are put someplace that is not noticeable, and they put it in an unexpected area. It is better to check every single sentence or word in the expected area and an inconspicuous area.
Trick Questions	When a user wants to subscribe but is asked to put answers before subscription.	When a user finds these answers or questions are not fully understandable or suspicious, then that can be considered as patterns. Usually, these answers or questions are represented with clear meaning so that there will be no confusion among users. The user should read the sentence carefully several times. If they are still unclear, then they do not need

		answers. It is better to stay away from that application.
Force Enrollment	When a user encounters a website that cannot browse without enrolling.	Usually, websites do not have mandatory enrollment. If the user finds such a website that is forcefully enrolling users before providing full browsing access to the application, he should carefully read the subscription form, terms, and conditions, payment issues, etc. If he finds everything satisfactory, he can continue by signing up for the application.

Table 7 describes about how to handle these dark patterns and where user can notice this kind of issues.

When in doubt for managing dark patterns, if you do not know about a specific choice, do not do it, search the web before you take your action about that site or application. Do not expect that the designers or site owners do not have any terrible goals since they may be and you are the next survivor of the dark patterns, do not simply address questions or give your credit card credentials without ensuring what you are doing and try not to manage these sorts of sites that utilize dark patterns. Though these patterns are not illegal only way to prevent organizations from using patterns is by increasing the awareness among users about these issues.

4.2.2 Detection Proposed Methods

This technique intends to tackle the issue utilizing cross-checking ways to match with specific keywords and identify dark patterns as conventionally as expected. The used model depends on input highlights from the Document Object Model (DOM), a tree-like portrayal of an HTML report, though every node of the DOM addresses a piece of the page's document. In this tree, every node can be treated as a likely possibility for addressing a dark pattern. To match the dark patterns, the scanned data has been checked in several ways. One is through a database by comparing with certain keywords stored on

the server-side before, and another one is by using HTML tag, attribute, attribute value, and content.

4.2.3 Systems' Requirement

To build any software, app, or anything that should have a required list. There must be a requirement list to build any kind of application. To make this work completely functional, it also needs some tools.

Server-side Development: There are many server-side programming languages. Some popular server-side programming languages are Python, Ruby, C#, PHP, and JavaScript (Nodejs). Developers can use any of these programming languages to write server-side code. This code has full access to the server operating system. The developer can choose any of these with a specific version.

In this work Python has been used in the server side to complete the task. There are so many reasons behind choosing Python as a server-side programming language.

Python: Python is a high-level programming language. Python is now ideal for rapid application development because of its built-in structure and combination with binding and dynamic typing. Python additionally offers support for packages and modules, which permits system code reuse and modularity.

It is one of the quickest programming languages as it requires not many lines of code. Its focus is on simplicity and readability, which attracts amateurs. These are the reasons behind choosing python as server-side development.

Django Framework: Django is an open-source, full-stack, and free web application structure written in Python. In short words, Django is a bunch of instant parts that assist the user with building sites with logical and clean structures. Among Python frameworks, Django is frequently viewed as the best framework for web application development.

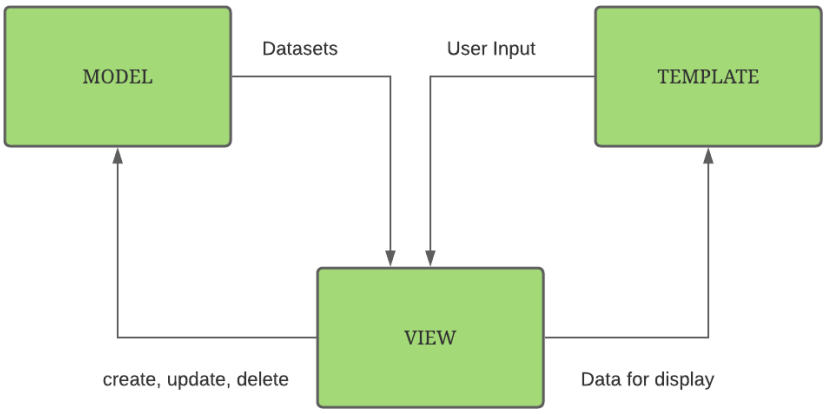

Figure 14. Data processing structure using Django framework

In this system for server-side development, I have used Django framework because this system also follows MVT architecture.

Figure 14 shows Django follows an MVT architecture which represents Model-View-Template. MVT is a Django variety of the renowned MVC structure; that is the reason the user will feel it is very comparable to how different frameworks work. At the point when the Django server gets a user request for the resource, Django works a controller and check to the available resources in URL. The URL router maps the request to the suitable view. The view at that point brings the information through the models, fills the format, and sends it back to the user.

Client-Side Development: Client-side code is the code that runs in the browser and is concerned about improving the appearance and conduct of a rendered web page. This incorporates choosing and styling UI parts, form validation, creating layouts, navigation, and so on.

Client-side code is composed utilizing HTML, CSS, and JavaScript. Inside the web browser, it is run and has practically zero access to the hidden operating system (counting restricted access to the file system). It is creating the code that is sent in the server's response. For this system I have used HTML, CSS, and JavaScript for 'TrackMe' extension and Bootstrap (open-source CSS framework), CSS, font-awesome for 'darktracker.xyz' web application.

Database: There are many free database engines users can use to learn SQL. The most popular are MySQL, SQLite3, Oracle Express, etc.

Each of these is an excellent database, and it is very difficult to recommend one. For this project, SQLite 3 is used to store user input data from darktracker website in the database, and for storing pattern category, patterns, and research report, CSV data file is used.

SQLite 3: SQLite is a database engine and programming library that gives a relational database management system. The light in SQLite implies lightweight regarding setup, database administration, and required resources. In a single file, a database is stored in SQLite — a quality that recognizes it from other database engines. This reality considers much openness: copying a database is not any more complicated than copying the file that stores the data.

In this system SQLite 3 has been used to store information except storing pattern type, patterns list, and scanned report. From web application the user request form is also saved in this database. It is used because it does not need a server. Due to the serverless architecture, users do not have to introduce or install SQLite prior to utilizing it. No server interaction should be started, configured, and halted.

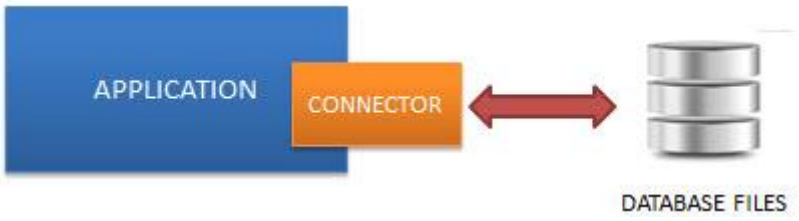

Figure 15. Serverless data processing using Sqlite

Figure 15 show that SQLite database is incorporated with the application that gets to the data set. The applications associate with the SQLite database, write, and read straightforwardly from the data set stored on disk. SQLite is usable in any environment, particularly in installed gadgets like Android, iPhones, game consoles, handheld media players, and so forth.

CSV and Pandas

For machine learning and data science datasets, one very popular format has been introduced: CSV (Comma Separated Values). A CSV record is a human-readable text file where every line has many fields, isolated by commas or some other delimiter. Users can consider each line a row and each field as a column. The CSV format has no norm; however, they are comparable enough that the CSV module will read by far most of CSV

records. For saving pattern type, patterns list and scanned report CSV data file is used in this system. From the CSV data file, it then showed in admin panel.

For Python programming, a new software library has been written and introduced name Pandas. One crucial feature of Pandas is its ability to write and read Excel, CSV, and many other types of files. Functions like the Pandas read_csv() method enable user to work with files effectively. User can use them to save the data and labels from Pandas objects to a file and load them later as Pandas Series or DataFrame instances. In this system to read and write the CSV data file Pandas has been used. All the queries in CSV are done using Pandas rules.

DOM Manipulation

When composing web pages and applications, perhaps the most well-known things user will need to do is control or manipulate the data structure in some way or another. This is typically done utilizing the Document Object Model (DOM), a bunch of APIs for controlling HTML and styling data that uses the Document object.

A Web page is a report. This report can be either shown in the browser window or as the HTML source. Be that as it may, it is a similar report in the two cases. The Document Object Model (DOM) addresses that equivalent document, so it very well may be controlled or manipulated. The DOM is an object-oriented portrayal of the site page, which can be altered with a scripting language like JavaScript.

In this system I have used this technique. But I have avoided scanning head. This system only scans body section.

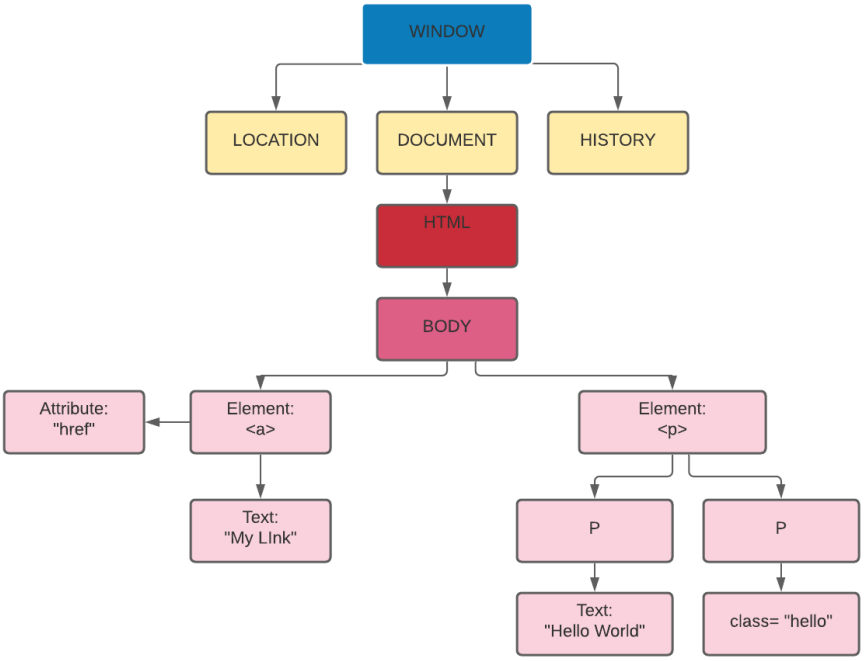

Figure 16. HTML DOM tree of object (excluding head section)

Figure 16 shows a representational tree and how the browser creates it. All the attributes can be found from a specific HTML element. The attribute class=" hello" is an attribute from the <p> element in the figure. In this system only body section will be scanned which is shown in the figure. Some pre-selected elements are listed as ignored scanning elements and scanning elements.

```
const blockElements = [
    'div', 'section', 'article', 'aside', 'nav',
    'header', 'footer', 'main', 'form', 'fieldset', 'table'
];
const ignoredElements = ['script', 'style', 'noscript', 'br', 'hr'];
```

Figure 17. Selected and ignored element list

Figure 17 describes which elements are selected and which are ignored elements while scanning the page. In this scanning div, section, article, aside, nav etc will be selected and script, style, noscript, br, hr will be ignored.

For example, the querySelectorAll method is specified by the standard DOM in the code below must return a list of all the <a> elements in the document:

```
const x = document.querySelectorAll("a");
// x[0] is the first <a> element
// x[1] is the second <a> element, etc.
alert(x[0].nodeName);
```

Manipulation Methods

The DOM has many methods that can manipulate content. They are the connection between events and nodes (elements). For this system I have used this manipulation methods to get data by Id, or by class name etc to match with the logic. To grab the elements, we can use both older and modern methods. Some of those are given below which I have used in this work as well.

getElementById()

It returns the element by checking the id. We all know, id is always unique, so it is very helpful to get element which we exactly want.

```
var helloWorld = document.getElementsById('hello');
```

getElementsByClassName()

This technique returns a HTML Collection of each one of those components containing the name class passed.

```
var containerBox = document.getElementsByClassName('box');
```

querySelector()

It returns the primary component that has the CSS selector passed. Simply recall that the selector ought to follow the CSS syntax. If the user does not have any selector, it brings null.

```
var submitButton = document.querySelector('#submit');
```

querySelectorAll()

Basically the same as the querySelector() technique, however, with a single distinction: it returns every one of the components that match with the CSS selector passed. The selector ought to likewise follow the CSS sentence structure. If the user does not have any selector, it brings null.

```
var submitButton = document.querySelector('#submit');
```

These are the basic procedures or methods that can use to manipulate DOM. There are so many methods which can be used to manipulate also.

To manipulate a component inside the DOM, user first need to choose it and store a reference to it inside a variable. Below is the code for this purpose.

```
const link = document.querySelector('a');
```

Having the element reference stored in a variable, we can start to manipulate it using properties and methods available to it. We can change the text inside the link by updating the value of the Node.textContent property. Below is the code that will fulfil this task.

```
link.textContent = 'Dark Pattern';
```

We can likewise change the URL the connection is highlighting to do not go to some wrong site when it is tapped on. Below is the code that will fulfil this task.

```
link.href = 'http://darktracker.xyz/';
```

5 Development of ‘TrackMe’ Extension and ‘darktracker.xyz’ Websites

After explaining today's web traps and dark patterns, it is time to describe the system that has been built to track the dark patterns trap. In this case, it will be an extension. One more web application that has been built will also be described. This web application will mainly create awareness among users about Dark Patterns traps. This application will have brief information about recent dark patterns that organizations use, one user question-asking question form, and the information that the web collected from the user. Main Section 5.1 will be an explanation about those applications, how its work, flowchart, in section 5.2, there will be user interface design, in section 5.3 there will be functionalities and the last part 5.4 will show the final result.

5.1 Analysis of the Systems

For analysis, the section will be divided into two parts. One part will be an analysis of the extension, and another will be an analysis of the web application.

Analysis of ‘TrackMe’ Extension

Identifying dark patterns requests Detector analysis of the user's website page. It will be taken care of by a web browser plugin. Web browser plugin functions as a detector that scans web page body components and then responses to this request. A plugin is an application introduced and executed over popular web browsers. Such applications are browser extensions, which add some usefulness to consumers.

There is not a lot of work directed in the broad field of pattern recognition on site pages. One special case for this is Mathur et al., who utilized automatic crawling to distinguish dark patterns on shopping sites. In any case, their emphasis was on making a dark patterns scientific classification and surveying how regular dark patterns are on the analysed shopping sites [6]. This work's goal is to create an overall structure to identify dark patterns.

There are so many challenges. Website pages are heterogeneous. There will be a multiple topic, e.g., news, games, or shopping. Now the pages' structure is heterogeneous, and it

is difficult to plan an overall way to deal with identify certain examples in an assorted arrangement of pages. Hence, it is required to build up a standard set to recognize patterns on a wide website page scope. However, this model should not be explicit to all dark patterns in any event to the regular ones without any problem. Considering the heterogeneity of website pages, unmistakably rule-based strategies are not adequate to handle dark patterns discovery. An extra issue is the uncommonness of dark patterns on pages rather than ordinary page components or other more successive patterns with no bad goal.

Thus, this work means tackling the issue utilizing Harry Brignull's [4] ways to identify dark patterns as conventionally as could be expected. The employed model depends on input highlights from the Document Object Model (DOM), a tree-like portrayal of an HTML archive. Interestingly, every node of the DOM addresses a part of the page document. In this work, just the body component segment of the HTML tree will be checked to follow dark patterns. To be more exact, one can likewise think about the entire tree of the DOM as an expected example; however, we limit the conversation to the body area in this work for simplicity.

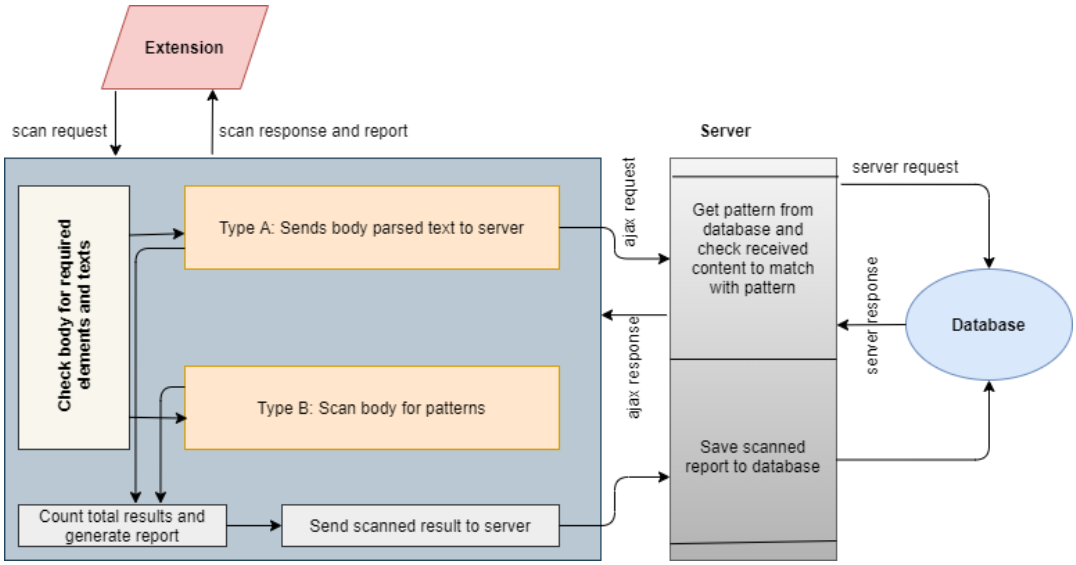

Figure 18. System architecture (extension)

Figure 18 shows the total architecture of this work. The extension will request for the body scan element to a web page by injecting js and web page get the response after being scanned. The user will get to see the scanned result into the extension within a second.

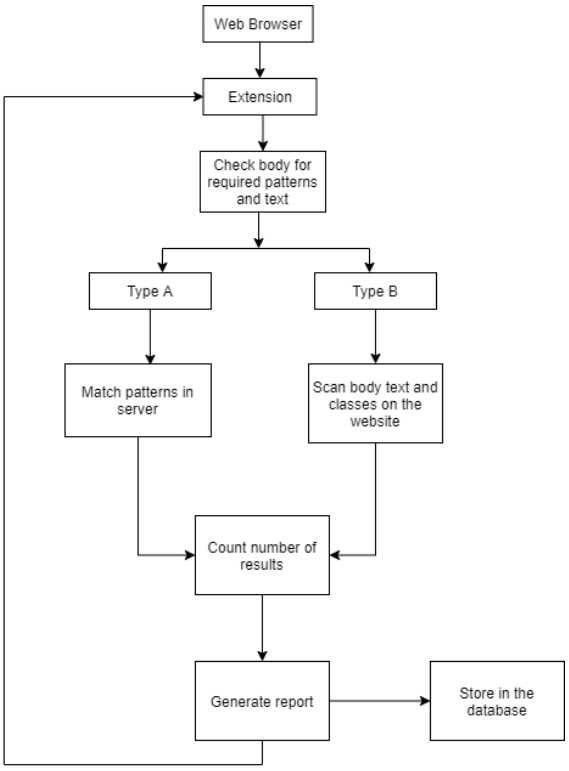

Figure 19. Flow chart (extension)

Figure 19 explains how the whole system's process work. When the extension is clicked, it starts to inject JS to that specific web page and checking the HTML body element for the required element or text.

The checking process divide into two parts. In type A, it sends body parsed text to the server to check and match with a previously stored pattern. This will be done at Ajax's request. And in type B, it scans the body for patterns by using elements, attributes, class matching logical conditions.

After checking dark patterns by both parts, it starts to count the patterns that match and finally generate a result and report.

This result will go to the extension and the database. The user will then see the output result in the extension, and they will get to know how many patterns this system encountered and what those patterns are after checking.

Analysis of 'darktracker' Website

There is not much work that creates awareness among users about dark pattern tricks and traps. This web application is a website where users can see detailed information about

that and will be a good solution to create awareness. They can also ask the system's administrator about dark patterns by filling up one question asking form.

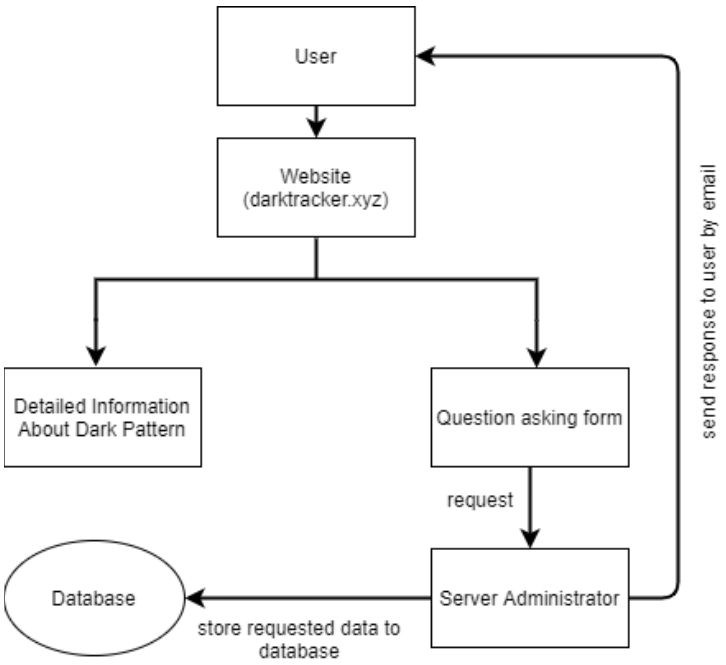

Figure 20. Flow chart (website)

Figure 20 shows the workflow diagram of the website, and it is described below.

When users visit this website, they will notice a home page and a dark patterns-related category page containing brief information about dark patterns.

There is one question asking form. The users can request the system's administrator by filling up the form if they notice any dark patterns or have doubts about some element or portion of web pages. This form has multiple fields such as source URL, list of patterns they suspect, picture uploading field, user's name, and email address.

After getting requests from the users, the administrator will analyze that specific portion of the web page. After analyzing, they will send the response to the users by using their email.

The system's administrator will use the user's requested data to enhance their database with more dark patterns-related information and update their dark tracker pattern logic. This will make this system more dynamic.

5.2 User Interface Design

Designers use the process named User Interface (UI) design to build interfaces in web apps, computerized devices, or software, focusing on style or the way it looks. The prime goal is to make this interface pleasurable, easy to use, and understand. This refers to graphical user interface and other forms such as gesture-based interfaces.

Since this project has two applications, it is necessary to describe both applications in this section.

'TrackMe' User Interface

To build the Dark Tracker extension, it was important to choose which browser may host this. There are so many browsers, but for this work, the most popular browser such as Mozilla Firefox, Google Chrome, Microsoft Edge, and Opera have been chosen. This Tracker is implemented as an extension. This decision was mainly because Mozilla, Chrome, Edge, and Opera is the most used and open-source web browser. It was easy to make a user interface, and implementation was also easy.

For content building and styling, I have used simple HTML and CSS. For the font, it was font-awesome-4.7.0 and font-family Poppin.

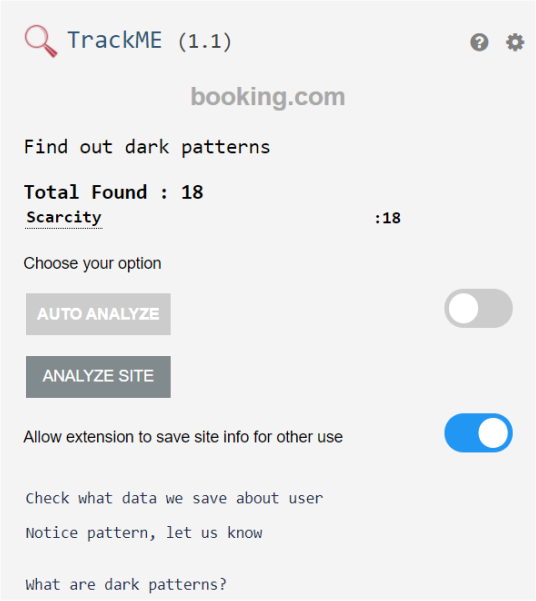

Figure 21. UI of TrackMe extension

Figure 21 shows the final visual representation of the user interface of the Dark Tracker, which looks nice and simple. This interface is undoubtedly user-friendly and purposeful.

'darktracker' User Interface

This is an information-based application that contains brief data about Dark Patterns. It was kept in mind that this website should be built to attract users easily and does not make users bore before starting to develop this website. This website creates a relationship with the user and system's administrator through a user question asking form. To build this site, I have used Bootstrap, CSS, and JavaScript. It has multiple pages with a variety of information and has menus. Some important views are shown below. This is an information-based application that contains brief data about Dark Patterns. It was kept in mind that this website should be built to attract users easily and does not make users bore before starting to develop this website. This website creates a relationship with the user and system's administrator through a user question asking form. To build this site, I have used Bootstrap, CSS, and JavaScript. It has multiple pages with a variety of information and has menus. Some important views are shown below.

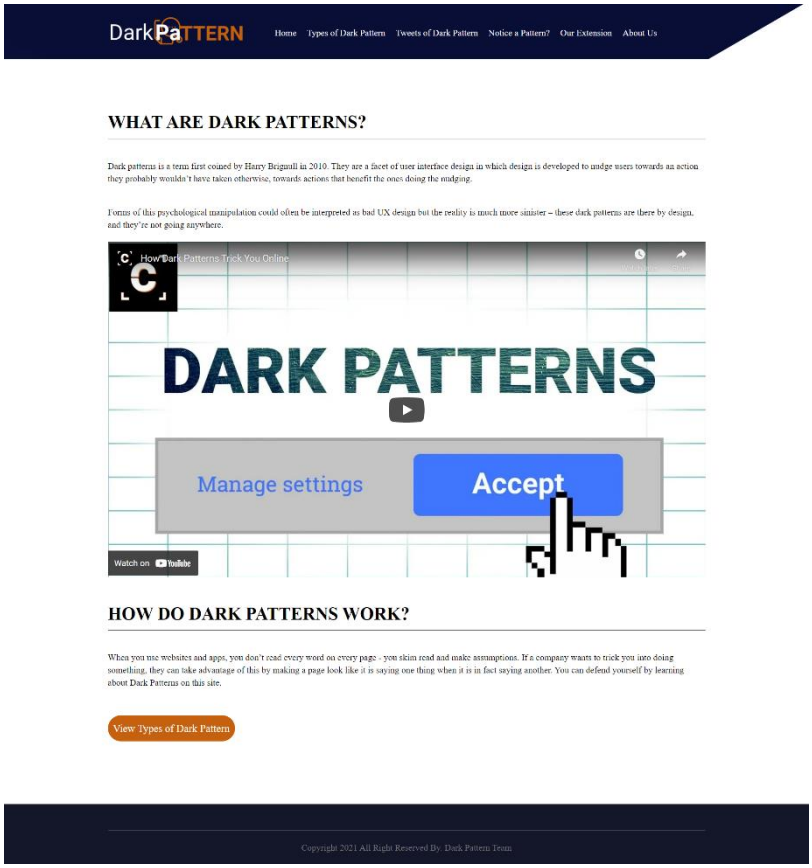

Figure 22. Front view of website (darktracker)

Figure 22 shows the home or front page of the website. Describing the basic information about dark pattern.

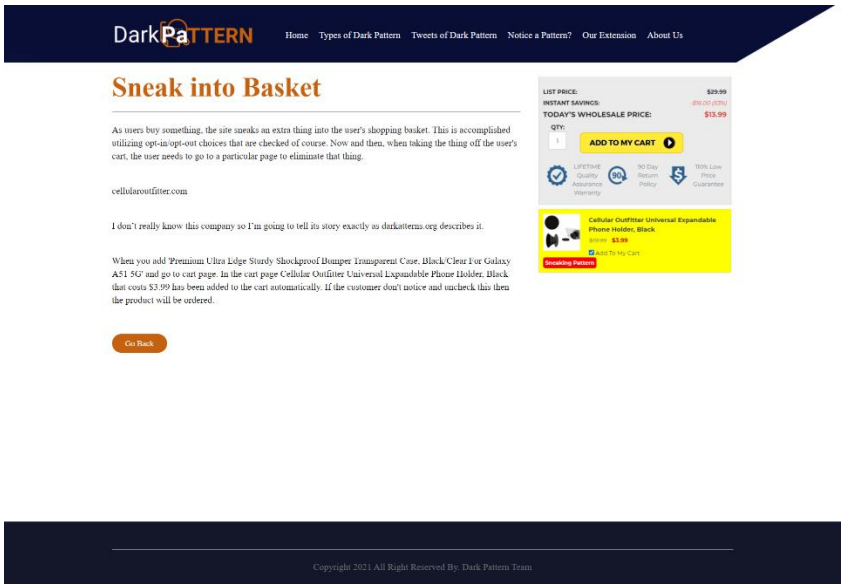

Figure 23. Content page (website)

Figure 23 shows the description page of one of the types of dark patterns. Here it shows 'Sneak into Basket' type, what this is and an image that will clear the idea about this to the users.

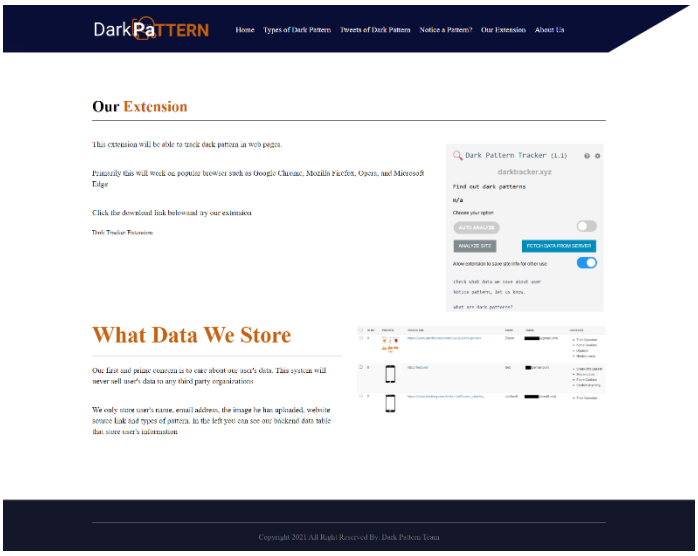

Figure 24. Our extension page (website)

Figure 24 shows the most important page of this website. This page contains the information about the extension, how it looks like, download link, and most importantly what data about users this extension is storing.

5.3 Functionalities

Functionality is how those features work to give the user the ideal result.

In software engineering, a functional prerequisite characterizes a framework or its segment. It depicts the functions a product should perform. A function is only inputs, its output, and its conduct. It might be information control, estimation, client cooperation, or whatever other explicit usefulness characterizes which function a system will probably perform.

This work required so many functionalities to get the desired output. Both extension and web application function properly after applying the necessary coding.

The extension has many functionalities which are fulfilling the desired outcomes. To add the functionality, this needed Python Django, packages such as matplotlib, nose, numpy, demjson, cycler, pandas, sqlparse, etc., for database sqlite3 and CSV as a data file.

The main functionality of this extension is to track dark patterns and show the result to the user through extension. But this main functionality is a combination of some small pieces of other function. Without these small functions, main functionality cannot be achieved. Here, the main functionality of the extension is divided into sub-functions such as web page scanning function, pattern matching function, show output result function, and storing output result into database function to make the system more understandable.

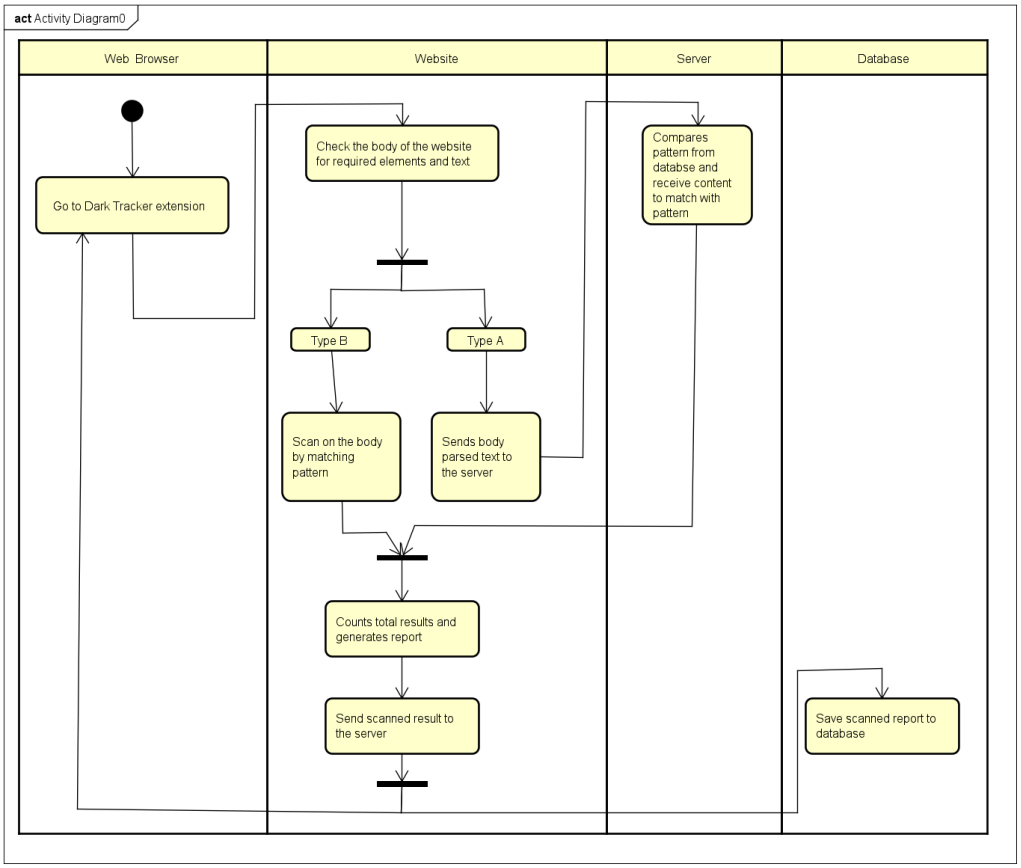

Figure 25. Activity diagram (extension)

In software engineering, a functional prerequisite characterizes a framework or its segment. It depicts the functions a product should perform. A function is only inputs, its output, and its conduct. It might be information control, estimation, client cooperation, or whatever other explicit usefulness characterizes which function a system will probably perform. Figure 25 shows the activity diagram of the 'TrackMe' extension. Here the extension is divided into four partition such as web browser, website, server, and database. Every partition has different functionalities. In the web browser partition initial node start the procedures with the block dark tracker extension. It will go to the website partition to check the body of the website for required elements and text. Checking body divided into two section type A and type B which is represented using fork node. Type A sends body parsed text to the server. Server get the request and send back the response. Type B scan on the body by matching patterns using js. Both the scan data and responded data from server join using the join node. Next object does count total results and generate report which will then pass to next block. This block sends scanned result to the server. Then the result then saved into database and send to dark pattern extension for the user.

As we can see from the above Figure 21 the UI of “TrackMe” extension. Output of the extension, having so many buttons. Each button has different functionalities.

Auto analyse: Sometimes users want to analyse the web page automatically. They do not want to click analyze site button every time. To get rid of this system has introduced this auto analyse button.

Allow extension button: This button is very important in this application. As we all know data privacy is now a big concern. Considering this issue in mind this application came up with this solution. In this system after scanning, the scanned report is saved in the database. Without user’s permission it should not save the scanned data. So, it is very important to get the permission from users and that is why this button has created. If the user does not enable the button, analyse page will not be functioned.

Data check button: Using this link user can see what type of data are stored by the system. If the user found this inappropriate, then they will be able to stay away from using this application.

Notice pattern: By pressing this link user will get to see the Notice pattern form. By this form user can ask system about the section of the page they are suspecting. User can clear their confusion about dark pattern.

Dark Pattern: This link will take user to the web application where user will get every necessary information about dark patterns, its type, how it works etc.

Main functionality of the web application ‘darktracker’ is to create awareness among users about dark patterns. This application consists of several pages which contains important information about dark pattern. User then be able to recognize the dark pattern if they notice something like this. Main functionality of this application is the user asking form which can build a relation between users and system administrator.

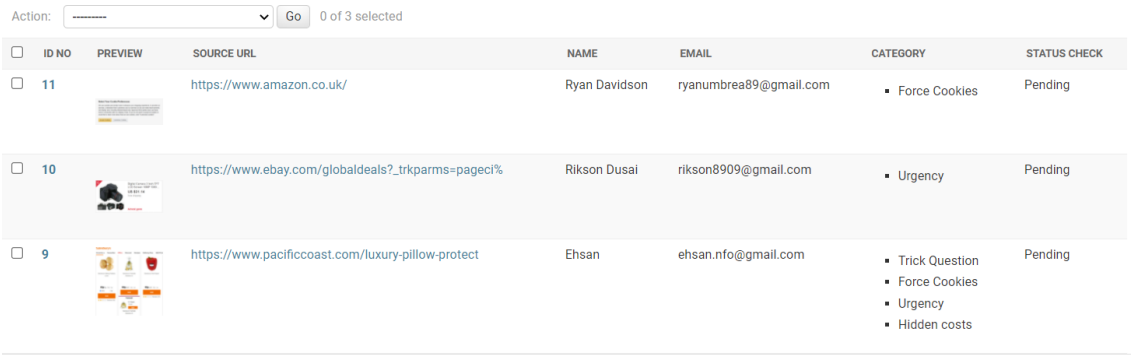

Figure 26. Request track data table (admin panel)

This form function using sqlite and store the request in the system admin panel database. Figure 26 shows the requested track data table that listed in admin panel. System administrations will analyse that web page, check the requested data and response after finishing analysing every single element of that. (Demonstrated clearly in Appendix 2 - Description and Manual)

5.4 Final Result

Data privacy is the big concern now in today's world. In order to keep users safe from being tracked and trapped by web owners or third party, new methods are developing every day. This work also tried to come up with solution that can help users and create awareness.

After spending days and nights working on this work finally the output results appear. After implementing all coding and functionalities this system works as expected.

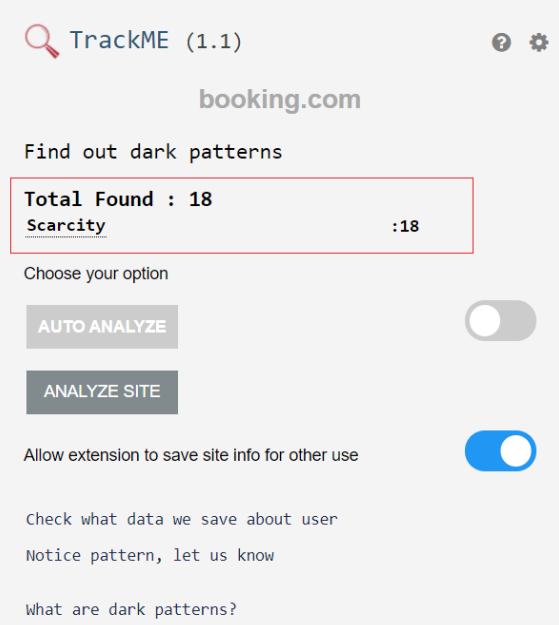

Figure 27. Scanned output result in extension

Data privacy is the big concern now in today's world. To keep users safe from being tracked and trapped by web owners or third party, new methods are developing every day. This work also tried to come up with solution that can help users and create awareness. In figure 27 we can see in the red box; the result is showing Scarcity. Here scarcity is the pattern name. These names are clickable which will take user to the brief description of that specific pattern name in the system's another web application 'darktracker.xyz'. The value is 18 that scarcity has been counted 18 times while scanning the page. Total result is all the valued counted together that is shown in the red marked box as well.

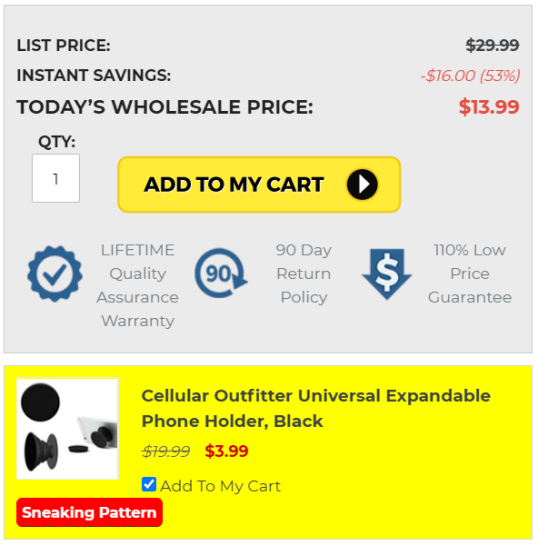

Figure 28. Scanned output result in webpage

Data privacy is the big concern now in today's world. To keep users safe from being tracked and trapped by web owners or third party, new methods are developing every day. This work also tried to come up with solution that can help users and create awareness. Figure 28 shows scanned output result in the web page. When user hit the analyze button, the portion of the dark pattern in web page become yellow and the small red box appears containing the pattern name. Here we can see it is showing sneaking pattern in the red highlighted box and the background of the area also become yellow.

These are the two main output results of this work.

In short, this work is:

- A windows-based extension for popular browsers developed that can track manipulation set by the web owners.
- Show the scanned report to the user to let them know about what patterns have been used in the page, its type, category, and brief description.
- A web application created to show latest information about dark patterns, how it is used, where it is used, its type and description.
- A user question asking form created to get request from users and store list of new web pages url for more analyses.

Although this work produces many results, still lot of works that should be done in the future.

6 Summary

There are numerous ways by which a user is being controlled online currently. Information accumulated explains the use of undeniable level innovation to take users' personal data and control the substance imparted to them. In this manner, impacting their sentiments, feelings just as insight effectively without their awareness.

This paper tends to a genuine concern identified with users' data privacy, online manipulation, and its ensuing exploitation. The reach and variety of tracking techniques are huge. Accordingly, it is not easy to try not to be tracked by any stretch of the imagination. The paper intends to familiarize the user with the distinctive tracking methods he or she experiences when browsing the web consistently and explains countermeasures against the strategies that are being utilized for online manipulation and tracking users' data. This paper additionally introduces two applications one is windows-based "TrackMe" extension and "darktracker" website that work against these issues and help users.

This work has some limitations that need to be acknowledged. First, a significant number of the dark patterns' archives are gotten from the current dark patterns writing. Nonetheless, a portion of these exist in a gray area, and in those cases deciding if a dark pattern is purposely deceptive or not can sometimes be difficult to recognize. Second, not a wide range of dark patterns as arranged by Gray et al. [7] will be distinguishable by the proposed system. For instance, bait and switch is a problem that normally is not explicit to web pages, however, is important for applications that cannot straightforwardly be addressed by the model. Third, here only text-based dark patterns and pattern matching by the value of element or attribute. Fourth, this system is not fully automated as it requires manual insertion of pattern in the database. So artificial intelligence is missing and in some cases error can appear. These limitations were not solved in this work, but this will be solved in the next version of this work in the future. There are varieties of opinions of dark patterns between and among users and experts. Explaining this dark side and building up how much these patterns are seen as manipulative by users can be additionally researched by future user studies. Future investigations could think about gathering these sorts of dark patterns from users and also work on limitations.

References

[1] M. G. Philip Hausner, “Dark Patterns in the Interaction with Cookie Banners,” [Online]. Available: https://arxiv.org/abs/2103.14956#:~:text=Dark%20patterns%20are%20interface%20designs,protect%20web%20users%20against%20them..

[2] M. D. T. H. a. N. P. Tobias Urban, “Beyond the Front Page: Measuring Third Party Dynamics in the Field,” 2020.

[3] P. Pedamkar. [Online]. Available: https://www.educba.com/what-is-gui/.

[4] H. Brignull. [Online]. Available: https://www.darkpatterns.org/.

[5] J. BROWNLEE. [Online]. Available: https://www.fastcompany.com/3060553/why-dark-patterns-wont-go-away.

[6] A. A. G. F. M. J. L. E. M. J. C. M. &. N. A. Mathur, “Dark Patterns at Scale. Proceedings of the ACM on Human-Computer Interaction,” 2019.

[7] Y. K. B. B. J. H. a. A. L. T. Colin M. Gray, “The Dark (Patterns) Side of UX Design,” 2018.

[8] T. C.-E. V. L. B.-R. &. B.-R. P. Bujlow, “A Survey on Web Tracking: Mechanisms, Implications, and Defenses,” 2017.

[9] S. V. D. V. &. S. K. Yu, “Effectively Protect Your Privacy: Enabling Flexible Privacy Control on Web Tracking,” 2017.

[10] A. K.-B. A. H. K. M. A. &. K. A. Hamed, “Evaluation of third party tracking on the web,” 2013.

[11] F. P. R. &. M. W. (. Fonseca, “Increasing User’s Privacy Control through Flexible Web Bug Detection,” 2005.

[12] A. &. B. A. H. K. Hamed, “Privacy risk assessment for Web tracking: A user-oriented approach toward privacy risk assessment for Web tracking,” 2016.

[13] I. U.-P. X. S. I. &. B. P. G. Sanchez-Rola, “The web is watching you: A comprehensive review of web-tracking techniques and countermeasures,” 2016.

[14] T. &. H. T. Li, “Privacy-Aware Web Services Selection and Composition,” 2014.

[15] A. &. A. H. K.-B. Hamed, “Privacy scoring and users’ awareness for Web tracking,” 2015.

[16] J. R. &. M. J. C. Mayer, “Third-Party Web Tracking: Policy and Technology,” 2012.

[17] A. M. H. &. M. Y. Yamada, “Web Tracking Site Detection Based on Temporal Link Analysis,” 2010.

[18] [Online]. Available: https://www.pewresearch.org/internet/2019/11/15/americans-and-privacy-concerned-confused-and-feeling-lack-of-control-over-their-personal-information/.

[19] [Online]. Available: https://www.salesforce.com/news/stories/state-of-the-connected-customer-report-outlines-changing-standards-for-customer-engagement/.

[20] [Online]. Available: https://www.cisco.com/c/dam/en/us/products/collateral/security/cybersecurity-series-2019-cps.pdf.

[21] [Online]. Available: https://unctad.org/page/data-protection-and-privacy-legislation-worldwide.

[22] [Online]. Available: https://iapp.org/resources/article/iapp-ey-annual-governance-report-2019/.

[23] [Online]. Available: https://www.emarketer.com/content/with-ccpa-days-away-many-companies-are-still-not-compliant.

[24] [Online]. Available: https://www.emarketer.com/content/with-ccpa-days-away-many-companies-are-still-not-compliant.

Appendix 1 – Non-exclusive licence for reproduction and publication of a graduation thesis[1]

I Md Ehsanur Rahman

1. Grant Tallinn University of Technology free licence (non-exclusive licence) for my thesis “Analyzing Web Traps and User’s Data Privacy”, supervised by Vladimir Viies, PhD, Associate Professor.
   1.1. to be reproduced for the purposes of preservation and electronic publication of the graduation thesis, incl. to be entered in the digital collection of the library of Tallinn University of Technology until expiry of the term of copyright;
   1.2. to be published via the web of Tallinn University of Technology, incl. to be entered in the digital collection of the library of Tallinn University of Technology until expiry of the term of copyright.
2. I am aware that the author also retains the rights specified in clause 1 of the non-exclusive licence.
3. I confirm that granting the non-exclusive licence does not infringe other persons' intellectual property rights, the rights arising from the Personal Data Protection Act or rights arising from other legislation.

09.05.2021

Md Ehsanur Rahman

1 The non-exclusive licence is not valid during the validity of access restriction indicated in the student's application for restriction on access to the graduation thesis that has been signed by the school's dean, except in case of the university's right to reproduce the thesis for preservation purposes only. If a graduation thesis is based on the joint creative activity of two or more persons and the co-author(s) has/have not granted, by the set deadline, the student defending his/her graduation thesis consent to reproduce and publish the graduation thesis in compliance with clauses 1.1 and 1.2 of the non-exclusive licence, the non-exclusive license shall not be valid for the period.

Appendix 2 - Description and Manual

This section provides the description about the functionality and how this system works. It contains a step-by-step process and guideline.

Each section has its own functionality, has its separate file.

1. This work can be downloaded through this link: https://github.com/ehsan-info/hub/tree/main

2. It has two applications one is ‘TrackMe extension’ and another is ‘darktracker’ web application.

3. Extension will work on windows for popular browser (Google Chrome, Mozilla Firefox, Microsoft Edge, and Opera) and web application can be performed in any browser.

4. Trackme extension folder is for extension. Inside this extension folder is for Google Chrome, Microsoft Edge, and Opera. Extension-firefox folder is for Mozilla Firefox. It contains most important files such popup.js and content.js. These two files used to scan web page and matching and checking pattern logic. Popup.html file is the main UI file.

5. Project_dark_tracker folder and public_html folder contains every data processing folder, web application folders and files.

Appendix 3 – Source Code

Extract Domain Url:

```
function extractRootDomain(url) {
    var domain = extractHostname(url),
        splitArr = domain.split('.'),
        arrLen = splitArr.length;

    //extracting the root domain here
    //if there is a subdomain
    if (arrLen > 2) {
        domain = splitArr[arrLen - 2] + '.' + splitArr[arrLen - 1];
        //check to see if it's using a Country Code Top Level Domain
(ccTLD) (i.e. ".me.uk")
        if (splitArr[arrLen - 2].length == 2 && splitArr[arrLen -
1].length == 2) {
            //this is using a ccTLD
            domain = splitArr[arrLen - 3] + '.' + domain;
        }
    }
    return domain;
}
```

Start scanning after receiving request message:

```
chrome.runtime.onMessage.addListener(
    function(request, sender, sendResponse) {
        if (request.message == 'analyze_site') {
            scrape();
        }
        else if (request.message == 'popup_open') {
            var element = document.getElementById('insite_count');
            var element2 =
document.getElementById('insite_count_details');
            if (element && element2) {
                sendDarkPatterns(element.value,element2.value);
            }else if (element) {
                sendDarkPatterns(element.value);
            }
        }
    }
);
```

Highlight function:

```
function highlight(element, type)
{
    element.classList.add('insite-highlight');

        var body = document.createElement("span");
        body.classList.add('insite-highlight-body');
        body.classList.add('dark-tracker-trace');

        var header = document.createElement("div");
        header.classList.add('dark-pattern-tracker-header');
        header.classList.add('darktracker-show-tooltip');

        var headerText = document.createElement("span");
        headerText.innerHTML = capitalizeFirstLetter(type) + '
Pattern';

headerText.setAttribute('onclick',"window.location.href='"+links[capit
alizeFirstLetter(type)]+"'")
        header.style.fontSize="13px";
        header.style.lineHeight= '15px';
        header.style.fontWeight="bolder";
        header.style.position="relative";
        header.style.zIndex = "99999999999999999999999";
        header.style.background = 'red';
        header.style.color = 'white';
        header.style.padding = '5px';
        header.style.borderRadius = '5px';
        header.style.cursor = 'pointer';
        header.appendChild(headerText);

        /* content */
        var content = document.createElement('span');
        content.classList.add('darktracker-show-tooltiptext');
        content.innerHTML = descriptions[capitalizeFirstLetter(type)];
        header.appendChild(content);

        body.appendChild(header);

    if(element.tagName.toLowerCase()=='img'){
        element.closest('div').insertBefore(body, undefined)
    }else{
        element.appendChild(body);
    }

}
```

Cookies

```
if(document.querySelectorAll(".darkTrackedCookie").length>0){

document.querySelectorAll(".darkTrackedCookie").forEach(function(evt){
                evt.classList.remove('darkTrackedCookie');
            });
}
        document.querySelectorAll("body div").forEach(function(e){
            if(e.parentElement.tagName=='BODY'){
                var text=e.textContent || e.innerText;
                if(text.toLowerCase().search("cookies")!=-1){
                    var hasClose=0;

e.querySelectorAll("a,button").forEach(function(b){
                        var btn_text=b.textContent || b.innerText;
                        if(btn_text.toLowerCase().trim()=='accept &
close' ||
                        btn_text.toLowerCase().trim()=='accept and
close' ||
                        btn_text.toLowerCase().trim()=='agree & close'
||
                        btn_text.toLowerCase().trim()=='agree and
close'||
                        btn_text.toLowerCase().trim()=='accept all
cookies' ||
                        btn_text.toLowerCase().trim()=='accept cookies
& close'){
                            b.classList.add('cookie-track-action');
                            return false;
                        }else if(btn_text.toLowerCase().trim()=="×" ||
                        btn_text.toLowerCase().trim()=="close" ||
                        btn_text.toLowerCase().trim()=="disagree"){ //
More close texts will be added here..
                            hasClose++;
                            return false;
                        }
                    });

                    if(hasClose==0){
                        if(e.querySelector('.cookie-track-action') &&
!e.querySelector('.cookie-track-
action').parentElement.closest('div').classList.contains('darkTrackedC
ookie')){
                            e.querySelector('.cookie-track-
action').parentElement.closest('div').classList.add('darkTrackedCookie
');
                            if(categoryListed['Forced
Cookie']!=undefined){
                                categoryListed['Forced
Cookie']=categoryListed['Forced Cookie']+1;
```

```
                            }else{
                                categoryListed['Forced Cookie']=1;
                            }
                            highlight(e, "Forced Cookie");
                            count++;
                        }
                    }

                }
            }
        });
```

Force action, Force Enrollment

```
if(document.querySelectorAll(".darkTrackedForced").length>0){

document.querySelectorAll(".darkTrackedForced").forEach(function(evt){
                evt.classList.remove('darkTrackedForced');
            });
        }
document.querySelectorAll("body>[class*='modal'],body>[class*='popup']
,body>[class*='overlay']").forEach(function(e){
            if(e.querySelectorAll('a,button').length>0){
                if(e.parentElement.tagName=="BODY"){
var hasClose=0;
e.querySelectorAll('a,button').forEach(function(b){
                        var text = b.textContent || b.innerText;
                        // More close texts will be added here..
                        if(text.toLowerCase().trim()=="×" ||
                        text.toLowerCase().trim()=="close"||
                        text.toLowerCase().trim()=="disagree"){
                            hasClose++;
                            return false;
                        }
                    });
if(hasClose==0){
if(!e.classList.contains('darkTrackedForced')){
                            e.classList.add('darkTrackedForced');
                            if(e.querySelector('form')){
                if(categoryListed['Forced
```

```
Enrollment']!=undefined){
                                    categoryListed['Forced
Enrollment']=categoryListed['Force Enrollment']+1;
                                }else{
                                    categoryListed['Forced
Enrollment']=1;
                                }
highlight(e, "Forced Enrollment");
                                count++;
                            }else{
                                if(categoryListed['Forced
Action']!=undefined){
                                    categoryListed['Forced
Action']=categoryListed['Forced Action']+1;
                                }else{
                                    categoryListed['Forced Action']=1;
                                }
                                highlight(e, "Forced Action");
                                count++;
                            }
                        }
                    }

                }
            }
        });
```

Tooltip

```
function add_tooltip_dark_tracker(){
    var styles  = '';
```

styles +='.darktracker-show-tooltip {position: relative;display: inline-block;border-bottom: 1px dotted black;}';

styles +='.darktracker-show-tooltip .darktracker-show-tooltiptext {visibility: hidden;width: 215px;background-color: #555;color: #fff;text-align: center;border-radius: 6px;padding: 5px 0;position: absolute;z-index: 1;bottom: 125%;left: 50%;margin-left: -110px;opacity: 0;transition: opacity 0.3s;font-size: 12px;}';

```
    styles +='.darktracker-show-tooltip .darktracker-show-
tooltiptext::after {content: "";position: absolute;top: 100%;left:
50%;margin-left: -5px;border-width: 5px;border-style: solid;border-
color: #555 transparent transparent transparent;}';
    styles +='.darktracker-show-tooltip:hover .darktracker-show-
tooltiptext {visibility: visible;opacity: 1;}';

    /* Create style element */
    var css = document.createElement('style');
    // css.type = 'text/css';

    if (css.styleSheet)
        css.styleSheet.cssText = styles;
    else
        css.appendChild(document.createTextNode(styles));

    /* Append style to the head element */
    document.getElementsByTagName("head")[0].appendChild(css);
}
```

Process Library

```
import re

import pandas
class Process:
    request = False
    def __init__(self,request):
        self.request=request

    def getCat(self,catId):
        cat = pandas.read_csv('data/categories.csv', skiprows=0,
error_bad_lines=False)

        categoryList = cat['category']
        categoryId = cat['id']

        for x in range(len(categoryId)):
            if categoryId[x] == catId:
                return categoryList[x]
```

```
    def checkPattern(self):

        ptn = pandas.read_csv('data/pattern.csv', skiprows = 0,
error_bad_lines=False)
        check_null=pandas.notnull(ptn['pattern']);
        ptn=ptn[check_null]
        pattern = ptn['pattern']
        pattern_len=len(pattern)
        category = ptn['category']

        resp={}
        i = 0

        for value in self.request:
            resp[i]=0
            for x in range(pattern_len):
                chck = re.search(pattern[x],value.lower().strip(),
re.IGNORECASE)
                if chck:
                    resp[i]=self.getCat(category[x])
            i = i+1

        return {'resp':resp}

    def saveList(self):
        item=self.request
        data = pandas.DataFrame([[item['url'],
item['count'],item['category']]], columns=['url', 'total_count',
'categories'])
        data.to_csv('data/collection.csv', index=False,
na_rep='Unknown', mode = 'a', header = False)
        return {'resp':'saved'}
```

Patter list, Pattern Category, Track List

class tracerList(APIView):

```
permission_classes = [AllowAny]

    def get(self, request):
        pattern1 = patterns.objects.all()
        serializer = patternSerializer(pattern1, many=True)
        return Response(serializer.data)
```

```
    def post(self, request):

        return Response(request.data)

class checkPattern(APIView):

    def post(self, request):
        category = Process(request.data.get('items'))
        response = Response(category.checkPattern())
        return response

    def get(self,request):
        return Response(request.data)

class saveTrackList(APIView):

    def post(self, request):
        savedata = Process(request.data.get('items'))
        response = Response(savedata.saveList())
        return response

    def get(self,request):
        return {}
class deleteCategory(APIView):

    def post(self,request):
        status_code=0
        form = request.POST
        cat=int(form['index'])
        categoryFile='data/categories.csv'
        collection = pandas.read_csv(categoryFile, skiprows=0,
error_bad_lines=False)
        df_new = collection[collection['id'] == cat ]
key_list = list(df_new['id'].keys())

        collection.at[key_list[0],'category']=''
        collection.to_csv(categoryFile, encoding='utf-8', index=False)
        status_code=1
return JsonResponse({
            'status_code':status_code,
            'message':"Successfull!"
        })
```

```
    def get(self,request):
        return {}

class updateCategory(APIView):

    def post(self,request):

        form = request.data
        cat=int(form['index'])
        categoryFile='data/categories.csv'
        collection = pandas.read_csv(categoryFile, skiprows=0,
error_bad_lines=False)
        df_new = collection[collection['id'] == cat ]

        key_list = list(df_new['id'].keys())

        short_desc = str(form['short_desc'])

collection.at[key_list[0],['category','short_desc','link']]=[form['new
Cat'],short_desc,form['link']]
        collection.to_csv(categoryFile, encoding='utf-8', index=False)

        return JsonResponse({
            'form':form,
            'ok':1,
            'status':'updated'

        })

    def get(self,request):
        return {}

class addCategory(APIView):

    def post(self,request):
        form = request.data

        cat = pandas.read_csv('data/categories.csv', skiprows=0,
error_bad_lines=False)
        categoryIDs = cat['id']
        last_id=categoryIDs[cat.index[-1]]
        newID = last_id + 1
        short_desc = str(form['desc'])
        data =
pandas.DataFrame([[newID,form['category'],short_desc,form['link']]],
columns=['id','category','short_desc','link'])
```

```
        data.to_csv('data/categories.csv', index=False,
na_rep='Unknown', mode = 'a', header = False)

        return JsonResponse({'ok':1,'status':'added'})

    def get(self,request):
        pass

class addPattern(APIView):

    def post(self,request):
        form = request.data

        pat = pandas.read_csv('data/pattern.csv', skiprows=0,
error_bad_lines=False)
        patternIDs = pat['id']
        last_id=patternIDs[pat.index[-1]]
        newID = last_id + 1
        data =
pandas.DataFrame([[newID,form['pattern'],form['category']]],
columns=['id','pattern','category'])
        data.to_csv('data/pattern.csv', index=False, na_rep='Unknown',
mode = 'a', header = False)

        return JsonResponse({'ok':1,'status':'added'})


    def get(self,request):
        pass

class deletePattern(APIView):

    def post(self,request):
        status_code=0
        form = request.POST
        patID=int(form['index'])
        ptnFile='data/pattern.csv'
        ptnRows = pandas.read_csv(ptnFile, skiprows=0,
error_bad_lines=False)
        df_new = ptnRows[ptnRows['id'] == patID ]

        key_list = list(df_new['id'].keys())

        ptnRows.loc[ptnRows['id'] == patID, ['pattern','category']] =
['','']
        ptnRows.to_csv(ptnFile, encoding='utf-8', index=False)
        status_code=1
```

```
        return JsonResponse({
            'status_code':status_code,
            'form':int(form['index']),
            'message':"Successfull!"
        })

    def get(self,request):
        return {}

class updatePattern(APIView):

    def post(self,request):

        status_code=0
        form = request.data
        patID=int(form['index'])
        ptnFile='data/pattern.csv'
        ptnRows = pandas.read_csv(ptnFile, skiprows=0,
error_bad_lines=False)
        df_new = ptnRows[ptnRows['id'] == patID ]

        key_list = list(df_new['id'].keys())

        ptnRows.loc[ptnRows['id'] == patID, ['pattern','category']] =
[form['newPat'],form['newCat']]
        ptnRows.to_csv(ptnFile, encoding='utf-8', index=False)
        status_code=1

        return JsonResponse({
            'status_code':status_code,
            # 'form':request.data,
            # 'patID':patID,
            'message':"Successfull!"
        })
```